Discussion Paper 44

Institute for Empirical Macroeconomics
Federal Reserve Bank of Minneapolis
250 Marquette Avenue
Minneapolis, Minnesota 55480

June 1991

WHEN UNIT ROOTS MATTER: EXCESS VOLATILITY AND EXCESS SMOOTHNESS OF LONG TERM INTEREST RATES

Peter Schotman*

Institute for Empirical Macroeconomics
Federal Reserve Bank of Minneapolis

and University of Limburg

ABSTRACT

This paper reexamines volatility tests of the expectations model of the term structure of interest rates. The restrictions of the model are studied in a general multivariate MA representation of the time series process of interest rates under various assumptions on the number of unit roots and the pattern of cointegration. Test statistics are computed by Bayesian techniques. We find that the long term interest rate overreacts to transitory shocks, and underreacts to permanent shocks.

*The research for this paper was undertaken while the author was a visitor of the Institute for Empirical Macroeconomics in Minneapolis (MN). An earlier version of this paper was presented at the 1990 World Congress of the Econometric Society, at a workshop of the IMA at the University of Minnesota, and at seminars at the Universities of Amsterdam, Leuven and Marseille (GREQE). I thank participants for their comments. All errors are of course my own.

This material is based on work supported by the National Science Foundation under Grant. No. SES-8722451. The Government has certain rights to this material.

1. INTRODUCTION.

The unit root issue has been of great concern in the econometrics literature the past few years. A lot of research has been devoted to determine whether (macro)economic time series are stationary in levels or stationary in first differences.[1,2] While the properties of Trend Stationary and Difference Stationary time series are vastly different, the empirical tests all have low power. Indeed, for a major U.S. macro economic time series as real per capita GNP, Christiano and Eichenbaum [1990] show that a trend stationary model can fit the data as well (and as parsimoniously) as a difference stationary model.[3] In line with work of Quah [1990] they further contend that the whole unit root issue might not be important anyway. Implications of dynamic economic models are not very sensitive to the presence of unit roots. In this paper we investigate this conjecture for the expectations model of the term structure of interest rates. The conclusion will be that this is an example of a substantive economic issue where the data are uninformative on the existence of a unit root, yet empirical results depend crucially on the decision on DS or TS. In particular the conclusion that long term interest rates are excessively volatile depends critically on the existence and number of unit roots in a VAR system with interest rates and other macroecomic variables.

Volatility and variance bounds tests have generated a voluminous literature by itself, starting with Shiller [1979] who observed that long

---

[1] For a selective overview see the special issue of the *Journal of Economic Dynamics and Control* [1988], and the special issue of the *Oxford Bulletin of Economics and Statistics* [1988]. Diebold and Nerlove [1989] provide a survey of the literature on testing for unit roots.

[2] 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.

[3] Cochrane [1990] and Blough [1990] further question the interpretation and usefulness of univariate unit root tests.

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.[4] 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. Kleidon [1986] shows in great detail why the early variance bounds tests are uninterpretable when the time series involved are non-stationary. Kleidon's analysis is formalized in Durlauf and Phillips [1988] who derive the asymptotic distributions of the variance bounds statistics in the unit root case. West [1988] proposes a variance inequality based on the variance of forecast errors, which can be tested with stationary as well as with integrated time series. Applied to stock market data West [1988] finds excess volatility under both assumptions. Campbell and Shiller [1987] explicitly impose cointegration between stock prices and dividends, or 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 [1987,1989]. 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 <u>V</u>ector <u>A</u>uto<u>R</u>egressions (VAR). Apart from a long and a short term interest rate our VAR contains three more macro economic variables: prices, real output and money. The purpose of extending the VAR is twofold. First, in a VAR with five variables the number of unit roots in the system can range from zero to five. Allowing for various patterns of cointegration between interest rates and macroeconomic variables provides

---

[4] See LeRoy [1984, 1989] for general surveys of the variance bounds literature.

greater flexibility in modelling the relative importance of permanent and transitory shocks over various forecast horizons. Identification of important macroeconomic news is the second advantage of using a five variable VAR. We can trace which shocks are responsible for the excess volatility. The presence of the additional variables also helps to investigate the sensitivity of implied annuity values of shocks.[5]

The paper presents several other modifications of the Campbell and Shiller [1987] model. Moving average (MA) time series representations of long term interest rates are derived for a very general data generating process of the short rate. The model of Campbell and Shiller arises as a special case when long and short term interest rates cointegrate, the long rate has infinite maturity, and the maturity of the short rate equals the time between subsequent observations of the short rate.[6] The general MA representations provide a tool to investigate the sensitivity of volatility tests with respect to assumptions on the order of integration and cointegration of interest rates and other variables.

Using Monte Carlo integration the empirical distribution of conditional variances can be computed as an exact transformation of the asymptotic distribution of the parameters of an unrestricted VAR. The distributions can also be interpreted as Bayesian posteriors obtained with a flat prior. The procedure is designed to overcome the statistical problems noted by Flavin [1983] when dealing with long horizon present value models and autoregressive

---

[5] In the context of the Permanent Income Hypothesis the annuity value is the amount that lifetime permanent income changes in reaction to a one dollar shock to income. They are another way to compare actual and implied volatility of consumption; see Deaton [1987]. The methods can also be applied to the term structure of interest rates.

[6] Campbell and Shiller [1989] also relax the assumptions on the maturities in their [1987] paper, but they use a slightly different form of the expectations model of the term structure and propose a different solution to the misalignment of the observation frequency and the shortest maturity that implicitly introduces complex "seasonal" unit roots.

roots close to unity.

The main objective of the paper is to investigate the sensitivity of empirical tests of excess volatility with respect to the presence of unit roots. It is not the intention to test all implications of the term structure. Several different test procedures have fairly well established that the expectations model is not strictly true.[7] The expectations model is clearly too simple, but it can still be a useful first approximation, and it helps in predicting interest rates.[8] Although time varying risk premia are frequently put forward, it has been found difficult to obtain an empirical model that performs uniformly better than the expectations model.

The paper is organized as follows. Section 2 describes the data and introduces the notation. Section 3 presents the basic facts relevant to volatility tests. Section 4 deals with the time series representation of long term interest rates and the assumptions underlying the volatility tests. Section 5 discusses the econometric techniques when the model is represented as a VAR. Section 6 contains empirical results. Section 7 concludes.

2. DATA AND NOTATION.

This section introduces the notation and describes the time series data on interest rates. 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 $R_t^{(1)}$ (see Shiller [1979]):

---

[7] See Shiller [1987] for a survey of the theory and tests of the term structure of interest rates.

[8] See Fama and Bliss [1987], or Mishkin [1988].

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

where $\phi^{(n)}$ represents a liquidity or risk premium that is assumed constant over time, and where $\delta = (1+\rho)^{-1}$ is the discount rate around which bond prices are linearized. The notation $E(\cdot|I_t)$ 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 $t+1$. Equation (1) expresses the long rate as a weighted average of the current and expected future short rate.

The interest rate data used in this paper consist of three time series of monthly interest rates for the United States, sampled on the last trading day of the month for the period January 1962 to June 1990.[9] The first series is the yield on a three month Treasury Bill. The maturity of the short rate (3 months) does not coincide with observation frequency (monthly). Although this creates some technical econometric problems, the 3-month series will be used in most of the empirical tests. The preferable 1-month rate was only available for the shorter period 1968 to 1990; this series will only be used in section 3 below. The long rate is the yield to maturity on 10 year government bonds. Figures 1A and 1B show the levels and first differences of the long rate and the 3-month rate. Two features of the data help in interpreting formal test results later. First, the levels of short and long rates have about the same sample variance over the full thirty year period. Yet the long rate is considerably smoother, since the standard deviation of its first difference is much smaller than the standard deviation of changes in the short rate.

The required modification of equation (1) is easily obtained by applying the expectations theory twice for bonds with maturities m and n, and assuming

---

[9] All data were kindly provided by the Federal Reserve Bank of Minneapolis.

that $k=n/m$ is an integer.

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

$$= \frac{1-\delta}{1-\delta^n}\sum_{h=0}^{k-1}\delta^{mh}\sum_{j=0}^{m-1}\delta^j E(R_{t+mh+j}^{(1)}|I_t) + \phi^{(n)}$$

$$= \sum_{h=0}^{k-1}\delta^{mh}E(R_{t+mh}^{(m)}|I_t) + \phi^{(n)} - \phi^{(m)}, \qquad (2)$$

linking an m-period rate to a longer n-period rate. Eq. (2) has the same structure as eq. (1). To simplify the 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 we use the shorthand $E_t(\cdot)$ for $E(\cdot|I_t)$ when there can be no confusion on the interpretation of the information set. In this notation eq. (2) simplifies to

$$R_t = \frac{1-\gamma}{1-\gamma^k}\sum_{i=0}^{k-1}\gamma^i E_t(r_{t+mi}) \qquad (3)$$

Since most of the discussion in this paper centers on the effects of imposing unit roots, we start by examining the results of a standard univariate unit root test. Results of the Phillips and Perron [1988] test are reported in table 1. The tests can not 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 co-integrated.[10]

Though a classical test can not reject the unit root, this by no means

---

[10] This is a standard result. See, for example, Stock and Watson [1988] or Campbell and Shiller [1987].

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.

3. STYLIZED VOLATILITY FACTS.

This section replicates two well-known volatility tests with our term structure data, and examines their sensitivity with respect to the assumption of unit roots. Subsection 3.1 considers Shiller's [1979] variance bound; subsection 3.2 deals with West's [1988] test.

3.1. Ex-post rational long rate

The most intuitive volatility test uses the concept of the ex-post rational long term interest rate R^*_t introduced by Shiller [1979], which is defined as

$$R^*_t = \frac{1-\gamma}{1-\gamma^k} \sum_{i=0}^{k-1} \gamma^i r_{t+mi}, \tag{4}$$

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 variables in agents' information set I_t. For any $H_t \subset I_t$ it therefore holds that

$$\mathrm{Var}(R^*_t | H_t) > \mathrm{Var}(R_t | H_t) \tag{5}$$

This variance inequality can easily be verified by constructing a time series of R^*_t. Since the time to maturity of the long rate is 10 years and the sample is long enough, the series can be constructed exactly without any further assumptions using (4). 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 (see Shea [1989b]). The drawback of the exact calculation is of course the loss of 10 years of observations at the end of the sample. Figure 2 shows the actual and ex-post rational long rate.

The smooth behavior of R^*_t and the sample unconditional variances of the two series given 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 = {constant}), we can remove the possible non-stationarity by conditioning on the past levels of the short and long rates.[11] 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.

3.2. *Variance bound test of West [1988]*

The variance bounds test of West [1988] is specifically designed to cope with nonstationarity. The test procedure consists of comparing the variance of two different expected present values: one using the market information set and the other using a univariate time series projection. The original version of the test is designed for stock prices, but it is also applicable to the term structure under the assumption that the maturity of the long term

---

[11] 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, p. 961-962]:"*confusion in interpretation of time-series plots of price and $p^*(t)$ stems from comparing the conditional variance of price, $var\{p(t)|p(t-k)\}$, with an inappropriate conditional variance of $p^*(t)$, $var\{p^*(t)|p^*(t-k)\}$*".

bond is infinite. We will apply the test to the one month Treasury Bill rate and the 10 year rate, pretending that 120 months is close to infinity.

Define the two present values x_{tI} and x_{tH} as

$$x_{tI} = E\left\{\sum_{i=0}^{\infty} \delta^i r_{t+i} | I_t\right\}, \tag{6}$$

$$x_{tH} = E\left\{\sum_{i=0}^{\infty} \delta^i r_{t+i} | H_t\right\}, \tag{7}$$

where $H_t \subset I_t$, and I_t denotes the information set used by agents in the market. West [1988] proves the inequality

$$\mathrm{Var}(x_{tH}|H_t) \geq \mathrm{Var}(x_{tI}|I_t) \tag{8}$$

The right hand side of the inequality will be estimated under the null and compared to an unrestricted estimate of $\mathrm{Var}(x_{tH}|H_t)$. Under the null hypothesis (1) with $n\to\infty$, we have that $x_{tI} = (1-\delta)^{-1}R_t$, and $E\{x_{t+1,I}|I_t\} = (x_{tI} - r_t)/\delta$, so that, as a slight modification of West [1988], the conditional variance can be estimated from the equation

$$R_t - r_t = \delta(R_{t+1} - r_t) + u_{t+1}, \tag{9}$$

where $u_{t+1} = -(1-\delta)(x_{t+1,I} - E\{x_{t+1,I}|I_t\})$. Under the null hypothesis the variables in (9) are stationary, even if individual interest rate time series are integrated. The parameters δ and σ_u^2 can be efficiently estimated by GMM with instruments dated time t or earlier. The right hand side of (8) is thus estimated by $\hat{\sigma}_I^2 = \hat{\sigma}_u^2/(1-\hat{\delta})^2$.

To obtain a high upper bound we take H_t = {constant, r_{t-1}}. The estimate of x_{tH} will differ between the stationary case and the integrated case. If the short rate is assumed stationary we estimate an AR(1), and obtain the right hand side of (8) as $\hat{\sigma}_H^2 = \hat{\sigma}_v^2/(1-\hat{\delta}\hat{\rho})^2$, with $\hat{\rho}$ the estimated AR(1) parameter, and $\hat{\sigma}_v^2$ the variance of the errors. Alternatively, with r_t integrated we set $\hat{\rho}$=1 and $\hat{\sigma}_v^2 = \mathrm{Var}(\Delta r_t)$. The long term interest rate is excessively volatile if the variance ratio $\hat{\sigma}_I^2/\hat{\sigma}_H^2 > 1$. Standard errors of the variance ratio are computed as in West [1988].

Table 3 reports empirical results. The discount factor δ is estimated imprecisely and implausibly low, as it implies an annual discount rate of $100(0.989^{-12}-1) = 14\%$.[12] The main result is that the outcome of excess volatility tests depends entirely on the stationarity assumption; under a unit root the bound is easily satisfied. Also, although the point estimates imply gross violation in the case of stationarity, the standard errors are very large and the null hypothesis (8) can not be rejected. These conclusions are not sensitive to the value of the discount factor δ; setting $\delta = 0.994$, corresponding to an annual discount rate of 7%, does not qualitatively affect the results.

4. ASSUMPTIONS AND TIME SERIES IMPLICATIONS OF THE THEORY.

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. Given a data generating process (DGP) of the short rate, we can explicitly calculate future expectations of the short rate. Misspecification of the DGP of the short rate gives also room, however, for

---

[12] Incidentally note that the low estimate of δ is contrary to the usual finding in empirical work. Mankiw [1986] and Shiller, Campbell and Schoenholz [1983], for instance, find $\hat{\rho} = 1/\hat{\delta} - 1 < 0$ for a large number of countries and maturities, implying $\hat{\delta} > 1$. The difference is due to the estimation technique employed in the West procedure. It can be shown that the GMM estimator with more instruments than just $S(t)$ will always produce lower estimates of δ than the OLS estimator used by Mankiw [1986] and others.

another explanation of rejection of the expectations model.

4.1. Time series representation.

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

$$\Delta r_t = \mathbf{c}(L)\varepsilon_t = \sum_{i=0}^{\infty} \mathbf{c}_i \varepsilon_{t-i}, \tag{10}$$

where L is the lag operator, $\mathbf{c}_i$ are $(1\times K)$ vectors of parameters, $\mathbf{c}(1)$ is bounded, and ε_t is a $(K\times1)$ serially uncorrelated vector of innovations with mean zero and identity covariance matrix. The contemporaneous covariances are modelled through $\mathbf{c}_0$. The presence of multiple shocks $(K>1)$ allows for a distinction between permanent and transitory shocks as in Quah [1990]. A shock ε_{it} (the i^{th} element of ε_t) is transitory if $\mathbf{c}_i(1) = 0$. If all shocks are transitory, the lag polynomial is divisible by the difference operator, implying overdifferencing of the original level time series. Although K sources of stochastic uncertainty are introduced, not all of these will be identifiable using interest rate data alone. Section 5 discusses the empirical identification of the shocks. Having multiple shocks also opens the possibility of Granger causality running from long to short as well as from short to long rates.

It will often 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 (3) as

$$S_t = \sum_{i=1}^{k-1} \frac{\gamma^i - \gamma^n}{1 - \gamma^n} E_t(\Delta_m r_{t+mi}), \tag{11}$$

where $\Delta_m r_t = \sum_{j=0}^{m-1} \Delta r_{t-j}$. Optimal forecasts of Δr_{t+h} can be obtained from (10) as

$$E_t(\Delta r_{t+h}) = \sum_{j=1}^{\infty} \mathbf{c}_j \varepsilon_{t+h-j} = \sum_{j=0}^{\infty} \mathbf{c}_{h+j} \varepsilon_{t-j} \tag{12}$$

Substituting (12) into (11) and rearranging one obtains the implied time series process for the spread as

$$S_t = \sum_{j=0}^{\infty}\left(\sum_{h=0}^{m-1}\sum_{i=1}^{k-1}\frac{\gamma^i - \gamma^k}{1-\gamma^k}\mathbf{c}_{mi+j-h}\right)\varepsilon_{t-j} \tag{13}$$

Noting that $\Delta R_t = \Delta S_t + \Delta r_t$, the long term interest rate is obtained by summing equations (13) and (10). Recollecting terms in ε_{t-j} gives

$$\begin{aligned}\Delta R_t &= \left(\mathbf{c}_0 + \sum_{i=1}^{k-1}\frac{\gamma^i - \gamma^k}{1-\gamma^k}\sum_{h=0}^{m-1}\mathbf{c}_{mi-h}\right)\varepsilon_t + \sum_{j=1}^{\infty}\left(\frac{1-\gamma}{1-\gamma^k}\sum_{i=0}^{k-1}\gamma^i\mathbf{c}_{mi+j}\right)\varepsilon_{t-j} \\ &= \sum_{j=0}^{\infty}\mathbf{g}_j\varepsilon_{t-j}\end{aligned} \tag{14}$$

The long-run impact of a shock to the system is defined as $\mathbf{c}(1)$ for the short rate process, and as $\mathbf{g}(1)$ for the n-period long term interest rate. Summing the coefficients in (14) establishes the important property

$$\mathbf{g}(1) = \mathbf{c}(1), \tag{15}$$

which implies that interest rates of all maturities cointegrate, whenever a single component of $\mathbf{c}(1)$ is non-zero. Otherwise all interest rates are stationary.

The coefficients of $\mathbf{g}(\mathrm{L})$ are finite sums of the $\mathbf{c}(\mathrm{L})$ coefficients. Representation (14) simplifies considerably if m=1 and/or $n\to\infty$. Working with $m\neq1$ is only a technical complication. The assumption of an infinite maturity, however, which is used in the term structure model of Campbell and Shiller [1987], while analytically convenient, might introduce severe dynamic misspecification. First, with a finite maturity n, the $\mathbf{g}_j$ coefficients depend only on the first j+n entries of the $\mathbf{c}(\mathrm{L})$ 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 $\mathbf{g}(L)$ would not be very sensitive with respect to unit roots. Second, if the actual maturity of the long bond is about 10 years, then a specification with $n\to\infty$ puts too much weight on expected short term interest rates in the distant future. Figure 3 shows the two different weighting schemes for a discount rate of 7.5%. The weights of the infinite maturity rate are almost twice those of the 10-year rate over the first 40 quarters, reflecting the correction factor $1-\gamma^k = 0.485$ in eq. (3) with k=40 and $\gamma=(1.075)^{-1/4}$. Assuming an infinite maturity when the actual maturity is "only" 10 years restricts the long term interest rate to behave more smoothly than necessary.

Representation (14) contains all conditions implied by the expectations model of the term structure. All cross equation conditions can in principle be tested by comparing the implied process in (14) with an unrestricted representation

$$\Delta R_t = \tilde{\mathbf{g}}(L)\varepsilon_t = \sum_{j=0}^{\infty}\tilde{\mathbf{g}}_j\varepsilon_{t-j} \tag{16}$$

A test of all the conditions implied by the expectations hypothesis entails that $\tilde{\mathbf{g}}(L) = \mathbf{g}(L)$. Before any implications can be tested we need to estimate the lag polynomials $\mathbf{c}(L)$ -- to obtain its implication $\mathbf{g}(L)$ -- and $\tilde{\mathbf{g}}(L)$. This will be discussed in section 5.

4.2. Volatility.

The volatility tests will be based on the conditional variance of the long rate, defined as

$$\sigma_j^2 = E_t\left[\left(R_{t+j} - E_t(R_{t+j})\right)^2\right] = \sum_{i=0}^{j-1}\psi_i\psi_i', \tag{17}$$

where $\psi_i = \sum_{h=0}^{i}\mathbf{g}_h$, and $\mathbf{g}_h$ is given by (14). 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 conditional variances defined in (17) match the conditional variances obtained from an unrestricted time series model of the long term interest rate like (16), say $\tilde{\sigma}_j^2 = \sum_{i=0}^{j-1}\tilde{\psi}_i\tilde{\psi}_i'$, with $\tilde{\psi}_i = \sum_{h=0}^{i}\tilde{\mathbf{g}}_h$.

Volatility tests based on comparing restricted and unrestricted estimates of (17) are not subject to some standard criticisms of variance bounds tests put forward in Kleidon [1986]. First, the analysis concentrates on conditional variances rather than the unconditional variance, which does not exist under the unit root hypothesis. Second, the finite maturity assumption avoids a terminal value problem.[13]

The conditional variances in (17) converge to the unconditional variance if $j\to\infty$. If the long rate is integrated, the unconditional variance is infinite and increases linearly in j for large j. Therefore, if we impose cointegration of the long and the short rate, which means $\tilde{\mathbf{g}}(1) = \mathbf{g}(1)$, the ratio $\tilde{\sigma}_j^2/\sigma_j^2$ will approach unity when $j\to\infty$, regardless of whether the other term structure restrictions hold. Under the maintained null hypothesis of cointegration, the variance ratio must be satisfied in the limit.

Volatility is but one aspect of the term structure. Since the scalar conditional variances σ_j^2 are a limited set of nonlinear functions of the original $\mathbf{g}(L)$ polynomial, the number of restrictions implied by (17) is less than the full set of restrictions implied by (16). The test can be more or less powerful than a test of all conditions depending on the actual way the data behave in deviation from the model.

Many "efficient market" tests are concerned with the unpredictability of excess holding period returns. To construct a time series of the return on

[13] See also sections 3.1 and 4.1 above.

holding an n-period bond for m periods over in excess of holding the short term m-period bond within the linearized framework of eq. (3) one needs data about the yield of m, n-m and n period bonds respectively.[14] A minor difference with the volatility tests is that tests based on (16) and (17) only use data from two different interest rates. Only if we pretend $n \to \infty$ can we compute excess returns from two interest rates.

In the rest of this discussion on the relation between different tests we consider the special case $m=1$ and $n \to \infty$. Excess returns can then be calculated directly as

$$Y_t = \frac{R_{t-1} - \delta R_t}{1 - \delta} - r_{t-1} = \frac{\delta c(\delta)}{1 - \delta}\varepsilon_t \tag{18}$$

The last equality sign relates the observable time series Y_t to shocks in the short term interest rate using the MA representations (10) and (14). The choice of variables z_t to include in the efficient market test regression $Y_t = z'_{t-1}\beta + u_t$ is arbitrary, as Y_t should be orthogonal to any information dated t-1 or earlier. A test of the cross-equation restrictions based on (16) can be rewritten as an efficient market test with z_{t-1} the set of lagged innovations $\{\varepsilon_{t-j}; j \geq 1\}$. An example is the VAR test of Campbell and Shiller [1987], which can be rewritten as a regression of Y_t on $\{\Delta r_{t-j}, S_{t-j}; j=1,\ldots,p\}$.[15]

Volatility tests can also be completely independent of the orthogonality tests. The variance of the excess return, and hence the innovation variance of the long term interest rate, is determined by the innovation variance and other time series properties of the short rate. The $(1\times K)$ vector $\delta c(\delta)/(1-\delta)$ in (22) describes the immediate transmission of the K different shocks in the short rate. An unrestricted estimate of the variance excess returns $(\tilde{\sigma}^2_Y)$ is

---

[14] See Shiller, Campbell and Schoenholz [1983], eq. (3).

[15] See Campbell and Shiller [1987, p. 1068 and footnote 9].

easily obtained, since Y_t is directly observable. This implies that the one period ahead conditional variance of the long rate is given by $\tilde{\sigma}_1^2 = \left(\frac{1-\delta}{\delta}\right)^2 \tilde{\sigma}_Y^2$. The restricted estimate follows from the assumed DGP of the short rate, $\sigma_1^2 = \mathbf{c}(\delta)\mathbf{c}(\delta)'$. This means that even if Y_t is unpredictable, *i.e.* even if all "orthogonality" conditions are satisfied, a volatility test can still lead to rejection of the model.

Assuming that both the agents in the market as well as the econometrician observe the past of the short and the long rate, they both share (under the null hypothesis) the same expectation of next period's long rate ($E_t(R_{t+1}) = R_t + \frac{1-\delta}{\delta}S_t$) and the observation of the innovation of the long rate ($\frac{1-\delta}{\delta}Y_t$), independent of the formulation of the vector polynomial $\mathbf{c}(L)$. But the theoretical innovation variance of the long rate,

$$\sigma_1^2 = \mathbf{c}(\delta)\mathbf{c}(\delta)', \tag{19}$$

is a function of $\mathbf{c}(L)$. Since $\mathbf{c}(L)$ can be a vector polynomial of unknown dimension, one would not expect that (19) would hold for any marginalization that the econometrician considers. This is the point that Quah [1990] makes against some excess smoothness results within the Permanent Income Hypothesis using univariate time series methods. Campbell and Shiller [1987] have shown, however, that for the term structure the marginalization does not affect the interpretation of the volatility tests. The only two conditions are that the information set of the econometrician and of the agents in the market both contain at least the past level of the two interest rates and that the information set of the econometrician is a subset of the information of the agents. These are the same conditions as were encountered in the variance bounds test of West [1988]; see section 3.2.

Before the volatility implications can be tested, we need to estimate the lag polynomials $\mathbf{g}(L)$, or $\mathbf{c}(L)$, and $\tilde{\mathbf{g}}(L)$. This is the subject of next section.

5. ECONOMETRIC CONSIDERATIONS.

The infinite MA representations were a general tool to derive all dynamic properties of interest rates of all maturities. An infinite MA model can, however, not be estimated, so some assumptions on the shape of the MA representation will be required before an econometric analysis can be undertaken.

5.1. Estimation.

The most convenient representation for estimation purposes is a VAR. The advantage of a VAR is that estimation is computationally straightforward, and also provides consistent and efficient estimates of the parameters. Imposing cointegrating vectors is also straightforward within a VAR framework.

The conditional variances of the long rate defined in eq. (17) can be computed as nonlinear functions of the coefficients of an MA representation of the p^{th} VAR

$$x_t = \sum_{i=1}^{p} \mathbf{A}_i x_{t-i} + \eta_t, \tag{20}$$

where x is a $(K\times 1)$ vector of observed time series including at least two interest rates with different maturities. For concreteness, let the first element in x_t be the short rate r_t and the second element in x_t the long rate R_t. The covariance matrix of η is denoted Σ. Let $\mathbf{F}$ be the companion matrix of the VAR, obtained by reformulating the VAR as a first order system:

$$z_t = \begin{pmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-p+1} \end{pmatrix} = \begin{pmatrix} A_1 & A_2 & \cdots & A_p \\ I & & & \\ & \ddots & & \\ & & I & \end{pmatrix} \begin{pmatrix} x_{t-1} \\ x_{t-2} \\ \vdots \\ x_{t-p} \end{pmatrix} + \begin{pmatrix} I \\ 0 \\ \vdots \\ 0 \end{pmatrix} \eta_t$$

$$= \mathbf{F} z_{t-1} + \mathbf{G}\eta_t \tag{21}$$

If the system contains unit roots, some of the eigenvalues of **F** will be equal to one. We assume that all other eigenvalues are strictly inside the unit circle. The MA representation of the time series x_t is given by

$$x_t = \mathbf{H} \sum_{j=0}^{\infty} \mathbf{F}^j \mathbf{G} \eta_{t-j}, \tag{22}$$

where $\mathbf{H} = (\mathbf{I}\ \mathbf{0}\ \ldots\ \mathbf{0})$, a $(K \times Kp)$ matrix. From (22) the MA representation of Δx_t follows as

$$\Delta x_t = \mathbf{HG}\eta_t - \mathbf{H} \sum_{j=1}^{\infty} (\mathbf{I} - \mathbf{F})\mathbf{F}^{j-1}\mathbf{G}\eta_{t-j} \equiv \sum_{j=0}^{\infty} \mathbf{B}_j \eta_{t-j} \tag{23}$$

The error vector η_t does not have an identity covariance matrix. Uncorrelated shocks are obtained from the variance decomposition

$$\eta_t = \mathbf{D}\varepsilon_t \tag{24}$$

with $E(\varepsilon_t \varepsilon_t') = \mathbf{I}$, and where **D** satisfies $\mathbf{DD}' = \Sigma$.[16] Let $\mathbf{C}_j = \mathbf{B}_j\mathbf{D}$. The first row of $\mathbf{C}_j$ contains the impulse responses of the short rate with respect to a $(K \times 1)$ vector of shocks ε_t, *i.e.* the first row contains the $\mathbf{c}_j$ defining the DGP of the short rate in eq. (10). From the $\{\mathbf{c}_j\}_{j=0}^{\infty}$ sequence we can compute the restricted MA polynomial $\mathbf{g}(L)$ of ΔR_t from eq. (14), and the restricted conditional variances of the long rate using eq. (17). The second row of $\mathbf{C}_j$

---

16 The choice of **D** is not unique. See Sims [1980], Bernanke [1986], and section *5.4.* below.

holds the unrestricted responses $\tilde{\mathbf{g}}_j$ of the long rate with respect to ε_t (see eq. (16)).

5.2. Monte Carlo Integration.

For statistical inference we need the distribution of this indirect estimator of the restricted and unrestricted conditional variances. Since σ_j^2 and $\tilde{\sigma}_j^2$ are functions $\mathbf{f}_j(\{\mathbf{A}_i\}_{i=1}^p,\Sigma)$ of the VAR parameters, we could in principle estimate the covariance matrix of the estimated $\hat{\sigma}_j^2$ and $\hat{\tilde{\sigma}}_j^2$ using the standard asymptotic approximation $\mathbf{V}(\hat{\sigma}_j^2,\hat{\tilde{\sigma}}_j^2) = \nabla\mathbf{f}_j'\mathbf{W}\nabla\mathbf{f}_j$ with $\mathbf{W}$ the covariance matrix of the VAR parameters and $\nabla\mathbf{f}$ the matrix of first order derivatives of $\hat{\sigma}_j^2$ and $\hat{\tilde{\sigma}}_j^2$ with respect to the VAR parameters. This procedure has two serious drawbacks, however. First, the computation of the standard errors will be very cumbersome due to the non-linearity of the functions $\mathbf{f}_j$. This will be especially important for the high order VAR that we will be estimating and for large j.[17] Second, the asymptotic approximation has been shown to be very poor for autoregressive time series models for interest rates.

This last point has been emphasized by Flavin [1983]. To illustrate the point intuitively, assume, as in Flavin [1983], that the short rate is generated by the AR(1) model $r_t = \vartheta r_{t-1} + \varepsilon_t$ with $|\vartheta|<1$. In this case long rates of all maturities are proportional to the short rate. For $n\to\infty$ the relation is given by $R_t = \alpha r_t$ with $\alpha = (1-\delta)/(1-\delta\vartheta)$. In empirical applications $\hat{\vartheta}$ will be close to unity, and have a 5% confidence interval that is open to the right at $\vartheta=1$.[18] A confidence interval of α based on asymptotic

[17] Lütkepohl [1989] shows how to construct the asymptotic covariance matrix of the impulse responses of a VAR. This is already a hard analytical exercise, but for the volatility tests we need nonlinear functions of at least the first 240 entries of the impulse responses (see section 6). The test would require inversion of a 1200×1200 matrix. It would also require the intractable first order derivatives $\nabla\mathbf{f}$.

[18] Here we maintain the assumption of stationarity. See Sims [1988] for the peculiar forms of confidence intervals near the unit root.

theory will also include values of $\alpha>1$, although these are theoretically ruled out. Further, the variance of the long rate is $\alpha^2\sigma_\varepsilon^2/(1-\vartheta^2)$. The variance of the long rate will be very sensitive to ϑ 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.

To overcome both problems with the asymptotics we will compute standard errors by Monte Carlo integration. It will be assumed that the asymptotic distribution provides a good approximation to the actual covariance matrix of the parameters of the unrestricted VAR.[19] The transformation from the asymptotic distribution of the VAR parameters to the sequence of MA parameters is performed exactly, taking into account the assumptions on stationarity or co-integration on the VAR.

For a sample of T observations the unrestricted VAR can be written compactly in matrix notation as

$$\mathbf{X} = \mathbf{Z}\mathbf{\Phi} + \mathbf{U}, \tag{25}$$

where $\mathbf{X} = (x_1,\dots,x_T)'$, $\mathbf{Z} = (z_0,\dots,z_{T-1})'$, $\mathbf{\Phi} \equiv (\mathbf{A}_1',\dots,\mathbf{A}_p')'$, and $\mathbf{U} = (\eta_1,\dots,\eta_T)'$. 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

$$\begin{cases} \mathrm{vec}(\hat{\Phi}) \sim \mathrm{Normal}(\mathrm{vec}(\Phi),\Sigma\otimes(\mathbf{Z}'\mathbf{Z})^{-1}) \\ \hat{\Sigma} \sim \mathrm{Wishart}(\Sigma,T,K) \end{cases} \tag{26}$$

To compute standard errors of all functions $\mathbf{f}_j$ of the VAR parameters a sequence of N random drawings is made from the distributions below.

---

[19] According to taste one can either adhere 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 results.

$$\begin{cases} \Sigma(i) \sim \text{Wishart}(\hat{\Sigma},T,K) \\ \text{vec}(\Phi(i)) \sim \text{Normal}(\text{vec}(\hat{\Phi}),\Sigma(i)\otimes(Z'Z)^{-1}) \end{cases} \tag{27}$$

For each $\Phi(i)$ we compute the roots of the VAR and check whether it is stable. Some drawings of the parameters must be discarded, since they produce an explosive system.

The Monte Carlo computation of standard errors can be applied both under the assumption of stationarity as well as for a co-integrated system. As a consequence of Granger's representation theorem (see Engle and Granger [1987]) any cointegrated VAR can be transformed to an Error Correction Model (ECM). Suppose there are M unit roots and $R = K-M$ cointegrating vectors contained in the $(K\times R)$ matrix β. Then the VAR (20) can be expressed alternatively as

$$\Delta x_t = \alpha\beta' x_{t-1} + \sum_{i=1}^{p-1} A_i^* \Delta x_{t-i} + \eta_t, \tag{28}$$

where α and β are both $(K\times R)$ matrices of full column rank. The Monte Carlo integration can be applied conditional on the cointegrating vectors β, and taking the asymptotic covariance matrix of $\hat{\alpha}$ and $\{\hat{A}_i\}_{i=1}^{p-1}$ from OLS estimation of eq. (28). In the empirical analysis the cointegrating vectors will mostly be specified *a priori* along with assumptions about the number of unit roots.

5.3. Further complications.

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. By choosing to work with a VAR we take the value of the discount factor $\delta = (1+\rho)^{-1}$ and the order of integration of interest rate time series as given, and ignore changes in policy regime that can possibly alter the dynamic structure of the model. Estimation of a constant ρ has been attempted by, *e.g.*, Mankiw [1986], while Engle and Watson [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* at $\delta = 0.994$. For monthly data this implies an annual discount rate of 7.5%.

The order of integration is probably the most important factor in interpreting volatility tests, as was shown in section 2 for the polar cases of I(0) and I(1). Shea [1989[a]] allows for fractionally integration (I(d) with $0 \leq d \leq 1$), and shows that quantifying the uncertainty about d can greatly reduce the significance of empirical violations of the variance bounds. Imposing the co-integration constraint $\mathbf{c}(1) = \mathbf{g}(1)$, however, poses severe technical problems, since estimation of fractionally co-integrated systems is not well developed yet. As a practical solution we will 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 won't 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 non-linear 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 "peso problem" points at another potential pitfall of standard regression tests, when there are infrequent regime shifts.[20] 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. (3), 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 sample a VAR will capture the second moments properties of the true DGP.

5.4. Identification of shocks.

In systems that contain some unit roots we can discriminate the different responses of interest rates due to permanent or transitory shocks. These shocks can be derived as a specific variance decomposition of the VAR, *i.e.* we will put structure on the transformation matrix **D** in (24). The distinction between permanent and transitory shocks follows from the properties of the long-run impact matrix, defined as the sum of the MA parameter matrices

$$\mathbf{B} = \sum_{j=0}^{\infty} \mathbf{B}_j = \mathbf{H}(\lim_{j\to\infty} \mathbf{F}^j)\mathbf{G} \tag{29}$$

---

[20] See Lewis [1991] for an application of the "peso problem" to the term structure.

By Granger's representation theorem the rank of $\mathbf{B}$ is equal to M, the number of unit roots in the system. The typical element $\mathbf{b}_{ij}$ of $\mathbf{B}$ measures the long-run effect of an innovation η_{jt} on output component x_{it}. In general, all innovations have permanent as well as transitory effects. The transformed, mutually uncorrelated shocks ε_{it} are chosen such that some of them only have transitory effects on all K variables in the VAR. The remaining elements of ε_t have persistent effects on at least some of the elements of the output series x_t. By definition a shock ε_{jt} is transitory if it does not affect any output x_{it} in the long run, i.e.

$$\mathbf{BD}_{.j} = \mathbf{0}, \tag{30}$$

where $\mathbf{D}_{.j}$ is the j^{th} column of $\mathbf{D}$. A shock is persistent if it is not transitory. If the system contains M unit roots, it is always possible to decompose η into M persistent shocks and $R = K\text{-}M$ transitory shocks. Partition $\varepsilon_t = (\varepsilon_{1t}' \ \varepsilon_{2t}')'$, where ε_{1t} contains the first R elements of ε and represents the transitory shocks. The transformation matrix $\mathbf{D}$ is partitioned accordingly as $\mathbf{D} = (\mathbf{D}_1 \ \ \mathbf{D}_2)$. The covariance matrix Σ is decomposed into a part due to purely transitory shocks and a part due to persistent shocks:

$$\Sigma \ = \ \mathbf{D}_1\mathbf{D}_1' + \mathbf{D}_2\mathbf{D}_2' \ = \ \Sigma_1 + \Sigma_2, \tag{31}$$

and where the definition of ε_1 as the transitory shocks implies that $\mathbf{D}_1$ must be chosen subject to

$$\mathbf{BD}_1 = \mathbf{0} \tag{32}$$

It is clear that condition (31) does not fully identify $\mathbf{D}$, since it is always possible to find orthogonal matrices $\mathbf{T}_1$ and $\mathbf{T}_2$ such that $\tilde{\mathbf{D}}_1 = \mathbf{D}_1\mathbf{T}_1$ and $\tilde{\mathbf{D}}_2 = \mathbf{D}_2\mathbf{T}_2$. The matrices Σ_1 and Σ_2 are exactly identified though, as will be shown below in the construction of the decomposition.

The decomposition can be constructed directly from 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. Using Granger's representation theorem the persistence matrix $\mathbf{B}$ has the following two properties:

$$\begin{aligned} &\text{(i)} \quad \mathbf{B}\alpha = \mathbf{0} \\ &\text{(ii)} \quad \beta'\mathbf{B} = \mathbf{0}, \end{aligned} \tag{33}$$

Property (i) establishes that $\mathbf{D}_1$ must be in the space spanned by the error correction parameters α, which have been estimated by OLS from the linear representation of the VAR in eq. (28). Let Γ be a $(R\times R)$ nonsingular matrix, then

$$\mathbf{D}_1 = \alpha\Gamma, \tag{34}$$

and the decomposition problem can be stated as finding a Γ and $\mathbf{D}_2$ such that

$$\Sigma = \alpha\Gamma\Gamma'\alpha' + \mathbf{D}_2\mathbf{D}_2' \tag{35}$$

Take $\tilde{\mathbf{D}}$ to be any Choleski decomposition of Σ. Then the decomposition is simply a Seemingly Unrelated Regression problem of the form $Y = XB + U$, with $Y=\tilde{\mathbf{D}}_1$, $X=\alpha$, $B=\Gamma$, $E(UU')=\Sigma$. Letting $\Gamma = (\alpha'\Sigma^{-1}\alpha)^{-1}\alpha'\Sigma^{-1}\tilde{\mathbf{D}}_1$, it follows that

$$\mathbf{D}_1 = \alpha\Gamma = \alpha(\alpha'\Sigma^{-1}\alpha)^{-1}\alpha'\Sigma^{-1}\tilde{\mathbf{D}} \tag{36}$$

$$\Sigma_1 = \alpha\Gamma\Gamma'\alpha' = \alpha(\alpha'\Sigma^{-1}\alpha)^{-1}\alpha', \tag{37}$$

and that $\mathbf{D}_2$ and Σ_2 can be obtained from the "residuals":

$$\mathbf{D}_2 = \left(\mathbf{I} - \alpha(\alpha'\Sigma^{-1}\alpha)^{-1}\alpha'\Sigma^{-1}\right)\tilde{\mathbf{D}} \tag{38}$$

$$\Sigma_2 = \Sigma - \alpha(\alpha'\Sigma^{-1}\alpha)^{-1}\alpha' \tag{39}$$

Neither Γ nor $\mathbf{D}_2$ is unique.[21] However, the matrices Σ_1 and Σ_2 are exactly identified, since they do not depend on the particular choice of $\tilde{\mathbf{D}}$.

With the above decomposition of Σ, the variance of the j-step ahead predictions of the long term interest rates is decomposed into a transitory part $\sigma^2_{\mathrm{T}j}$ and a permanent part $\sigma^2_{\mathrm{P}j}$ as

$$\begin{aligned} \sigma^2_{\mathrm{T}j} &= \sum_{i=0}^{j-1} \psi_{\mathrm{T}i}\psi'_{\mathrm{T}i} \\ \sigma^2_{\mathrm{P}j} &= \sum_{i=0}^{j-1} \psi_{\mathrm{P}i}\psi'_{\mathrm{P}i}, \end{aligned} \tag{40}$$

where $\psi_i = (\psi_{\mathrm{T}i}\ \psi_{\mathrm{P}i})$ is partitioned analogously to $\mathbf{D}$. For example, the unrestricted transitory conditional variance of the long rate uses the elements $\psi_{\mathrm{T}i} = \sum_{h=0}^{i}\mathbf{g}_{\mathrm{T}h}$, with $\mathbf{g}_{\mathrm{T}h}$ equal to the second row of $\mathbf{B}_h\mathbf{D}_1$. The other conditional variances ("restricted *permanent*", "*restricted transitory*", and "unrestricted permanent") are defined similarly.

6. EMPIRICAL RESULTS.

The interest rate data have been introduced in section 2. In addition three macro economic variables are included in the VAR: Industrial Production (y), Money stock M1 (M), and the Consumer Price Index (p). All data are seasonally adjusted monthly series taken from the Citibase tape. Series enter the VAR in

---

[21] The matrix $\mathbf{D}$ is fully identified only if K=2 and M=1, the case studied in Blanchard and Quah [1989].

logarithms and after detrending. The macroeconomic series are detrended by regressing on a constant and time trend if a series is assumed trend stationary, and by regressing first differences on a constant in case of I(1) series. The two interest rates are in deviation from the sample mean. No constant term or trend is included in the VAR. All differences in the empirical results are entirely due to assumptions about the dynamic specification.

Five different VAR's are compared:

(A) A 24^{th} bivariate VAR in levels, containing only the two interest rates. Interest rates are I(0) in this model.

(B) A 12^{th} order VAR with all five variables; all series are assumed stationary.

(C) A 24^{th} order bivariate VAR with a single unit root. Both interest rates are I(1) but cointegrate with cointegrating vector $\beta = (1 \ \ -1)'$.

(D) A 12^{th} 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 $\beta = (1 \ \ -1)'$.

(E) A 12^{th} order VAR with all five variables and 3 unit roots. All individual series are I(1), and there is one additional cointegrating relation, linking the short rate to velocity: $y_t + p_t - M_t - 0.066r_t$ is assumed to be a stationary series.[22]

Comparing the two bivariate models (A) and (C) gives insight in the

---

[22] 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 in a bivariate cointegrating regression test (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]). Within the five variable VAR Johansen's unrestricted ML estimator produces rather different cointegrating vectors. It is beyond the scope of this paper to provide a detailed analysis of the cointegration relations among the major U.S. macroeconomic variables.

importance of the assumption that interest rates cointegrate. 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. The most important difference between model (E) and the other VAR's is the cointegrating vector which relates the long-run properties of the interest rates to macro economic variables. Table 4 gives summary statistics of the different VAR's, and suggests that the unit root assumptions do not greatly affect the fit of the model. The five variable VAR's provide some improvement over the bivariate VAR for the equation of the short rate.[23] The estimates roots of the VAR also that the unit root restrictions are empirically plausible. The largest roots of the stationary VAR are very close to unity. In a univariate model with unknown intercept and trend a unit root can be rejected at the 10% level if the estimated root is smaller than 0.947.[24]

Figure 3 gives a first impression of the volatility implications of the different VAR's. The three broken 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,

---

[23] Formal Granger causality tests (not reported in the tables) have been performed to 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 matrix in the following matrix (A + 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.

[24] See Fuller [1976, table 8.5.1]), $\hat{\rho}$=1-18.1/341 (T=341 is the sample size for the VAR).

3 and 4 unit roots, respectively. The ranking is clear: the more unit roots in the system, the higher the implied standard deviation of the long rate. The solid line in the figure shows the unrestricted estimates $\tilde{\sigma}_j$ implied by a stationary VAR. 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, but they will of course slowly diverge to infinity if the forecast horizon j increases. Figure 3 illustrates the sensitivity of the excess volatility results with respect to the presence of unit roots.

In the figure the differences between the alternative models appear small for the one period ahead innovation variances, the starting point of the four lines. To investigate the statistical significance of the deviations between actual and implied volatility we used the Monte Carlo Integration; results are in table 5. Part I of the table shows that the actual volatility $\tilde{\sigma}_1$ -- 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, indicating that all the VAR's fit the long rate about equally well. The implied volatilities σ_1 are estimated less precisely and differ substantially across the models. The third column in table 5 presents the estimated probabilities of excess volatility for each of the models. The estimates confirm the pattern of figure 3. There is (significant) excess volatility if we believe in a stationary model. Imposing a common stochastic trend in the two interest rates, a seemingly innocuous restriction, leads to completely opposite results. The variance inequality is reversed, with the probability of excess volatility falling to 0.15 and 0.04 respectively.[25] The model with 3 unit roots (and the cointegration between the

[25] 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 non-stationarity 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 macroeconomic series. Also, a model with stationary interest rates, but

short rate and velocity) takes a middle position just as in figure 3.

The last three columns of the table focus on the endpoints in the curves of figure 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 (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 4.2. 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. Parts II and III of table 5 show that this is not true. Using the decomposition in transitory and persistent shocks described in section 5.4 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. For model E the excess volatility with respect to the transitory shocks happens to cancel to excess smoothness due to the permanent shocks for the innovations variance ratio. Over longer horizons the permanent components will eventually dominate all series, and any evidence against excess volatility or excess smoothness will disappear in the limit. The two transitory components are 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 the DGP of the short rate either

---

integrated macro economic variables is similar to the stationary VAR.

explains all variance (the stationary models A and B), is almost absent (C and D), or somewhere in between (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 the long rate. This is evident in the last line of table 4, which tells that the long rate must react less to the transitory shocks, and relatively more to the persistent shocks. This is another way of stating the conclusion that followed from table 5.

Three general explanations are possible of the volatility puzzle. First, the expectations model fails; second expectations of economic agents are irrational or the bond market is inefficient; third, a linear VAR is seriously misrepresents the DGP of the short rate. While the first two options can not be excluded, I am more inclined to misspecification of the different VAR's. Table 6 provide some diagnostic statistics. The first line shows that the errors are conditionally heteroskedastic. Correcting for ARCH will lead to time dependent volatility statistics and maybe to less significant variance ratio statistics. But unless the correlation structure of the error covariance matrix varies a lot over time ARCH will not affect the general conclusions. The significant skewness of the short rate is a first indication of possible nonlinear effect. The other two diagnostics in the table provide further evidence of nonlinearity, which seems to be present despite the overparameterization of the linear effects.[26] 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. In the continuous time literature the interest rate process of Cox, Ingersoll and Ross [1985], and Marsh and

---

[26] The residual diagnostics are similar for all VAR's. Only the test for nonlinear error correction is special for this model.

Rosenfeld [1983] introduce nonlinear mean reversion. Extending these models to the multivariate case seems a worthwile alternative to the linear cointegrated VAR.

7. CONCLUSION.

One main conclusion emerges from the empirical results. Whatever the form of a VAR that is fitted to interest rate data, the long rate will overreact to pure transitory shocks and underreact 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 for time series of U.S. interest rates.

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 small sample effects that might invalidate 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 zero to 98%, results are not sensitive to the assumed "size of the random walk".

If the sensitivity with respect to unit roots would extend to other data sets, the value of a VAR as a tool to test models of forward looking behavior would be much reduced. The diagnostics tests of the linear VAR's analyzed in this paper reveal important non-linearities.

TABLE 1: UNIVARIATE UNIT ROOT TESTS

Series	Sample	p = 0	p = 12
1 month	68:1 - 90:5	-2.89*	-2.75
3 months	62:1 - 90:5	-2.33	-2.28
10 years	62:1 - 90:5	-1.64	-1.71
spread (10 yr./ 1 m.)	68:1 - 90:5	-4.17*	-4.14*
spread (10 yr./ 3 m.)	62:1 - 90:5	-3.60*	-3.59*
spread (3 m. / 1 m.)	68:1 - 90:5	-11.80*	-13.47*

Test statistic is the adjusted t-statistic as proposed in Phillips and Perron [1988]; p is the number of additional lags. Critical value at the 1%, 5% and 10% level are -3.46, -2.88 and -2.57 respectively (see Fuller [1976, table 8.5.2]).

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^* constructed using actual short rates as in equation (4). The conditional variance is the residual variance of the regression of $x_t=\alpha+\beta_1 r_{t-1}+\beta_2 R_{t-1}+e_t$, where x is R and R^* respectively.
Sample period: 62:1 to 80:9.

TABLE 3: WEST'S [1988] VARIANCE BOUNDS TEST

GMM estimation (standard error in parenthesis):

$$R_t - r_t = \underset{(0.017)}{0.989} (R_{t+1} - r_t) + u_{t+1} \qquad \sigma_u^2 = 0.181$$

d	δ	$\hat{\sigma}_I^2/\hat{\sigma}_H^2$	s.e.
0	0.989	9.78	10.67
1	0.989	0.26	0.96
0	0.994	23.61	19.82
1	0.994	0.26	0.60

NOTES: d is order of integration of the short term interest rate; δ is the monthly discount factor; $\hat{\sigma}_I^2/\hat{\sigma}_H^2$ is the variance ratio described in the text; s.e. is an estimate of its standard error, computed as in West [1988].

TABLE 4: SUMMARY STATISTICS OF VAR'S

(I) Measures of fit and 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
F: 3 unit roots, (stationary interest rates)				1	1	1	0.980	0.980

(II) 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

NOTES: Specification of different VAR models (A to E) is given in the text. Model F has unit roots in the macro economic variables, but stationary interest rates. σ_r is the innovation standard error of the short rate; σ_R the innovation standard error of the long rate; ρ_{rR} is the correlation between the innovations; λ_i (i=1,...,5) are the largest roots of the system. The variance decomposition is described in section 5.4 in the text.

TABLE 5: VOLATILITY OF LONG TERM INTEREST RATES

	$E(\tilde{\sigma}_1)$	$E(\sigma_1)$	$\Pr\left(\frac{\tilde{\sigma}_1}{\sigma_1}>1\right)$	$E(\tilde{\sigma}_{120})$	$E(\sigma_{120})$	$\Pr\left(\frac{\tilde{\sigma}_{120}}{\sigma_{120}}>1\right)$
(I) Total variance						
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)*	0.540 *(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 *(0.125)*	0.500 *(0.013)*	4.077 *(1.353)*	3.820 *(1.957)*	0.773 *(0.011)*
(II) Variance due to transitory components						
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)*
(III) Variance due to permanent components						
A: Bivariate I(0)	0	0	---	0	0	---
B: Stationary VAR	0	0	---	0	0	---
C: Bivariate I(1)	0.249 *(0.058)*	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)*	0.994 *(0.002)*	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)*

NOTES: $E(\tilde{\sigma}_j)$ and $E(\sigma_j)$ are 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; numerical accuracy of the estimates is in parentheses. All entries are based on 1500 Monte Carlo replications.

TABLE 6: 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*

NOTES: LMARCH is a $\chi^2(4)$ LM test for 4^{th} order ARCH. Nonlinear ECM is a $\chi^2(4)$ test for nonlinear error correction by adding the variables S^2_{t-1}, S^3_{t-1}, z^2_{t-1}, z^3_{t-1}, 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 = $\sum \hat{u}^3_t/T\hat{\sigma}^3$, Kurtosis = $\sum \hat{u}^4_t/T\hat{\sigma}^4 - 3$. An asterisk (*) denotes significance at the 5% level.

REFERENCES.

Bernanke, B., [1986], Alternative explanations of the money-income correlation, *Carnegie Rochester Conference Series on Public Policy*, 25, 49-100.

Blanchard, O.J, and D. Quah [1989], The dynamic effects of aggregate demand and supply disturbances, *American Economic Review*, 79, 655-673.

Blough, S.R., [1990], Unit roots, stationarity, and persistence in finite smaple macroeconometrics, working paper #241, Johns Hopkins University.

Campbell, J.Y., and R.J. Shiller [1987], Cointegration and tests of present value models, *Journal of Political Economy*, 95, 1062-1088.

Campbell, J.Y., and R.J. Shiller [1989], Yield spreads and interest rate movements: a bird's eye view, unpublished.

Christiano, L.J., and M. Eichenbaum [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], A critique of the application of unit root tests, *Journal of Economic Dynamics and Control*, 15, 275-284.

Cox, J.C., J.E. Ingersoll and S.A. Ross [1985], A theory of the term structure of interest rates, *Econometrica*, 53, 145-166.

Deaton, A., [1987], Life-cycle models of consumption: Is the evidence consistent with the theory?, chapter 14 in: T.F. Bewley (ed.), *Advances in Econometrics*, Cambridge University Press (Cambridge).

Durlauf, S.N., and P.C.B. Phillips [1988], Trends versus random walks in time series analysis, *Econometrica*, 56, 1333-1354.

Engle, R.F., and C.W.J. Granger [1987], Co-integration and error correction: representation, estimation, and testing, *Econometrica*, 55, 251-276.

Engle, R.F., and M.W. Watson [1987], The Kalman filter: applications to forecasting and rational-expectations models, chapter 7 in: T.F. Bewley (ed.), *Advances in Econometrics*, Cambridge University Press (Cambridge).

Fama, E.F., and R.R. Bliss [1987], The information in long-maturity forward rates, *American Economic Review*, 77, 680-692.

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 Statistica Time Series*, Wiley (New York).

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.

Johansen, S., and K. Juselius [1990], Maximum likelihood estimation and inference on cointegration -- with applications to the demand for money, *Oxford Bulletin of Economics and Statistics*, 52, 169-210.

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*, XXVII, 1583-1621.

Lewis, K.K., [1991], Was there a "peso problem" in the U.S. term structure of interest rates: 1979-1982?, *International Economic Review*, 32, 159-173.

Lütkepohl, H., [1989], A note on the asymptotic distibution of impulse response functions of estimated VAR models with orthogonal residuals, *Journal of Econometrics*, 42, 371-376.

Mankiw, N.G., [1986], The term structure of interest rates revisited, *Brookings Papers on Economic Activity*, 1986:1, 61-109 (with discussions by S.M. Goldfeld and R.J. Shiller).

Marsh, T.A., and E.R. Rosenfeld [1983], Stochastic processes for interest rates and equilibrium bond prices, *Journal of Finance*, 38, 635-646.

Mishkin, F.S., [1988], The information in the term structure: some further results, *Journal of Apllied Econometrics*, 3, 307-314.

Phillips, P.C.B., and P. Perron [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.

Shea, G.S., [1989[a]], Uncertainty and implied variance bounds in long-memory models of the interest rate term structure, unpublished, Pennsylvania State University.

Shea, G.S., [1989[b]], Ex-post rational price approximations and the empirical reliability of the present value relation, *Journal of Applied Econometrics*, 4, 139-159.

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.

Shiller, R.J., [1987], The term structure of interest rates, unpublished, Cowles Foundation, Yale University.

Shiller, R.J., J.Y. Campbell and K.L. Schoenholz [1983], Forward rates and future policy: interpreting the term structure of interest rates, *Brookings Papers on Economic Activity*, 1983:I, 173-223.

Sims, C.A., [1980], Macroeconomics and reality, *Econometrica*, 48, 1-48.

Sims, C.A., [1988], Bayesian skepticism on unit root econometrics, *Journal of Economic Dynamics and Control*, 12, 463-474.

Stock, J.H., and M.W. Watson, [1988], Testing for common trends, *Journal of the American Statistical Association*, 83, 1097-1107.

West, K.D., [1988], Dividend innovations and stock price volatility, *Econometrica*, 56, 37-61.

FIGURE 1A: Interest rates (levels)

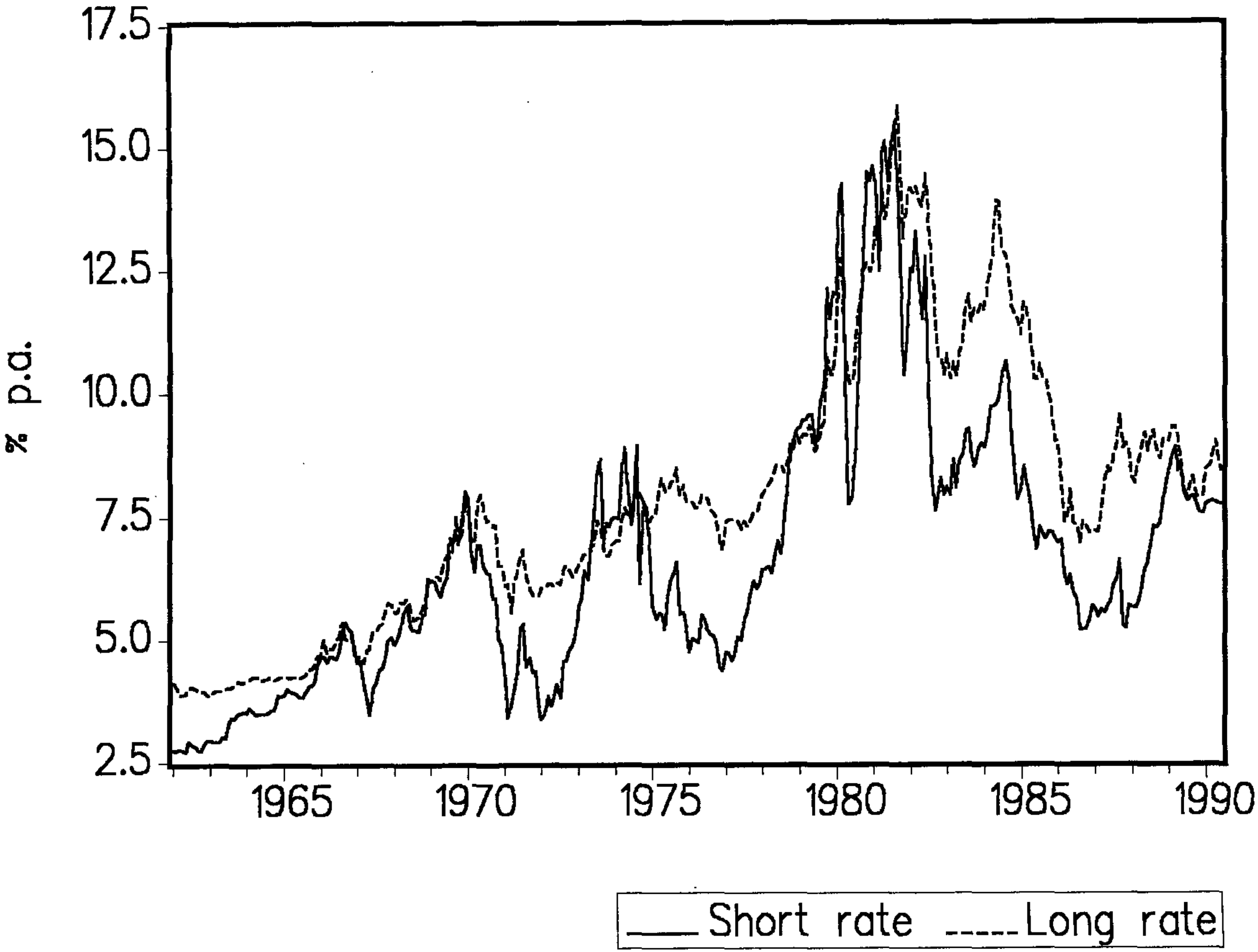

FIGURE 1B: Interest rates (differences)

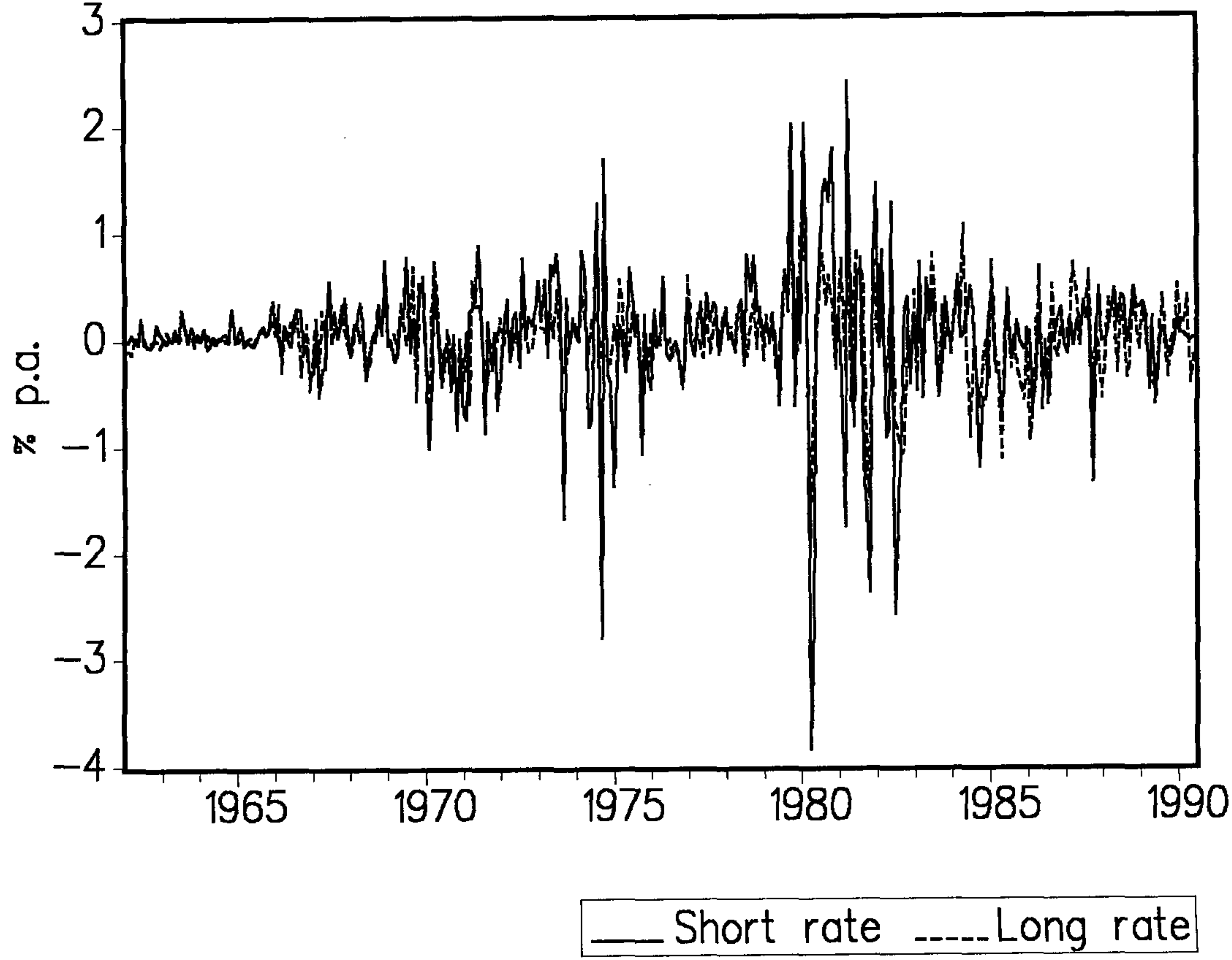

FIGURE 2: Long term interest rate

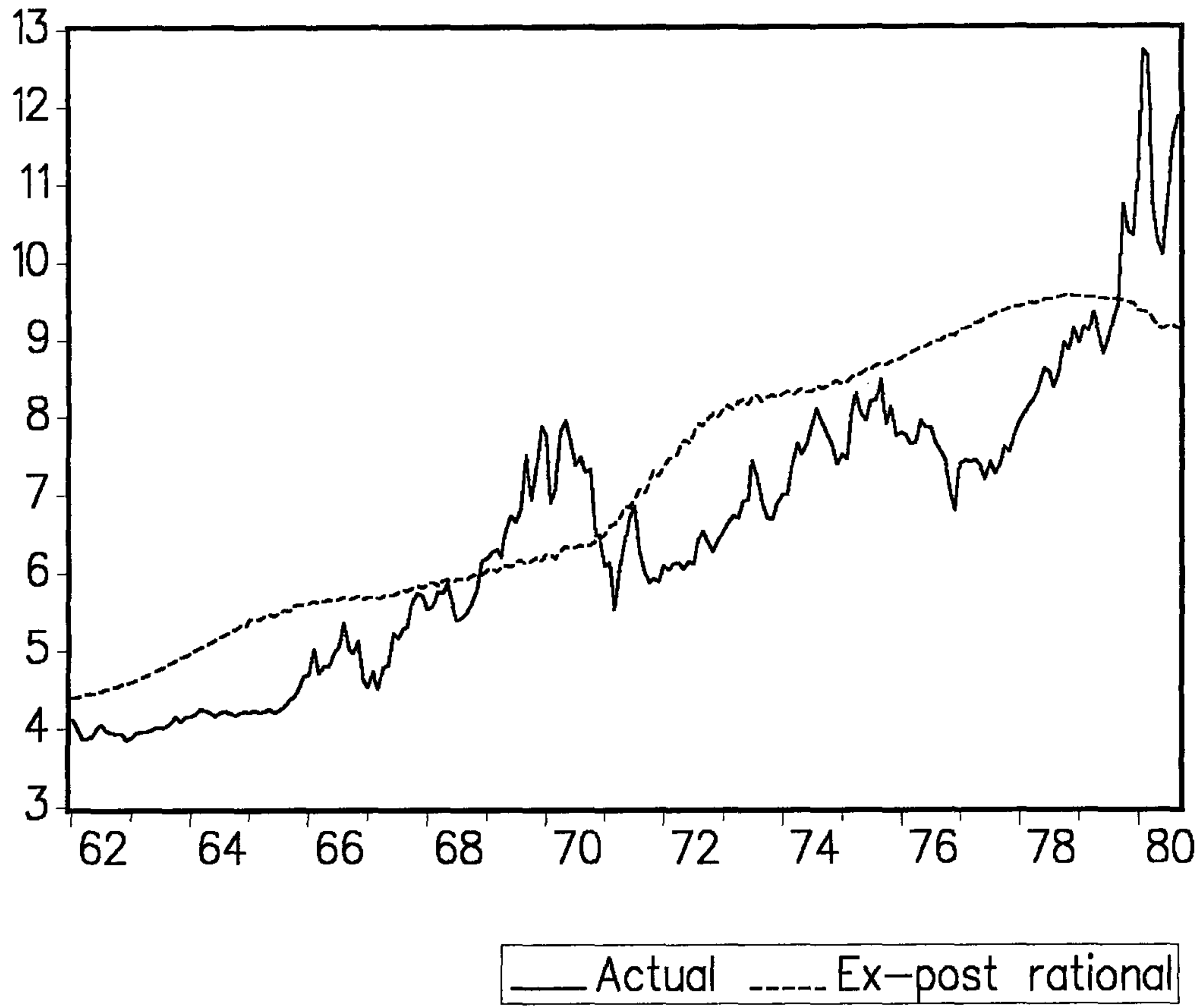

FIGURE 3: Weighting scheme for future expectations

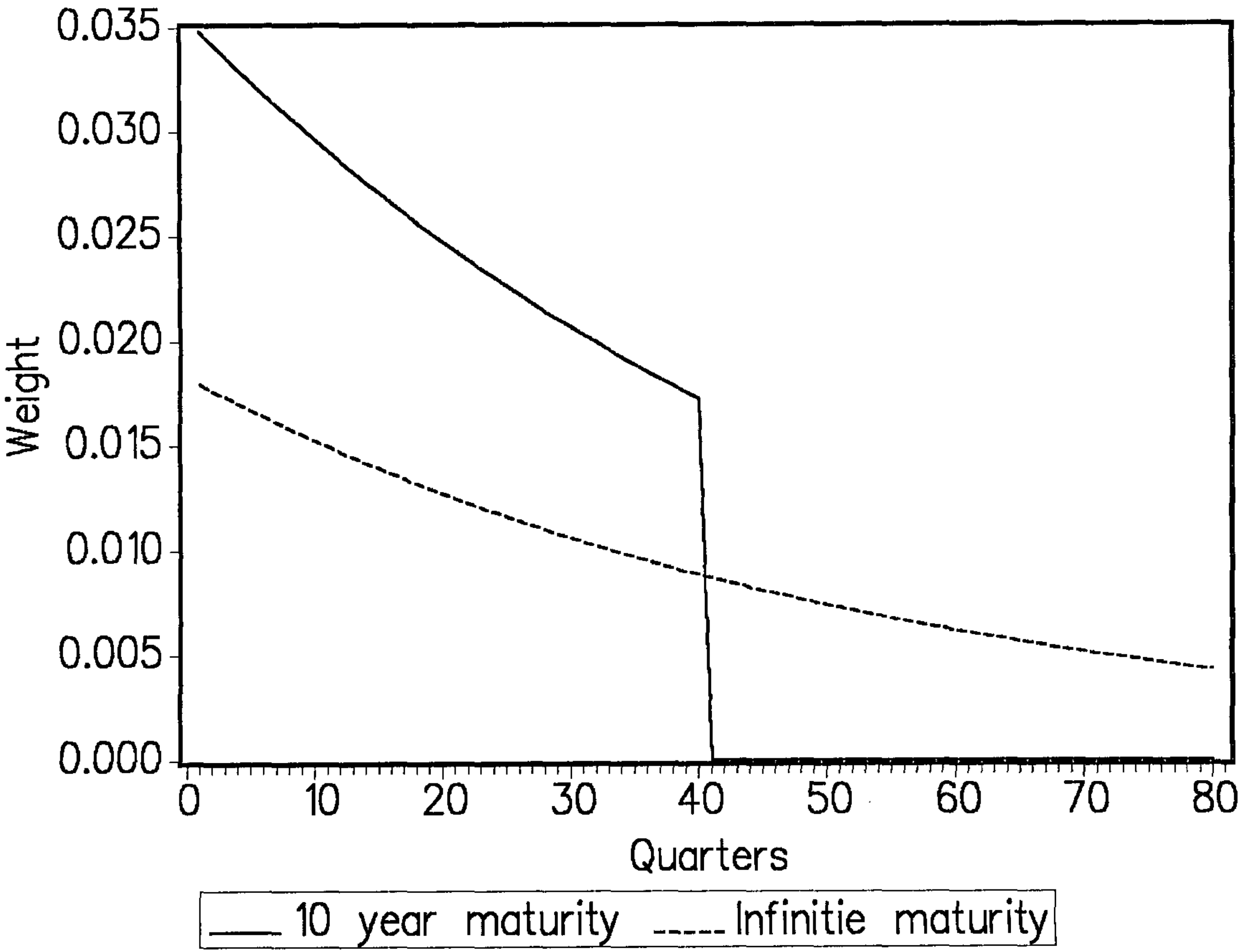

FIGURE 4: IMPLIED VOLATILITY