

When units roots matter: excess volatility and excess smoothness of long-term interest rates

Peter C. Schotman *

*Limburg Institute of Financial Economics, Maastricht University, P.O. Box 616,
6200 MD Maastricht, Netherlands*

Abstract

This paper re-examines volatility tests of the expectations model of the term structure of interest rates. In a multivariate vector autoregression (VAR) including interest rates, prices, money and output, we find that the long-term interest rate overreacts to all transitory shocks, and underreacts to all permanent shocks, irrespective of the number of unit roots and the cointegration structure in the system. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Cointegration; Efficient markets; Term structure; Unit roots; Vector autoregression; Volatility

1. Introduction

In the econometrics literature, unit roots and cointegration continue to be of great concern since the mid-1980s. A lot of research has been devoted to determine whether macroeconomic and financial time series are stationary in levels or first differences.¹ While the properties of *Trend Stationary* and *Difference Stationary* time series are vastly different in population, it is empirically difficult to discriminate between them. Indeed, for a major U.S. macroeconomic time series as real per capita GNP, Christiano and Eichenbaum (1990) show that a

* Tel.: +31-43-388-3862; fax: +31-43-388-4875.

E-mail address: P.Schotman@BERFIN.Unimaas.nl (P.C. Schotman).

¹ The term “stationary” will in most cases be used as synonymous with “integrated of order zero” (I(0)), or abbreviated as TS (“trend stationary”) when there is no confusion on the meaning. Similarly, “nonstationary” will mostly be a synonym for “integrated of order one” (I(1)) or “difference stationary”, abbreviated DS. Exact definitions are stated when necessary.

trend stationary model can fit the data as well (and as parsimoniously) as a difference stationary model. In line with work of Quah (1990), they further contend that the whole unit root issue might not be important at all. Implications of dynamic economic models need not be very sensitive to the presence of unit roots, once we move away from very tightly parameterized models. In this paper, we investigate this conjecture for the expectations model of the term structure of interest rates. Although it is confirmed that the data are uninformative on the unit root issue, we will conclude that empirical implications depend critically on the decision to treat interest rates and other macroeconomic variables as DS or TS.

The paper is closely related to the literature on volatility and variance bounds tests, which started with Shiller's (1979) observation that long-term interest rates seemed to fluctuate too much to accord with rational expectations or efficient markets. The volatility tests have been criticized on many grounds, but the most pervasive issue revolves around whether or not interest rates have unit roots.² Early contributions by Sargent (1979) and Shiller (1981) already demonstrated the difference in empirical conclusions when time series are first differenced or assumed stationary. Under stationarity the theory seemed to be rejected overwhelmingly. Campbell and Shiller (1987) explicitly impose cointegration between long- and short-term interest rates. Within this framework there remains no evidence of excess volatility in the term structure.

The approach in this paper extends the methodology of Campbell and Shiller (1987, 1991) to provide further insights in what exactly is wrong with the expectations theory. The actual volatility of long-term interest rates is compared to the volatility implied by the net present value model of the term structure within several vector autoregressions. Apart from a long- and a short-term interest rate, our VAR's contain three more macroeconomic variables: prices, real output and money. This enables us to analyze the impact of different macroeconomic shocks. By putting restrictions on the VAR, we can make alternative assumptions on the relative size of the permanent and transitory shocks. We can also link the permanent shocks by assuming various patterns of cointegrating relations. All these models will fit the empirical data about equally well. Yet the models differ enormously in their implications regarding volatility. There is, however, one dominant regularity that all these models share. Irrespective of the number of unit roots and the pattern of cointegration, it appears that the long-term interest rate overreacts to every transitory shock and underreacts in response to every permanent shock.

The paper is organized as follows. Section 2 describes the data and introduces the notation. It also presents the stylized volatility facts and reproduces the variance bounds statistics of Shiller (1979) and West (1988) for our dataset. These statistics do not depend on a specific time series model of the short-term interest

² See LeRoy (1984, 1989) for general surveys of the variance bounds literature.

rate, and therefore show that the sensitivity with respect to unit roots is not an artifact of working with a VAR. Section 3 deals with the time series representation of long-term interest rates and the assumptions underlying the volatility tests. Section 4 contains empirical results based on comparing a variety of VAR's. Econometric issues are discussed in Appendices A and B. In particular we deal with several drawbacks of standard asymptotic statistical tests for long horizon present value models with autoregressive roots close to unity. To overcome them we will use Monte Carlo integration to obtain the empirical distribution of forecast error variances. Section 5 contains conclusions.

2. Stylized volatility facts

In a discrete time setup, the linearized form of the expectations model of the term structure relates a long-term interest rate $R_t^{(n)}$ with maturity n periods to the one-period short-term interest rate $R_t^{(1)}$. For the yield to maturity on coupon bonds that are linearized around par, the model specializes to³

$$R_t^{(n)} = \frac{1 - \delta}{1 - \delta^n} \sum_{i=0}^{n-1} \delta^i \mathbf{E}[R_{t+i}^{(1)} | I_t] + \phi^{(n)}, \quad (1)$$

where $\phi^{(n)}$ represents a liquidity or risk premium that is assumed constant over time, and where $\delta = (1 + \mu)^{-1}$ with μ the interest rate around which bond prices are linearized. The notation $\mathbf{E}[\cdot | I_t]$ denotes conditional expectations with respect to the market information set I_t . The short-term interest rate $R_t^{(1)}$ is known at the end of period t and applies to the period from t to $t + 1$. Eq. (1) expresses the long rate as a weighted average of the current and expected future short rate.

The interest rate data in this paper consist of two time series of monthly interest rates for the United States, sampled on the last trading day of the month for the period January 1962 through June 1990.⁴ The first series is the yield on a 3-month Treasury Bill. The long rate is the yield to maturity on 10-year government bonds. Fig. 1 shows the levels and first differences of the long rate and the 3 months rate. Two features of the data help in interpreting formal tests results later. First, the levels of short and long rates have about the same sample variance over the full 30-year period. Yet, the long rate is considerably smoother, since the standard deviation of its difference is much smaller than the standard deviation of changes in the short rate.

The maturity of the short rate (3 months) does not coincide with the observation frequency (monthly). This creates some technical econometric problems due to overlapping data that can most easily be dealt with by modifying Eq. (1). The

³ See Campbell and Shiller (1991) and Shiller (1979).

⁴ All data were kindly provided by the Federal Reserve Bank of Minneapolis.

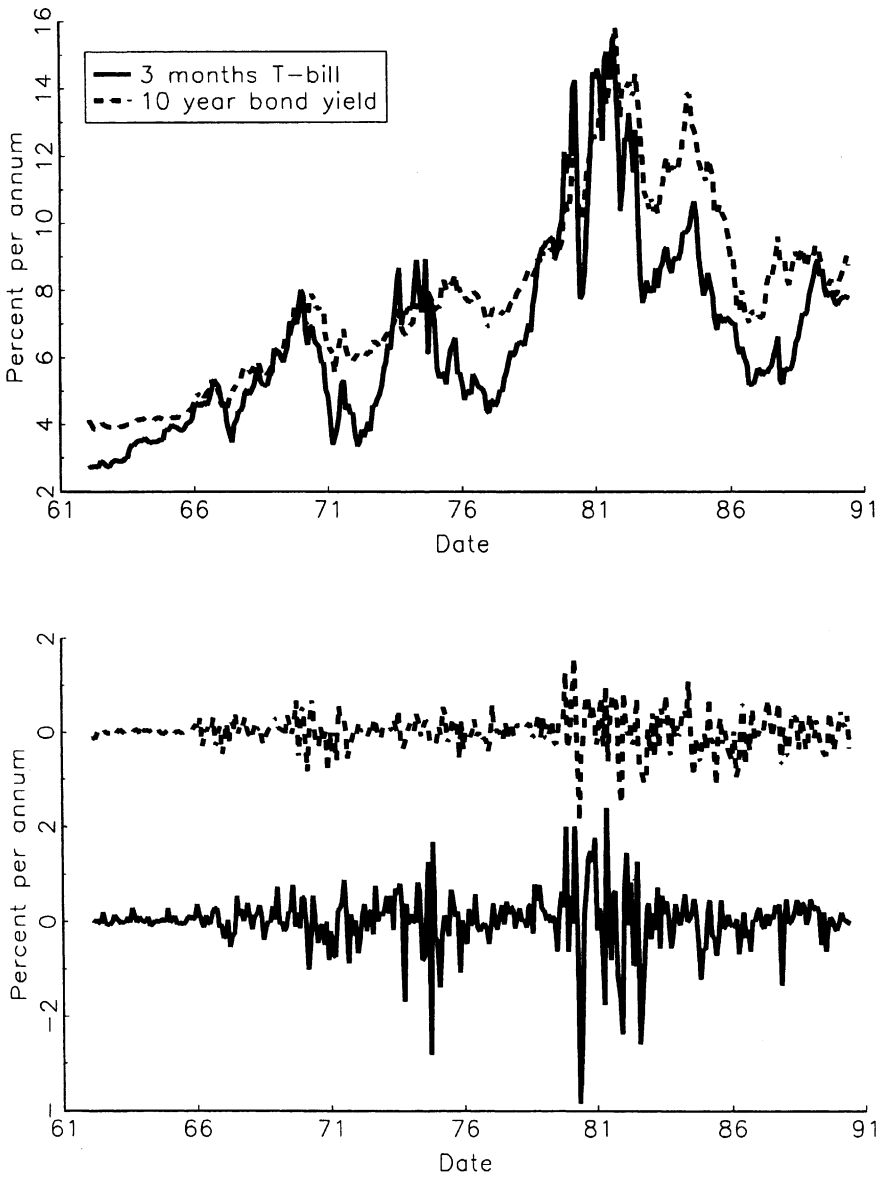


Fig. 1. Levels and first differences of interest rates. The figure shows monthly time series of the US 3-month T-bill rate and the 10-year government bond yield for the period January 1962–June 1990. The top panel is levels, the bottom panel first differences.

Table 1
Univariate unit root tests

Series	$p = 0$	$p = 12$
3-Month T-bill rate	–2.33	–2.28
10-Year bond yield	–1.64	–1.71
Spread	–3.60 *	–3.59 *

Test statistic is the adjusted t -statistic in Phillips and Perron (1988); p is the number of additional lags. Critical values at the 1%, 5% and 10% level are –3.64, –2.88 and –2.57, respectively (see Fuller 1976, Table 8.5.2). An asterisk denotes significance at the 5% level.

required modification of Eq. (1) is easily obtained by applying the expectations hypothesis twice for bonds with maturities m and n , and assuming that $k = n/m$ is an integer:

$$R_t^{(n)} = \frac{1 - \delta^m}{1 - \delta^{mk}} \sum_{h=0}^{k-1} \delta^{mh} \mathbf{E}[R_{t+mh}^{(m)} | I_t] + \phi^{(n)} - \phi^{(m)}, \quad (2)$$

linking an m -period rate to a longer n -period rate. Eq. (2) has the same structure as Eq. (1). To simplify notation, the superscripts will from now on be omitted. The long rate is represented as $R_t = R_t^{(n)}$, the short rate is called $r_t = R_t^{(m)}$, and the discount factor becomes $\gamma = \delta^m$. We also drop the risk premium, though a constant is always included in the empirical work. Finally the shorthand $\mathbf{E}_t[\cdot]$ is used for $\mathbf{E}[\cdot | I_t]$ when there can be no confusion on the interpretation of the information set. In this notation Eq. (2) simplifies to

$$R_t = \frac{1 - \gamma}{1 - \gamma^k} \sum_{i=0}^{k-1} \gamma^i \mathbf{E}_t[r_{t+mi}]. \quad (3)$$

Since most of the discussion in this paper centers on the effects of imposing unit roots, we start by examining the results of standard univariate unit root tests. Results of the Phillips and Perron (1988) test are reported in Table 1. The tests cannot reject the null hypothesis of a unit root in the level of the three interest rate time series. But the spread between any two interest rates is stationary according to the test. So if interest rates are integrated, they are also cointegrated.⁵

Although a classical test cannot reject the unit root, this by no means implies that we must accept the existence of a unit root. The tests take the unit root as the null hypothesis, and have notoriously low power. The size of the type II error is so large that the possibility that interest rates are $I(0)$ cannot be ignored. This is one of the points that motivated the Bayesian analysis in Schotman and Van Dijk (1991a,b).

⁵ This is a standard result. See, for example, Stock and Watson (1988), Campbell and Shiller (1987) or Hall et al. (1992).

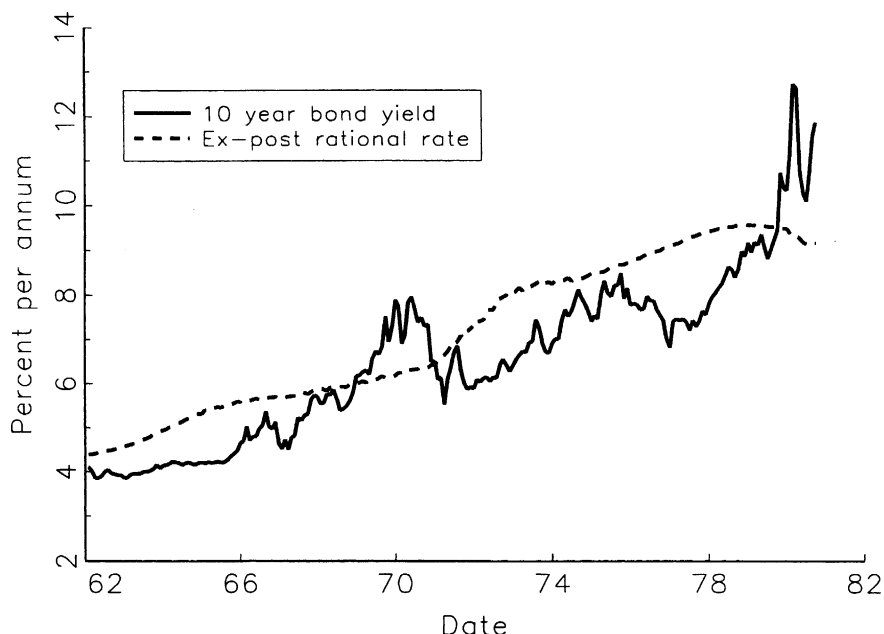


Fig. 2. Ex-post rational rate. The figure shows monthly time series of the 10-year government bond yield and the ex-post rational long rate for the period January 1962–May 1980.

The effect of unit roots on term structure tests can be seen most easily using the concept of the ex-post rational rate R_t^* introduced by Shiller (1979), which is defined as

$$R_t^* = \frac{1 - \gamma}{1 - \gamma^k} \sum_{i=0}^{k-1} \gamma^i r_{t+mi}, \quad (4)$$

and differs from the actual long rate only in replacing expectations by realizations of the short rate. Under rational expectations the forecast error $v_t = R_t^* - R_t$ is uncorrelated with all other variables in agents' information set I_t . West (1988) notes that for any information set $H_t \subset I_t$, it therefore holds that:

$$\text{Var}[R_t^* | H_t] > \text{Var}[R_t | H_t]. \quad (5)$$

This variance inequality can be easily verified by constructing a time series of R_t^* and projecting both R_t and R_t^* on the H_t . Since the time to maturity of the long rate is 10 years and the sample is long enough, the series R_t^* can be constructed exactly without any further assumptions using Eq. (4).⁶ The drawback of the exact

⁶ The exact calculation of the ex-post rational long rate circumvents the problems with the usual backward recursion $R_t^* = \gamma R_{t+m}^* + (1 - \gamma)r_{t+m}$, which requires some terminal condition like $R_t^* = R_t$. The volatility tests for the term structure are thus simpler than the analogous tests for the stock market, where the present value relation has an infinite horizon.

Table 2

Variance of ex-post rational long rate

Series	Unconditional variance	Conditional variance given R_{t-1} and r_{t-1}
R_t	5.36	0.096
R_t^*	2.91	0.604

R_t^* is the ex-post rational long rate constructed using actual short rates as in Eq. (4). The conditional variance is the residual variance of the regression:

$$x_t = \alpha + \beta_1 r_{t-1} + \beta_2 R_{t-1} + e_t,$$

where x is R and R^* , respectively. Sample period is 62:1 to 80:9.

calculation is the loss of 10 years of observations at the end of the sample. Fig. 2 shows the actual and ex-post rational long rate. It has been assumed that $\delta = 0.994$, which corresponds to an annual discount rate of 7.5%, approximately the sample mean of the short rate. In the linearized framework, this gives the 10-year bond an effective duration of $D(n) = (1 - \delta^n)/(1 - \delta) = 7.14$ years.

The smooth behavior of R_t^* and the sample unconditional variances of the time series reported in Table 2 give the impression that the variance bound is grossly violated. But the sample unconditional variances are uninterpretable if interest rates have a unit root. However, since the variance inequality must hold for any information set H_t (not only $H_t = \{\text{constant}\}$), we can remove the possible nonstationarity by conditioning on the past levels of the short and long rates.⁷ Table 2 shows that the variance bound is now easily satisfied. The violation of the bound thus seems closely related to the existence of unit roots. As emphasized by Cochrane (1991), a formal volatility test is equivalent to an orthogonality test, testing the null hypothesis that $R_t^* - R_t$ predictable with information dated time t or earlier. These orthogonality tests will depend on the information set used, and in particular whether we will allow levels variables in H_t . The remainder of the paper considers the effects of changing the information structure H_t and the sensitivity of results with respect to the unit root assumptions in a VAR framework.

3. Time series assumptions and implications

To gain further insight in the way the presence of unit roots affects the volatility of long-term interest rates, we need additional assumptions. In particular, it will be necessary to specify the behavior of short-term interest rates in more

⁷ Both R and R^* are conditioned on the same information set $H_t = \{r_{t-1}, R_{t-1}, \text{constant}\}$ and therefore not subject to the criticism of Kleidon (1986, pp. 961–962): “confusion in interpretation of time series plots of price and $p^*(t)$ stems from comparing the conditional variance of price, $\text{var}\{p(t) | p(t-k)\}$, with an inappropriate conditional variance of $p^*(t)$, $\text{var}\{p^*(t) | p^*(t-k)\}$ ”.

detail. Given a data generating process (DGP) of the short rate, we can explicitly calculate future expectations of the short rate.

The assumed DGP of the short rate is a general moving average process,

$$\Delta r_t = c(L) \epsilon_t = \sum_{i=0}^{\infty} c_i \epsilon_{t-i}, \quad (6)$$

where L is the lag operator, c_i are $(1 \times K)$ vectors of parameters, $c(1)$ is bounded, and ϵ_t is a $(K \times 1)$ vector of innovations with mean zero and identity covariance matrix. The contemporaneous covariances of the shocks are modeled through c_0 . The presence of multiple shocks ($K > 1$) allows for distinction between permanent and transitory shocks. A shock ϵ_{it} (the i th element of ϵ_t) is transitory if $c_i(1) = \sum_{j=0}^{\infty} c_{ij} = 0$. If all shocks are transitory, the lag polynomial is divisible by the difference operator $(1 - L)$, implying overdifferencing of the original level time series. Having multiple shocks also opens the possibility of Granger causality running from long to short rates as well as in the opposite direction. Although K sources of stochastic uncertainty are introduced, not all of these will be identifiable using interest rate data alone. Campbell and Shiller (1987) have stressed that to avoid inconsistencies between the actual information set of agents and the information set of the econometrician, we must at least include the lagged values of the long- and short-term interest rates in our empirical information set. This entails that the innovation in the long rate is one of the shocks in the model. We must further assume that the agents have more information than the econometrician. Section 4 discusses the empirical identification of the shocks in more detail.

It will be convenient to work with the spread $S_t^{(n,m)} = R_t^{(n)} - R_t^{(m)}$. Omitting the superscripts as in Section 2, we get $S_t = R_t - r_t$. The relation between the spread and the short rate follows from Eq. (3) as:

$$S_t = \sum_{i=1}^{k-1} \frac{\gamma^i - \gamma^n}{1 - \gamma^n} \mathbf{E}_t[\Delta_m r_{t+mi}], \quad (7)$$

where $\Delta_m r_t = \sum_{j=0}^{m-1} \Delta r_{t-j}$. Optimal forecasts of Δr_{t+h} can be obtained from Eq. (6) as:

$$\mathbf{E}_t[\Delta r_{t+h}] = \sum_{j=h}^{\infty} c_j \epsilon_{t+h-j} = \sum_{j=0}^{\infty} c_{h+j} \epsilon_{t-j}. \quad (8)$$

Substituting Eq. (8) into Eq. (7) and rearranging, one obtains the implied time series process for the spread as:

$$S_t = \sum_{j=0}^{\infty} \left(\sum_{h=0}^{m-1} \sum_{i=1}^{k-1} \frac{\gamma^i - \gamma^k}{1 - \gamma^k} c_{mi+j-h} \right) \epsilon_{t-j}. \quad (9)$$

Noting that $\Delta R_t = \Delta S_t + \Delta r_t$, the long-term interest rate is obtained by summing Eqs. (9) and (6). Recollecting term in ϵ_{t-j} gives:

$$\begin{aligned}\Delta R_t &= \left(c_0 + \sum_{i=1}^{k-1} \frac{\gamma^i - \gamma^k}{1 - \gamma^k} \sum_{h=0}^{m-1} c_{mi-h} \right) \epsilon_t + \sum_{j=1}^{\infty} \left(\frac{1 - \gamma^k}{1 - \gamma^k} \sum_{i=0}^{k-1} \gamma^i c_{mi+j} \right) \epsilon_{t-j} \\ &\equiv \sum_{j=0}^{\infty} g_j \epsilon_{t-j}.\end{aligned}\quad (10)$$

The long-run impact of a shock to the system is defined as $c(1)$ for the short rate process, and as $g(1)$ for the n -period interest rate. Summing the coefficients in Eq. (10) establishes the important property,

$$g(1) = c(1), \quad (11)$$

which implies that interest of all maturities cointegrate, whenever a single component of $c(1)$ is non-zero. If $c(1) = 0$, all interest rates are stationary.

Representation (10) simplifies considerably if $m = 1$ and/or $n \rightarrow \infty$. Although the infinite maturity assumption $n \rightarrow \infty$ would be analytically convenient, it might introduce severe dynamic misspecification. First, with finite maturity n , the coefficient g_j depends on the first $j + n$ entries of the $c(L)$ polynomial. If unit roots would only restrict the very long memory properties of the short rate without affecting the short and medium term dynamics, the first few entries of $g(L)$ would not be very sensitive with respect to unit roots. Second, if the actual maturity of the long-term bond is about 10 years, then a specification with $n \rightarrow \infty$ puts too much weight on expected short-term interest rates in the distant future. With a discount rate of 7.5%, the sum of the factors over the first 40 quarters is only $(1 - \gamma) \sum_{i=0}^{39} \gamma^i = 0.51$. The first 10 years of expected interest rates make up only half of the weights of an infinite maturity rate. Assuming an infinite maturity when the actual maturity is “only” 10 years restricts the long-term interest rate to behave more smoothly than necessary, especially if the assumed DGP of the short rate is stationary.

Representation (10) contains all conditions implied by the expectations model of the term structure. All cross equation restrictions can in principle be tested by comparing the implied process in Eq. (10) with an unrestricted representation,

$$\Delta R_t = \tilde{g}(L) \epsilon_t = \sum_{j=0}^{\infty} \tilde{g}_j \epsilon_{t-j}. \quad (12)$$

A test of all the conditions implied by the expectations hypothesis entails that $\tilde{g}(L) = g(L)$. Here we will consider violations of the model in a particular direction, and thus focus on a subset of all the implications. One departure from the expectations model is the excess volatility phenomenon, which means that the variance of the long rate is larger than the variance implied by the expectations theory.

The volatility tests will be based on the variance of the j -step forecast errors of the long rate, defined as

$$\sigma_j^2 = \mathbf{E}_t \left[\left(R_{t+j} - \mathbf{E}_t [R_{t+j}] \right)^2 \right] = \sum_{i=0}^{j-1} \psi_i \psi_i', \quad (13)$$

where $\psi_i = \sum_{h=0}^i g_h$, and g_h is given by Eq. (10). The parameters σ_j^2 are still functions of the parameters that describe the time series process of the short-term interest rate r_t . The expectations hypothesis is capable of explaining the volatility of the long-term interest rate, if the predictive variances defined in Eq. (13) match the predictive variances obtained from an unrestricted time series model of the long-term interest rate like Eq. (12), say $\tilde{\sigma}_j^2 = \sum_{i=0}^{j-1} \tilde{\psi}_i \tilde{\psi}_i'$, with $\tilde{\psi}_i = \sum_{h=0}^i \tilde{g}_h$. Under the null hypothesis $g_j = \tilde{g}_j$, and we must have $\sigma_j^2 = \tilde{\sigma}_j^2$ for all j .

The predictive variances in Eq. (13) converge to the unconditional variance of the long rate when $j \rightarrow \infty$. If the long rate is integrated, the unconditional variance is infinite and the limiting behavior of the predictive variances satisfies:

$$\lim_{j \rightarrow \infty} \frac{\sigma_j^2}{j} = g(1)g(1)',$$

$$\lim_{j \rightarrow \infty} \frac{\tilde{\sigma}_j^2}{j} = \tilde{g}(1)\tilde{g}(1)'. \quad (14)$$

Therefore, if we impose cointegration of the long and the short rate, which means $g(1) = \tilde{g}(1)$, the ratio $\sigma_j^2 / \tilde{\sigma}_j^2$ will approach unity when $j \rightarrow \infty$, regardless of whether the other term structure restrictions hold. Under the maintained hypothesis of cointegration, the variance ratio implications are automatically satisfied in the limit.⁸ Under stationarity, there is no automatic restriction on the limiting unconditional variance of the long-term rate.

Volatility is but one aspect of the term structure. Since the scalar variances σ_j^2 are a limited set of nonlinear functions of the original $g(L)$ polynomial, the number of restrictions implied by Eq. (13) is less than the full set of restrictions implied by Eq. (10). The test can be more or less powerful than a test of all conditions depending on the actual deviation between the data and the model.

Unit roots enter the model because some of the shocks to the short rate can be persistent; others might be purely transitory. Suppose we arrange the shocks so that we can partition ϵ_t as $(\epsilon'_{1t}, \epsilon'_{2t})'$, where the first K_1 shocks are transitory, and the remaining $K_2 = K - K_1$ shocks are persistent. That means that we restrict the first K_1 elements of $c(1)$ to be equal to zero. Partitioning the vectors c_j , g_j and ψ_j

⁸ See also Phillips and Perron (1988) on the limiting behavior of these variance bound ratios.

conformably with ϵ_t we can write the variance of the j -period ahead forecast error as the sum of the variance due to the transitory shocks, $\sigma_j^2(T)$, and the variance due to persistent shocks, $\sigma_j^2(P)$:

$$\sigma_j^2 = \sigma_j^2(T) + \sigma_j^2(P) = \sum_{i=0}^{j-1} \psi_{1i} \psi'_{1i} + \sum_{i=0}^{j-1} \psi_{2i} \psi'_{2i}. \quad (15)$$

An analogous decomposition holds for the unrestricted predictive variances. Departures from the expectations model can show up in both components of the total volatility.

For the empirical analysis, we need some assumptions on the shape of the infinite MA representation. The most convenient representation for estimation purposes is a vector autoregression (VAR). The advantage of a VAR is that estimation, and imposing cointegrating vectors, is computationally straightforward. Once the VAR has been estimated, we can compute the sequences of predictive variances as functions of the parameters of the VAR. The details are spelled out in Appendix A. By a suitable decomposition of the error covariance matrix of the VAR, the first K_1 shocks are constructed to be purely temporary. The remaining K_2 shocks are orthogonal to the temporary shocks and have permanent effects. In Appendix A, it is also proven that for any given VAR, the decomposition of the total predictive variance into a part due to pure transitory shocks and one due to persistent shocks is unique at every forecast horizon. In other words, the quantities $\sigma_j^2(T)$ and $\sigma_j^2(P)$ in Eq. (15) are unique (although ψ_{1i} and ψ_{2i} are subject to the usual rotation indeterminacy).

Statistical inference is carried out by Monte Carlo integration starting from the asymptotic covariance matrix of the VAR parameters. Monte Carlo integration is an exact numerical method to obtain the distribution of transformations of random variables. Computational details are in Appendix B.

4. Empirical results

The interest rate data have been described in Section 2. In addition, three macroeconomic variables are included in the VAR: industrial production (y), Money Stock (M), and the Consumer Price Index (p). All data are seasonally adjusted series taken from the Citibase tape. These three series enter the VAR in logarithms and after detrending. We decided to first detrend the data in order to make the results comparable across different VAR's. The macroeconomic series are detrended by regressing on a constant and time trend if a series is assumed stationary, and by regressing first differences on a constant in case a series is assumed I(1). The two interest rates are in deviation from the sample mean. No constant term or trend is included in the VAR. All differences in test results are

Table 3
Summary statistics of VAR's

(A) Measures of fit and system roots

Model	σ_r	σ_R	ρ_{rR}	λ_1	λ_2	λ_3	λ_4	λ_5
A: Bivariate I(0)	0.593	0.371	0.629	0.982	0.955			
B: Stationary VAR	0.588	0.371	0.633	0.994	0.994	0.984	0.984	0.967
C: Bivariate I(1)	0.551	0.357	0.585	1	0.957			
D: 4 unit roots	0.571	0.363	0.586	1	1	1	1	0.985
E: 3 unit roots	0.565	0.360	0.578	1	1	1	0.982	0.968

(B) Proportion of variance due to transitory components

Model	Short rate	Long rate (unrestricted)	Long rate (restricted)
A: Bivariate I(0)	1.00	1.00	1.00
B: Stationary VAR	1.00	1.00	1.00
C: Bivariate I(1)	0.02	0.45	0.00
D: 4 unit roots	0.02	0.53	0.00
E: 3 unit roots	0.72	0.65	0.18

Specification of different VAR models (A to E) is explained in the text. σ_r is the innovation standard error of the short rate; σ_R is the innovation standard error of the long rate; ρ_{rR} is the correlation between the innovations; λ_i , ($i = 1, \dots, 5$) are the largest roots of the system. The variance decomposition is described in the text.

thus entirely due to differences in the dynamic specification of the VAR, and not due to differences in estimated growth rates and long-term means.⁹

Five different VAR's are compared:

- (A) 24th order bivariate VAR in levels, containing only the two interest rates. Interest rates are I(0) in this model.
- (B) A 12th order VAR with all five variables; all series are assumed stationary.
- (C) A 24th order bivariate VAR with a single unit root. Both interest rates are I(1), but cointegrate with cointegrating vector $\beta = (1 - 1)'$.
- (D) A 12th order VAR with all five variables and 4 unit roots. All individual time series are I(1), but the two interest rates cointegrate with cointegrating vector $\beta = (1 - 1)'$.

⁹ The results with trends and constants included unrestrictedly in the VAR are qualitatively similar, except when a time trend is allowed for the two interest rates.

- (E) A 12th order VAR with all five variables and 3 unit roots. All individual series are $I(1)$, and there is one additional cointegrating vector, linking the short rate to velocity: $y_t + p_t - M_t - 0.066r_t$ is assumed to be a stationary series.¹⁰

Comparing the two bivariate models (A) and (C) shows the effects of imposing a unit root on the interest rate time series. Models (B) and (D) convey the same information within a multivariate model. Comparing the two pairs of models can highlight the sensitivity with respect to the information set used to forecast interest rates. In model (E) the macroeconomic variables are more than just another source of information, since the cointegration relation restricts the permanent part of interest rate shocks by linking them to a linear combination of the macroeconomic shocks.

Table 3 gives summary statistics of the different VAR's, indicating that the unit root restrictions do not greatly affect the fit of the models. The five variable VAR's provide some improvement over the bivariate VAR for the equation of the short rate.¹¹ The estimated roots of the VAR also suggest that the unit root restrictions are empirically plausible. The largest roots of the stationary VAR are very close to unity. Again, although we cannot reject the unit root hypothesis, we also cannot reject stationarity.

Fig. 3 gives an overview of the volatility implications of the different multivariate VAR's. The lines in the figure show the standard deviation of the long rate (σ_j) over various forecast horizons (j) implied by the expectations model of the term structure and conditional on a VAR with 0, 3 and 4 unit roots, respectively.

¹⁰ The cointegrating vector has been estimated by an OLS regression of velocity on the short-term interest rate. The Phillips–Perron test of the residuals, adjusted for 12 lags, gives a t -statistic of -3.08 , which is close to being significant at the 5% level (see Engle and Granger, 1987). The ML test for cointegration between velocity and the short rate within a bivariate system (24 lags) yields $\lambda_{\max} = 6.6$, which is significant at the 10% level (see Johansen and Juselius, 1990, Table A3).

¹¹ Formal Granger causality tests (not reported in the tables) reveal more about the dynamic structure of the VAR's. For all five VAR's there is Granger causality in both directions between the two interest rates. In the cointegrated model, the spread is significantly error correcting in the equation for the change in the long rate. The causality pattern in the five variable VAR's depends on the number of unit roots and the parameterization of the VAR. Model (E) provides an example. Using F -tests and a 10% significance level, the causality structure can be summarized in the following matrix (+ denotes significance):

	$y + p - M - \alpha r$	$R - r$	Δy	Δp	ΔM	Δr	ΔR
Δy	–	–	+	–	–	–	–
Δp	–	–	–	+	–	–	–
ΔM	–	–	–	+	+	+	+
Δr	+	+	+	–	+	+	+
ΔR	+	+	+	–	+	+	+

Most of the error correction takes place through the interest rates.

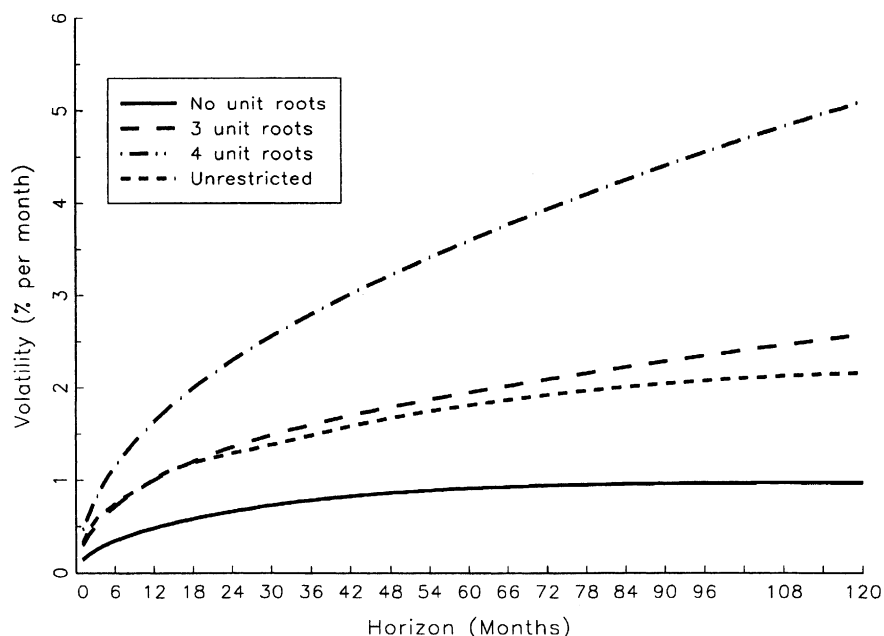


Fig. 3. Implied volatility. The figure shows the actual volatility of the long rate and volatility of the long rate implied by three different VAR's: models B, D and E in the text. The unrestricted volatility shown in the figure is computed with model B.

The ranking is clear: the more unit roots in the system, the higher the implied standard deviation of the long rate. The figure illustrates the sensitivity of the volatility measures with respect to the presence of unit roots. The unrestricted estimates $\hat{\sigma}_j$ are close to the VAR with 3 unit roots. The unrestricted estimates for a VAR with 3 or 4 unit roots (not shown in the figure) almost coincide with those of the stationary VAR for the first 80 periods, again illustrating that the estimated VAR's are almost indistinguishable statistically. Of course the $\hat{\sigma}_j$ sequences for the models with unit roots will eventually diverge to infinity as the forecast horizon j increases.

To investigate the statistical significance of the deviations between actual and implied volatility, we used the Monte Carlo integration; results are in Table 4. Panel A of the table shows that the actual volatility $\tilde{\sigma}_1$ —the one period ahead forecast error variance, which is just the standard error of the residuals of the equation of the long rate in the VAR—is estimated quite precisely. It does not vary greatly over the alternative models. However, the implied volatilities σ_1 are estimated less precisely and differ substantially across the models. The third column in Table 4 presents the estimated probabilities of excess volatility for each of the models. The estimates confirm the point estimates in Fig. 3. There is significant excess volatility if we believe in a stationary model. The probabilities

Table 4
Volatility of long-term interest rates

	$E[\tilde{\sigma}_1]$	$E[\sigma_1]$	$\Pr[\tilde{\sigma}_1 > \sigma_1]$	$E[\tilde{\sigma}_{120}]$	$E[\sigma_{120}]$	$\Pr[\tilde{\sigma}_{120} > \sigma_{120}]$
<i>(A) Total volatility</i>						
A: Bivariate I(0)	0.345 (0.015)	0.213 (0.079)	0.939 (0.006)	2.715 (0.859)	1.151 (0.786)	1.000 (–)
B: Stationary VAR	0.327 (0.014)	0.159 (0.048)	0.988 (0.003)	2.559 (0.668)	1.266 (0.700)	0.997 (0.001)
C: Bivariate I(1)	0.345 (0.015)	0.473 (0.160)	0.155 (0.009)	4.879 (1.753)	5.074 (2.522)	0.544 (0.013)
D: 4 unit roots	0.333 (0.014)	0.640 (0.171)	0.042 (0.005)	5.872 (2.017)	6.499 (2.970)	0.264 (0.011)
E: 3 unit roots	0.329 (0.015)	0.360 (0.124)	0.500 (0.013)	4.077 (1.353)	3.920 (1.957)	0.773 (0.011)
<i>(B) Volatility due to transitory components</i>						
A: Bivariate I(0)	0.345 (0.015)	0.213 (0.079)	0.939 (0.006)	2.715 (0.859)	1.151 (0.786)	1.000 (–)
B: Stationary VAR	0.327 (0.014)	0.159 (0.048)	0.988 (0.003)	2.559 (0.668)	1.266 (0.700)	0.997 (0.001)
C: Bivariate I(1)	0.224 (0.061)	0.023 (0.020)	0.991 (0.002)	0.748 (0.261)	0.094 (0.127)	1.000 (–)
D: 4 unit roots	0.220 (0.048)	0.024 (0.023)	0.994 (0.002)	0.611 (0.209)	0.138 (0.152)	0.997 (0.001)
E: 3 unit roots	0.259 (0.038)	0.111 (0.064)	0.973 (0.004)	1.471 (0.647)	0.592 (0.470)	0.989 (0.003)
<i>(C) Volatility due to permanent components</i>						
A: Bivariate I(0)	0	0	–	0	0	–
B: Stationary VAR	0	0	–	0	0	–
C: Bivariate I(1)	0.249 (0.047)	0.462 (0.160)	0.051 (0.006)	4.814 (1.753)	5.073 (2.520)	0.448 (0.012)
D: 4 unit roots	0.241 (0.047)	0.538 (0.170)	0.002 (0.000)	5.836 (2.017)	6.497 (2.968)	0.228 (0.010)
E: 3 unit roots	0.194 (0.050)	0.334 (0.131)	0.083 (0.007)	3.738 (1.376)	3.748 (1.949)	0.653 (0.012)

$E[\tilde{\sigma}_j]$ and $E[\sigma_j]$ are the posterior means of the innovation standard error; posterior standard errors are in parentheses. $\Pr(\tilde{\sigma}_j > \sigma_j)$ is the posterior probability of excess volatility. All entries are based on 1500 Monte Carlo replications.

that the actual volatility is higher than the implied volatility are 0.94 and 0.99 for the two stationary models. Imposing a common stochastic trend in the two interest rates, a seemingly innocuous restriction, leads to completely opposite conclusions. The actual predictive variances are now significantly below the implied predictive variances. The variance inequality is reversed, with probabilities of excess smoothness equal to 0.85 and 0.96, respectively.¹² The model with 3 unit roots (and the cointegration between the short rate and velocity) takes a middle position just as in Fig. 3.

The last three columns of the table focus on the endpoints in the curves of Fig. 3, i.e. the long-term (120 periods ahead) variances. For the stationary VAR, the variances have converged to the unconditional variances. The Monte Carlo results provide strong evidence of excess volatility for the two stationary models ((A) and (B)). No clear evidence of excess volatility is obtained if interest rates are assumed to be cointegrated (models (C), (D) and (E)). The latter result is in line with the theory in Section 3. The variance ratio converges to unity when the forecast horizon goes to infinity.

The results suggest that the cointegration between the interest rates and velocity somehow resolves the volatility puzzle, as for this model the volatility restrictions seem to hold almost exactly. Panels B and C of Table 4 show that this is not true. Using the decomposition in transitory and persistent shocks, it appears that long rates overreact to transitory shocks, but underreact to permanent shocks. The probabilities of excess volatility with respect to transitory shocks are very high for all five models. In contrast, the probabilities of excess volatility with respect to permanent shocks are low for all five models. In model (E), the excess volatility with respect to permanent shocks happens to cancel out against the excess smoothness due to permanent shocks. Over longer horizons, the permanent components will eventually dominate, and any evidence against volatility will disappear in the limit. The two transitory components are, however, still important in the last VAR, even over a horizon of 120 months.

The variance decomposition in the second part of Table 4 shows that by varying the number of unit roots in the VAR we have succeeded in obtaining models where the transitory component in the DGP of the short rate either explains all variance (the stationary models (A) and (B)), is almost absent (models (C) and (D)), or somewhere in between (model (E)). In contrast, a purely transitory shock always has a sizable effect on the long rate. Since the long rate should reflect long-term expectations of the short rate, the expectations model implies that the permanent shocks must take account of most of the innovations to

¹² The results do not depend on assumptions about the number of unit roots in the macroeconomic variables, as long as they are not related to the nonstationarity in the interest rates. For example, the results for a VAR with a single unit root in the two interest rates and trend stationary macroeconomic variables are similar to model (D) with difference stationary macro variables. Also, a model with stationary interest rates, but integrated macroeconomic variables, is similar to the stationary VAR.

the long rate. This is evident in the last line of Table 3, which tells that the long rate must react less to the transitory shocks, and relatively more to the persistent shocks. This is another way stating the conclusion that follows from Table 4.

The tests of the expectations model do not formally incorporate all sources of model uncertainty, and will thus overstate the evidence against the expectations model. We have conditioned on the value of the discount factor $\delta = 1/(1 + \mu)$ and the order of integration of interest rate time series, and ignored changes in policy regime that can possibly alter the dynamic structure of the model. Estimation of a constant μ has been attempted by, e.g. Mankiw (1986), while Watson and Engle (1987) consider time varying discount rates. Both studies failed to obtain precise estimates of the discount factor. Since point estimates of δ larger than one (as found by Mankiw, 1986) would preclude any further analysis of the term structure based on discounted sums of MA parameters we have chosen to fix the discount factor a priori. The results are, however, robust to alternative values of δ . Even with δ as low as 0.90, corresponding to an annual discount rate of more than 100%, we continue to find the same variance inequalities. Setting δ larger than 0.994 makes the results even stronger.

We also only considered the polar cases of $I(0)$ and $I(1)$ time series, and have shown that the implied volatility is either much too big or much too small. Shea (1991) allows for the intermediate case of fractional integration ($I(d)$ with $0 < d < 1$), and argues that quantifying the uncertainty about d can greatly reduce the significance of empirical violations of the variance bounds. The theoretical MA relations in Section 3 continue to hold under fractional integration, but a VAR will misrepresent the long-run dynamics, so this is possibly a valid criticism of our approach. It is hard to investigate the alternative of fractional integration explicitly, however, since imposing the co-integration constraint $c(1) = g(1)$ poses severe technical problems, and since estimation of fractionally co-integrated systems is not well developed yet. As a practical solution we have set the lag length of the VAR at the rather high values $p = 12$ or $p = 24$ in order to restrict the MA representation as little as possible at the cost of some overparameterization.

A high order VAR requires long time series, so that it will not be feasible to look at particular subperiods, or to deal with changes in regime as in Hamilton (1988). The stochastic structural breaks modeled by Hamilton (1988) lead to nonlinear responses of the long rate to some shocks of the short rate. Again, using a long vector autoregression we hope to capture the nonlinearities by additional flexibility in the linear effects. The results of Hamilton (1988) suggest that there has been a temporary shift in the mean and the error variance of the U.S. short-term interest rate between the third quarter of 1979 and the first quarter of 1982, without any further changes in the parameters of the system. Also, the probability of regime shifts is estimated to be virtually zero after 1982. The possibility of a regime shift will thus have a very limited effect on the calculation of discounted sums of expected future short-term interest rates. The main effect of

introducing a structural break in the model is to make the model more stationary, which will lead to more excess volatility for the stationary models.

The “peso problem” points at another potential pitfall of the standard regression model in a situation when there are infrequent regime shifts. Economic agents might rationally anticipate a major regime shift, but it takes some time before the switch actually occurs. In a short sample, time series tests will find significant systematic deviations of the orthogonality conditions and hence wrongly reject the expectations model. The “peso problem” is another way of expressing the need for a nonlinear model of the short-term interest rate. It does not in any way invalidate the expectations model as represented in Eq. (1), but it suggests that a VAR might produce inadequate forecasts of future short rates. The VAR approximation might be especially poor in relatively small samples. The importance of the “peso problem” can be judged from the residuals of an estimated VAR. Around a regime shift leading to higher interest rates the equation for the long rate in a VAR will show small positive residuals before the break, and a large residual after the shift has taken place. In large samples a VAR will capture the second moments properties of the true DGP.

Table 5 provides some diagnostic statistics for general nonlinear effects. The first line shows that the errors are conditionally heteroskedastic. Introducing ARCH will lead to time dependent volatility statistics and maybe to less significant variance ratio statistics. But under the null hypothesis the covariance matrix of the errors is unrestricted, so our results are still interpretable as evidence about the average volatility. The significant skewness of the short rate is a first indication of possible nonlinear effects. The other two diagnostics in the table provide further evidence of nonlinearity, which seems to be present despite the overparameterization of the linear effects. The nonlinear reaction of the short rate to the lagged spread implies that the impulse responses will become time-dependent, with volatility depending on the current slope of the term structure. Exploring the exact of nonlinearity is beyond the scope of the paper though.

Table 5
Diagnostics of cointegrated VAR (model E)

	ΔY_t	Δp_t	ΔM_t	Δr_t	ΔR_t
LMARCH	8.82	13.1*	4.96	22.8*	15.9*
Skewness	0.06	0.54	−0.12	−0.62	0.24
Kurtosis	2.03	3.60	0.41	4.16	2.16
Nonlinear ECM	2.58	7.59	4.84	13.1*	5.02
RESET	6.82	6.80	7.14	36.2*	15.2

LMARCH is a $\chi^2(4)$ LM test for 4th order ARCH. Nonlinear ECM is a $\chi^2(4)$ test for nonlinear error correction by adding the variables S_{t-1}^2 , S_{t-1}^3 , z_{t-1}^2 , z_{t-1}^3 , where z_t is the residual of the cointegrating regression of velocity on the short-term interest rate. RESET is a $\chi^2(10)$ test for nonlinear effects formed by adding fitted values of all five equations raised to the second and third power. Skewness = $\Sigma \hat{u}_t^3 / T\hat{\sigma}^3$, Kurtosis = $\Sigma \hat{u}_t^4 / T\hat{\sigma}^4 - 3$.

5. Conclusions

One main conclusion emerges from the empirical results. Whatever the form of a VAR that is fitted to interest rate data, the long rate overreacts to pure transitory shock and underreacts to permanent shocks. The assumptions about which variables contain a unit root and how they cointegrate largely determine the empirical results with respect to volatility. But these assumptions are extremely hard, if not impossible, to test.

The econometric procedure that we used excludes a number of possible explanations for this conclusion. First, the finite maturity of 10 years of the long-term interest rate avoids transversality problems with infinite horizon models. Second, the Monte Carlo integration technique controls for some distortions due to asymptotic approximations in models with near unit roots. Third, the results hold for various specifications of the VAR, both bivariate and with additional macroeconomic variables. Fourth, since the importance of the permanent shocks in the short-term interest rate ranges from 0% to 98%, results are not sensitive to the assumed “size of the random walk” component. It remains an open question, however, how the results stand up against explicit modeling of nonlinear effects, which appear still significant even in a highly overparameterized VAR.

6. Epilogue

The paper appears as it was in February 1992. Much has happened since then, both in the econometrics as well as in the finance literature. Still, rejecting a unit root in interest rate time series and empirically determining the number of unit roots in a system is as difficult as it was. Results in this paper are all conditional on the number of unit roots and the cointegrating vectors. In the Bayesian literature, Kleibergen and Van Dijk (1994) made some progress on inference incorporating the uncertainty about the cointegration structure and the number of unit roots.

In the finance literature, the importance of mean reversion for term structure models is well understood, especially in the literature on derivatives pricing. Rather than estimating the mean reversion from historical short rate time series, the implied mean reversion of the short rate is modeled through the term structure of volatilities, reversing the methods of this paper. For example, in Black et al. (1990), the volatility of interest rates with different maturities provides information about the dynamics of the short rate. The prime example of this approach is the Heath et al. (1992) model.

The empirical results in this paper suggest that nonlinearities could be highly relevant for term structure models. The same conclusion emerges from the recent work by Dai and Singleton (2000) on affine term structure models, who point at “omitted nonlinearity”. Pfann et al. (1996) looked at simple nonlinear models, and

found that threshold regime switching models generate dynamics for the short-term interest rate that are in between $I(0)$ and $I(1)$ and that these models imply long-term interest volatility that is close to the observed volatility. Ang and Bekaert (in press) explicitly consider term structure models within a regime switching framework.

Acknowledgements

The original draft is in my dissertation. Most of the empirical research for the September 1990 version was undertaken while I was a visitor at the Institute of Empirical Macroeconomics in Minneapolis. The second revision was completed while I was a visitor at the Woodrow Wilson School of Public and International Affairs at Princeton University. I thank these institutions for their hospitality. Many people have commented on the paper. In particular I would like to thank participants at the World Congress of the Econometric Society in 1990 and at the CEPR Finance Meetings in Gerzensee in 1991.

Appendix A. VAR variance decompositions

This appendix shows how to compute the point estimates of the j -step ahead forecast error variances of the long rate defined in Eq. (8). The volatility statistics are nonlinear functions of the estimated coefficients of the p th order VAR

$$x_t = \sum_{i=1}^p A_i x_{t-i} + \eta_t, \quad (16)$$

where x is a $(K \times 1)$ vector. Let the first element in x_t be the short rate r_t , and the second element the long rate R_t . The covariance matrix of η_t is denoted Σ . Let F be the companion matrix of the VAR, obtained by reformulating the VAR as a first order system

$$\begin{aligned} z_t = \begin{pmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-p+1} \end{pmatrix} &= \begin{pmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I & & & O & O \\ & \ddots & & & \vdots \\ & & \ddots & & \vdots \\ O & & & I & O \end{pmatrix} \begin{pmatrix} x_{t-1} \\ x_{t-2} \\ \vdots \\ x_{t-p} \end{pmatrix} + \begin{pmatrix} I \\ O \\ \vdots \\ O \end{pmatrix} \\ &= F_{zt-1} + G\eta_t. \end{aligned} \quad (17)$$

If the system contains unit roots, some of the eigenvalues of F will be equal to 1.

We assume that all other eigenvalues are strictly inside the unit circle. Forecasts of future values of z_t are obtained from the prediction formula:

$$\mathbf{E}_t[z_{t+1}] = \mathbf{F}^i z_t, \quad (18)$$

and in particular $\mathbf{E}_t[r_{t+i}] = \mathbf{h}_1' \mathbf{F}^i z_t$, with $\mathbf{h}_\ell = (0, \dots, 0, 1, 0, \dots, 0)'$ selecting the ℓ th element of a vector of length Kp . The theoretical long-term interest rate is given by:

$$\begin{aligned} R_t &= \left(\frac{1 - \gamma^k}{1 - \gamma} \right) \left(\sum_{i=0}^{k-1} \gamma^i \mathbf{E}_t[r_{t+mi}] \right) = \left(\frac{1 - \gamma^k}{1 - \gamma} \right) \mathbf{h}_1' \left(\sum_{i=0}^{k-1} \gamma^i \mathbf{F}^{mi} \right) z_t \\ &= \left(\frac{1 - \gamma^k}{1 - \gamma} \right) \mathbf{h}_1' (\mathbf{I} - \gamma \mathbf{F}^m)^{-1} (\mathbf{I} - \gamma \mathbf{F}^n) z_t. \end{aligned} \quad (19)$$

The j -period ahead forecast error of z_t is calculated as

$$z_{t+j} - \mathbf{E}_t[z_{t+j}] = \sum_{i=0}^{j-1} \mathbf{F}^i \mathbf{G} \eta_{t-i}, \quad (20)$$

which has the $(Kp \times Kp)$ covariance matrix:

$$\boldsymbol{\Omega}_j = \sum_{i=0}^{j-1} \mathbf{F}^i \mathbf{G} \boldsymbol{\Sigma} \mathbf{G}' (\mathbf{F}')^i. \quad (21)$$

Since the theoretical long rate is a linear function of z_t , its j -period ahead forecast error variance follows from Eq. (19) as:

$$\begin{aligned} \sigma_j^2 &= \left(\frac{1 - \gamma^k}{1 - \gamma} \right)^2 \mathbf{h}_1' (\mathbf{I} - \gamma \mathbf{F}^m)^{-1} (\mathbf{I} - \gamma \mathbf{F}^n) \boldsymbol{\Omega}_j (\mathbf{I} - \gamma \mathbf{F}^n)' \\ &\quad \times (\mathbf{I} - \gamma (\mathbf{F}')^m)^{-1} \mathbf{h}_1. \end{aligned} \quad (22)$$

The corresponding expression for the unrestricted long-term interest rate is simpler, since the long rate is one of the variables directly included in the VAR:

$$\tilde{\sigma}_j^2 = \mathbf{h}_2' \boldsymbol{\Omega}_j \mathbf{h}_2. \quad (23)$$

In systems that contain some unit roots, we can separate the different responses of interest rates due to permanent or transitory shocks. These shocks can be derived from a specific variance decomposition of the VAR. The derivation starts from the invertible MA representation of the time series x_t , given by:

$$x_t = \mathbf{H} \sum_{j=0}^{\infty} \mathbf{F}^j \mathbf{G} \eta_{t-j}, \quad (24)$$

where $\mathbf{H} = (\mathbf{I} \cdot \mathbf{O} \cdot \dots \cdot \mathbf{O})$, a $(K \times Kp)$ matrix. From Eq. (24), the MA representation of Δx_t follows as:

$$\Delta x_t = \mathbf{H}\mathbf{G}\eta_t - \mathbf{H} \sum_{j=1}^{\infty} (\mathbf{I} - \mathbf{F}) \mathbf{F}^{j-1} \mathbf{G}\eta_{t-j} \equiv \sum_{j=0}^{\infty} \mathbf{B}_j \eta_{t-j}. \quad (25)$$

The long-run impact matrix, defined as the sum of the MA parameter matrices, is the key to separating transitory and permanent effects:

$$\mathbf{B} = \sum_{j=0}^{\infty} \mathbf{B}_j = \mathbf{H} \left(\lim_{j \rightarrow \infty} \mathbf{F}^j \right) \mathbf{G}. \quad (26)$$

When K_2 eigenvalues of $\mathbf{F}$ are equal to 1, the limit will be nonzero. Because of Granger's representation theorem (see Engle and Granger, 1987), the rank of $\mathbf{B}$ is equal to K_2 . The error vector η_t in Eq. (25) does not have an identity covariance matrix, and all elements η_{it} of the innovations can in general have both permanent and transitory effects. We therefore need a transformation:

$$\eta_t = \mathbf{D}\epsilon_t, \quad (27)$$

with $\mathbf{D}\mathbf{D}' = \mathbf{\Sigma}$, so that $\mathbf{E}[\epsilon_t \epsilon_t'] = \mathbf{I}$. The first K_1 elements of ϵ_t are pure transitory, and the last K_2 elements are permanent. The transformation matrix $\mathbf{D}$ is partitioned $(\mathbf{D}_1 : \mathbf{D}_2)$ conformably with the shock vector $\epsilon_t = (\epsilon'_{1t} : \epsilon'_{2t})'$, i.e. $\mathbf{D}_1$ is a $(K \times K_1)$ matrix and $\mathbf{D}_2$ is $(K \times K_2)$. The covariance matrix $\mathbf{\Sigma}$ is decomposed into a part due to pure transitory shocks and a part due to persistent shocks

$$\mathbf{\Sigma} = \mathbf{D}_1 \mathbf{D}_1' + \mathbf{D}_2 \mathbf{D}_2' = \mathbf{\Sigma}_1 + \mathbf{\Sigma}_2. \quad (28)$$

By the definition of transitory shocks, the long-term impact of ϵ_{1t} on all components of x_t must be 0:

$$\mathbf{B}\mathbf{D}_1 = 0. \quad (29)$$

This condition can be translated back to the parameters of the VAR written in error correction form, since the persistence matrix is closely related to the cointegrating vectors β and the error correction parameters α of the system. With K_2 unit roots and K_1 cointegrating vectors, Granger's representation theorem implies that the VAR (16) can be expressed alternatively as (see also Johansen, 1991):

$$\Delta x_t = \alpha \beta' x_{t-1} + \sum_{i=1}^{p-1} \mathbf{A}_i^* \Delta x_{t-i} + \eta_t, \quad (30)$$

where α and β are both $(K \times K_1)$ matrices of full column rank. Again by Granger's representation theorem, the persistence matrix $\mathbf{B}$ satisfies:

$$\mathbf{B}\alpha = 0. \quad (31)$$

This property establishes that D_1 must be in the space spanned by the error correction parameters α . Therefore, if P is a $(K_1 \times K_1)$ symmetric positive definite matrix, then the variance due to the transitory shocks is

$$\Sigma_1 = \alpha P \alpha'. \quad (32)$$

Further, since $D' \Sigma^{-1} D = I$, we have in particular that $D_1' \Sigma^{-1} D_2 = 0$, and $\alpha' \Sigma^{-1} D_2 = 0$. Applying these results to Eq. (28) we get

$$(\alpha' \Sigma^{-1}) \Sigma (\Sigma^{-1} \alpha) = \alpha' \Sigma^{-1} \alpha = (\alpha' \Sigma^{-1} \alpha) P (\alpha' \Sigma^{-1} \alpha) + 0, \quad (33)$$

from which we obtain the solution $P = (\alpha' \Sigma^{-1} \alpha)^{-1}$. The decomposition can now be stated formally as:

Proposition 1. *Let Σ be a $(K \times K)$ symmetric positive definite matrix, and let α be a $(K \times K_1)$ matrix of full column rank $K_1 < K$. Then there exists a unique decomposition $\Sigma = \Sigma_1 + \Sigma_2$, such that:*

1. $\text{rank}(\Sigma_1) = K_1$, and $\text{rank}(\Sigma_2) = K - K_1$;
2. $\Sigma_1 = \alpha P \alpha'$, with P a $(K_1 \times K_1)$ symmetric positive definite matrix.

The decomposition is given by:

$$\begin{aligned} \Sigma_1 &= \alpha (\alpha' \Sigma^{-1} \alpha)^{-1} \alpha', \\ \Sigma_2 &= \Sigma - \alpha (\alpha' \Sigma^{-1} \alpha)^{-1} \alpha'. \end{aligned}$$

The proposition is a generalization of a result in Blanchard and Quah (1989), who considered the case $K = 2$ and $K_1 = 1$. This is the only case where not only Σ_1 and Σ_2 are unique, but also the transformation matrices D_1 and D_2 . In general, we cannot identify the individual shocks ϵ_{it} , but only the joint effect of the stationary and permanent components.

Appendix B. Monte Carlo integration

For statistical inference, we need the distribution of the restricted and unrestricted conditional variances. Since σ_j^2 and $\tilde{\sigma}_j^2$ are functions $f_j(\{A_i\}_{i=1}^p, \Sigma)$ of the VAR parameters, we can use direct Monte Carlo integration to obtain the distribution of the estimated $\hat{\sigma}_j^2$ and $\hat{\tilde{\sigma}}_j^2$. It will be assumed that the asymptotic distribution provides a good approximation to the actual covariance matrix of the parameters of the unrestricted VAR.¹³ The transformation from the asymptotic distribution of the VAR parameters to the sequence of predictive variances is

¹³ According to taste, one can either adhere to the classical interpretation of the Monte Carlo integration as described in the text, or favor a Bayesian interpretation. A Bayesian analysis with a flat prior leads to the same numerical quantities, though with a different interpretation.

performed exactly, taking into account the assumptions on stationarity or co-integration in the VAR.

Using the notation in Appendix A, we can write the unrestricted VAR for a sample of T observations compactly in matrix notation as:

$$\mathbf{X} = \mathbf{Z}\Phi + \mathbf{U}, \quad (34)$$

where $\Phi \equiv (\mathbf{A}'_1, \dots, \mathbf{A}'_p)$, $\mathbf{X} = (x_1, \dots, x_T)'$, and $\mathbf{U} = (\eta_1, \dots, \eta_T)'$ and $\mathbf{Z}$ are defined analogously. The OLS estimator of Φ is denoted $\hat{\Phi}$, and the covariance matrix of the residuals is estimated by $\hat{\Sigma} = \frac{1}{T - Kp} \hat{\mathbf{U}}' \hat{\mathbf{U}}$. The sample size T is assumed sufficiently large to allow the asymptotic approximation:

$$\begin{cases} \text{vec}(\hat{\Phi} | \Sigma) & \sim \text{Normal}(\text{vec}(\Phi), \Sigma \otimes (\mathbf{Z}'\mathbf{Z})^{-1}) \\ \hat{\Sigma} & \sim \text{Wishart}(\Sigma, T, K) \end{cases} \quad (35)$$

To compute standard errors of all functions f_j of the VAR parameters, a sequence $i = 1, \dots, N$ random drawings is generated from the distributions:

$$\begin{cases} \Sigma(i) & \sim \text{Wishart}(\hat{\Sigma}, T, K) \\ \text{vec}(\Phi(i)) & \sim \text{Normal}(\text{vec}(\hat{\Phi}), \Sigma(i) \otimes (\mathbf{Z}'\mathbf{Z})^{-1}) \end{cases} \quad (36)$$

For each $\Phi(i)$ we compute the roots of the VAR and check whether they are stable. If some drawing produces an explosive system, we discard it.

The Monte Carlo computation of the standard errors can be applied both under the assumption of stationarity and under a cointegrated system. Any cointegrated VAR can be transformed to an Error Correction Model (ECM). The Monte Carlo integration is applied conditional on the cointegrating vectors β , and taking the asymptotic covariance matrix of $\hat{\alpha}$ and $\{\hat{\mathbf{A}}^*_i\}_{i=1}^{p-1}$ from the OLS estimation of Eq. (30).

Monte Carlo integration has a number of advantages in the present application. In principle, we could use the standard asymptotic approximation $V(\hat{\sigma}_j^2, \tilde{\sigma}_j^2) = \nabla f'_j \mathbf{W} \nabla f_j$ to estimate the covariance matrix of the estimators $\hat{\sigma}_j^2$ and $\tilde{\sigma}_j^2$, with $\mathbf{W}$ the covariance matrix of the VAR parameters and ∇f the matrix of first order derivatives of σ_j^2 and $\tilde{\sigma}_j^2$ with respect to the VAR parameters. But the computation of the standard errors would be very cumbersome due to the nonlinearity of the functions f_j . This will be especially important for the high order VAR's that we estimate and for large j . Second, the Wald test of the restrictions of the expectations model is not numerically invariant with respect to the form in which we test the hypothesis.¹⁴ P -values computed by Monte Carlo integration do not have this problem.

¹⁴ See Campbell and Shiller (1987, fn. 9 and 27), where it appears that the expectations model is strongly rejected in one form (p -value less than 0.005%) and only marginally in an algebraically equivalent form (p -value 8.4%).

The third reason for preferring the Monte Carlo integration is that it might be better in dealing with autoregressive roots close to unity. To illustrate the point intuitively, assume, as in Flavin (1983), that the short rate is generated by the AR(1) model $r_t = \rho r_{t-1} + \epsilon_t$ with $\rho < 1$. In this case, long rates of all maturities are proportional to the short rate. For $n \rightarrow \infty$, the relation is given by $R_t = \alpha r_t$ with $\alpha = (1 - \delta)/(1 - \delta\rho)$. In empirical applications, $\hat{\rho}$ will be close to unity, and have a 5% confidence interval that is open to the right at $\rho = 1$.¹⁵ A confidence interval of α based on asymptotic theory will also include values of $\alpha > 1$, although these are theoretically ruled out. Further, the variance of the long rate is $\alpha^2 \sigma_\epsilon^2 / (1 - \rho^2)$. The variance of the long rate will be very sensitive to ρ if it is close to unity, causing the asymptotic standard error to be a poor approximation of the true uncertainty about the volatility of the long rate.

References

- Ang, A., Bekaert, G., 2001. Short rate nonlinearities and regime switches. *Journal of Economic Dynamics and Control*, in press.
- Black, F., Derman, E., Toy, W., 1990. A one factor model of interest rates and its application to treasury bond options. *Financial Analysts Journal* 46, 33–39.
- Blanchard, O.J., Quah, D., 1989. The dynamic effects of aggregate demand and supply disturbances. *American Economic Review* 79, 655–673.
- Campbell, J.Y., Shiller, R.J., 1987. Cointegration and tests of present value relations. *Journal of Political Economy* 95, 1062–1088.
- Campbell, J.Y., Shiller, R.J., 1991. Yield spreads and interest rate movements: a bird's eye view. *Review of Economics and Statistics* 58, 495–514.
- Christiano, L.J., Eichenbaum, M., 1990. Unit roots in real GNP: do we know and do we care? *Carnegie Rochester Conference Series on Public Policy* 32, 7–62.
- Cochrane, J.H., 1991. Volatility tests and efficient markets: a review essay. *Journal of Monetary Economics* 27, 463–485.
- Dai, Q., Singleton, K., 2000. Specification analysis of affine term structure models. *Journal of Finance* 55, 1943–1978.
- Engle, R.F., Granger, C.W.J., 1987. Co-integration and error correction: representation, estimation and testing. *Econometrica* 55, 251–276.
- Flavin, M.A., 1983. Excess volatility in the financial markets: a reassessment of the empirical evidence. *Journal of Political Economy* 91, 929–956.
- Fuller, W.A., 1976. *Introduction to Statistical Time Series*. Wiley, New York.
- Hall, A.D., Anderson, H.M., Granger, C.W.J., 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.
- Heath, D., Jarrow, R., Morton, A., 1992. Bond pricing and the term structure of interest rates: a new methodology for contingent claim valuation. *Econometrica* 60, 77–105.

¹⁵ Here we maintain the assumption of stationarity. See Sims (1988) for the peculiar forms of confidence intervals near the unit root.

- Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59, 1551–1580.
- Johansen, S., Juselius, K., 1990. Maximum likelihood estimation and inference on cointegration—with applications to the demand for money. *Oxford Bulletin of Economics and Statistics* 52, 169–210.
- Kleibergen, F., van Dijk, H.K., 1994. On the shape of the likelihood/posterior in cointegration models. *Econometric Theory* 10, 514–551.
- Kleidon, A.W., 1986. Variance bounds tests and stock price valuation models. *Journal of Political Economy* 94, 953–1001.
- LeRoy, S.F., 1984. Efficiency and the variability of asset prices. *American Economic Review* 74, 183–187.
- LeRoy, S.F., 1989. Efficient capital markets and martingales. *Journal of Economic Literature* 27, 1583–1621.
- Mankiw, N.G., 1986. The term structure of interest rates revisited. *Brookings Papers on Economic Activity* I, 61–109.
- Pfann, G., Schotman, P.C., Tschernig, R., 1996. Nonlinear interest rate dynamics and implications for the term structure. *Journal of Econometrics* 73, 149–176.
- Phillips, P.C.B., Perron, P., 1988. Testing for a unit root in time series regression. *Biometrika* 75, 335–346.
- Quah, D., 1990. Permanent and transitory movements in labor income: an explanation for “Excess Smoothness” in consumption. *Journal of Political Economy* 98, 449–475.
- Sargent, T.J., 1979. A note on maximum likelihood estimation of the rational expectations model of the term structure. *Journal of Monetary Economics* 5, 133–143.
- Schotman, P.C., van Dijk, H.K., 1991a. A Bayesian analysis of the unit root hypothesis in real exchange rates. *Journal of Econometrics* 49, 195–238.
- Schotman, P.C., van Dijk, H.K., 1991b. On Bayesian routes to unit roots. *Journal of Applied Econometrics* 6, 387–401.
- Shea, G.S., 1991. Uncertainty and implied variance bounds in long-memory models of the interest rate term structure. *Empirical Economics* 16, 287–312.
- 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.
- Shiller, R.J., 1981. Alternative tests of rational expectations models: the case of the term structure. *Journal of Econometrics* 16, 71–87.
- Sims, C.A., 1988. Bayesian Skepticism on unit root econometrics. *Journal of Economic Dynamics and Control* 12, 463–474.
- Stock, J.H., Watson, M.W., 1988. Testing for common trends. *Journal of the American Statistical Association* 83, 1097–1107.
- Watson, M.W., Engle, R.F., 1987. The Kalman Filter: applications to forecasting and rational expectations models. In: Bewley, T.F. (Ed.), *Advance in Econometrics*. Cambridge Univ.
- West, K.D., 1988. Dividend innovations and stock price volatility. *Econometrica* 56, 37–61.