Journal of Monetary Economics 33 (1994) 285–318. North-Holland

Do stationary risk premia explain it all?

Evidence from the term structure*

Martin D.D. Evans
New York University, New York, NY 10012-1126, USA

Karen K. Lewis
University of Pennsylvania, Philadelphia, PA 19104-6367, USA

Received October 1991, final version received November 1993

Predictable variations in excess returns have often been attributed to the presence of time-varying risk premia. In this paper, we use an insight based upon new techniques from time series analysis to test whether stationary risk premia can alone explain the behavior of excess returns to long bonds relative to rolling over short rates. Surprisingly, we reject this hypothesis using U.S. T-bill returns. We then show that either permanent shocks to the risk premia and/or rationally anticipated shifts in the interest rate process could produce anomalous results.

Key words: Interest rates; Financial markets

JEL classification: E43; G1

1. Introduction

A great deal of research has explored the information in the term structure of interest rates. This research examines whether the relationship between interest

Correspondence to: Karen K. Lewis, Finance Department, The Wharton School, University of Pennsylvania, 2300 Steinberg Hall/Dietrich Hall, Philadelphia, PA 19104-6367, USA. Tel: 215-898-7622, Fax: 215-898-6200.

* We are grateful to Jim Stock for useful comments as well as participants in seminars at the NBER Summer Institute, the Wharton School, the University of Michigan, and the University of Washington. Lewis gratefully acknowledges research support from the Olin Foundation and the National Science Foundation.

rates at different maturities helps to explain future movements in rates.[1] Some of the empirical research has focused upon the 'expectations theory' that relates the yield on long-term bonds to expected future short rates. This theory has frequently been tested and rejected using regression tests of the slope of the yield curve.[2] Explanations for the rejections have focused on the presence of either time-varying risk premia or biased forecast errors in the regression residuals. Either explanation implies that excess bond returns are predictable, since they are the sum of risk premia and forecast errors by definition.

In this paper, we ask whether time-varying risk premia that are covariance-stationary can alone explain the behavior of excess bond returns.[3] For this purpose, we use an insight based upon new techniques from time series analysis. When interest rates contain unit roots and when forecasts are unbiased, the coefficient from the cointegrating regression of the level of a forward interest rate upon the level of its corresponding future interest rate will be contaminated by the presence of a time-varying risk premium only if this risk premium contains a unit root. Thus, the cointegrating coefficient can tell us whether the risk premium contains a unit root component or not.

After reviewing the standard regression tests of the 'expectations theory' in section 2, section 3 of the paper uses this methodology to test whether the risk premia in excess U.S. T-bill returns are I(0)-stationary.[4] The test is rejected for many maturities less than one year. These results point to the surprising conclusion that excess returns are subject to permanent shocks.

We then investigate the source of these findings with a series of Monte Carlo experiments. In section 4 we examine whether our results are due to treating bond yields as variables with unit roots when in fact they are stationary. Our experiments calculate the empirical distribution of the coefficient estimates assuming stationary processes for short-term interest rates and the risk premia. In this case, the results demonstrate that there is a less than 1% probability of observing our coefficient estimates, given the observed degree of autocorrelation in the risk premia. Thus, it is highly unlikely that our results are due to incorrectly assuming that interest rates are subject to permanent disturbances.

In section 5 we conduct a second set of experiments to examine whether our results are consistent with the observed autocorrelation in excess returns. We show that it is possible for risk premia with a small unit root component to generate both our cointegrating regression results and a first-order

[1]Mishkin (1988), Campbell and Shiller (1987), and Fama (1984a), among many others, find that the term structure contains information about future rates.

[2]For a list of references, see Campbell and Shiller (1991) or Shiller (1987).

[3]Risk premia in the term structure are typically treated as stationary variables both in the theoretical and empirical investigations. For a theoretical example, see Backus, Gregory, and Zin (1989) and for empirical examples see footnote 1.

[4]We follow the literature in calling variables that follow stationary processes as I(0) and those that follow processes with permanent disturbances as I(1).

autocorrelation in excess returns as low as we observe in the data. In particular, when permanent shocks contribute to between 5% and 10% of the variability of the risk premia, we are likely to observe the cointegrating results we find in the data but are very unlikely to detect the presence of a unit root in excess returns using standard methods. Our results are therefore completely consistent with the observation that excess returns appear stationary.

Under standard rational expectations, our cointegration results imply that the risk premium contains a unit root component, since excess returns are comprised of a risk premium and a stationary forecast error. In section 6 we offer an alternative explanation for the presence of permanent shocks to excess returns. In particular, we show how rational expectations of a shift in the process for interest rates can produce forecast errors that appear to contain unit roots. Thus, our cointegrating results may be attributable to either permanent shocks to the risk premium, or shifts in the interest rate process, or both.

In section 7 we re-examine the standard regression tests of the 'expectations theory' in the light of our cointegration results. We show how the regression coefficients are affected by the presence of unit roots in excess returns. Interestingly, the coefficients from different regression tests have distinct large-sample distributions that appear consistent with the estimates we observe in the data.

The paper ends with some concluding remarks.

2. Standard regression tests

To illustrate how our cointegrating tests below relate to the previous literature, we begin by briefly reviewing two standard regression tests of the 'expectations theory' of the term structure. As has been documented in earlier studies, some forms of the regression tests tend to reject the theory while others do not. In the final section of this paper, we will show how our cointegration results lead to a reinterpretation of these standard findings.

2.1. The framework

The 'expectations theory' of the term structure of interest rates relates the equilibrium yield of long bonds to the expected value of short rates over the maturity of the bond. Since we will be discussing these results in the presence of time-varying risk premia, we consider a framework that allows for risk premia and incorporates the expectations theory as a special case.[5] For the case of pure

[5]As emphasized by Cox, Ingersoll, and Ross (1981), the expectations hypothesis has several forms. They show that only the 'local expectations hypothesis' is consistent with equilibrium asset pricing models such as in Cox, Ingersoll, and Ross (1985). If excess holding returns were constant, then discretely compounded excess returns would yield our regression restrictions below. See, for example, the model in Vasicek (1977) with constant risk premia. Campbell (1986) and Shiller, Campbell, and Schoenholtz (1983) argue that the expectations hypothesis may be linearly approximated, an approach that we follow below.

discount bonds, this relationship is

$$R_t^k = (1/k) \sum_{i=0}^{k-1} \mathrm{E}_t R_{t+i}^1 + (1/k)\,\theta_t^k, \tag{1}$$

where R_t^k is the yield on a k-period bond purchased at time t, E_t denotes the market's expectations conditional upon information available at time t, and θ_t^k is a time-varying risk premium on holding the k-period bond relative to rolling over one-period bonds. Tests of the expectations theory examine a special case of eq. (1) where the risk premium is equal to zero.

A slightly different form of the risk premia will also prove expositionally useful for the investigation below. In particular, we define the one-period holding premium as

$$\phi_t^k \equiv kR_t^k - (k-1)\mathrm{E}_t R_{t+1}^{k-1} - R_t^1, \tag{2}$$

where ϕ_t^k is the time-varying premium relative to the risk-free one-period rate R_t^1 on a risky position of holding a k-period bond for one period and then selling the proceeds at the prevailing rate. Iterating (2) forward verifies that $\theta_t^k = \sum_{i=0}^{k-1} \mathrm{E}_t \phi_{t+i}^{k-i}$, or that the risk premium on holding a long bond relative to rolling over short bonds is equal to the expected value of the sum of holding premia from today until the maturity period of the long bond. We will use this basic framework to review standard regression tests in order to motivate our cointegrating regression tests below.

2.2. *Yield spread regressions*

According to the expectations theory, the difference between the long k-period bond and the short one-period bond is the market's forecast of the change in the long bond. Hence, tests of the expectations theory have frequently used this yield spread as a regressor.[6] These regression tests have typically taken two forms. In the first, the yield spread predicts the one-period change in the long bond:

$$(k-1)(R_{t+1}^{k-1} - R_t^k) = a_0 + a_1(R_t^k - R_t^1) + u_{1,t+1}. \tag{3}$$

We can re-express (3) in terms of the framework in (1) and the definition in (2) by setting $a_1 = 1$, so that

$$u_{1,t+1} \equiv -\phi_t^k + (k-1)(R_{t+1}^{k-1} - \mathrm{E}_t R_{t+1}^{k-1}) - a_0, \tag{4}$$

[6]Another strand of the literature uses forward premia as regressors. See Evans and Lewis (1990) for a discussion of these studies.

and a_0 equals the mean of the risk premium. Notice that $u_{1,t+1}$ can be interpreted as the *ex post* excess return on the k-period bond relative to the one-period rate. Standard regression tests of the expectations hypothesis would test $a_1 = 1$ and $a_0 = 0$.[7] As (3) and (4) show, however, if risk premia are time-varying and correlated with the yield spread, then the estimates of a_1 will deviate from one due to standard omitted variables problems.

The second form of the yield spread regression relates the yield spread to the *ex post* changes in the short rate over the maturity of the bond:

$$(1/k)\sum_{i=1}^{k-1} R^1_{t+i} - [(k-1)/k]R^1_t = b_0 + b_1(R^k_t - R^1_t) + u_{2,t+k}. \tag{5}$$

This regression can be re-expressed in terms of (1) and (2) by setting $b_1 = 1$, and

$$u_{2,t+k} \equiv -(1/k)\theta^k_t + (1/k)\sum_{i=1}^{k-1}(R^1_{t+i} - \mathrm{E}_t R^1_{t+i}) - b_0, \tag{6}$$

where b_0 equals the mean of the risk premium on holding the bond over its maturity. $u_{2,t+k}$ can be viewed as the *ex post* excess return on holding the k-period bond relative to rolling over short bonds. Standard regression tests of the expectations hypothesis test $b_1 = 1$ and $b_0 = 0$ in eq. (5). Again, if risk premia are time-varying and correlated with the yield spread, (5) and (6) imply that the estimates of b_1 will deviate from one.

The first and second columns of table 1 report the coefficient estimates of a_1 and b_1, respectively, for one-month through eleven-month U.S. T-bill rates. These data correspond to the series first constructed by Fama (1984) using the U.S. Government Securities File of the Center for Research in Security Prices at the University of Chicago. The data set provides prices of U.S. Treasury bills for the end of the month over the period of availability from June 1964 to December 1988. The estimates we report are based upon the midpoints between the bid and ask rates. We will discuss the sensitivity of our results to this choice below.

Table 1 shows that the parameter estimates for a_1 are negative and increase in absolute value with maturity k. All of the coefficients are significantly less than the hypothesized value of one and even become significantly negative as the maturity horizon lengthens. Furthermore, this pattern continues with longer maturities up to ten years, as described in Campbell and Shiller (1991). This regression test points to a strong rejection of the expectations theory.

By contrast, the estimates for b_1 are positive, though less than one at values near 0.4. Since the residuals contain k overlapping forecast errors under the null hypothesis, the reported standard errors are corrected for a moving average

[7]Campbell and Shiller (1991) also examine a modified expectations theory where $a_1 = 1$ and $a_0 = -\phi^k$.

Table 1

Regressions tests of short-horizon change in long bond and long-horizon change in short bond on current yield spread.[a]

This table reports two regressions results. First, the change in the return on the k-period bond is regressed on the spread between a k-period bond and a one-period bond,

$$(k-1)(R_{t+1}^{k-1} - R_t^k) = a_0 + a_1(R_t^k - R_t^1) + u_{1,t+1}, \tag{3}$$

where R_t^k is the annualized return on a k-month U.S. T-bill. Second, the k-period return on rolling over one-period bond returns is also regressed on the spread between a k-period bond and a one-period bond,

$$(1/k)\sum_{i=1}^{k-1} R_{t+i}^1 - [(k-1)/k]R_t^1 = b_0 + b_1(R_t^k - R_t^1) + u_{2,t+k}. \tag{5}$$

Equations are estimated by OLS using Hansen's (1982) estimate of the variance–covariance matrix to allow for conditional heteroskedasticity and moving average error terms. Standard errors are in parentheses. Data are monthly U.S. T-bill rates from the CRSP for the period June 1964 to December 1988.

	Regression coefficients	
	(1)	(2)
Maturity k in months	a_1 (S.E.)	b_1 (S.E.)
2	-0.17^b	0.42^b
	(0.24)	(0.12)
3	-0.43^b	0.32^b
	(0.48)	(0.19)
4	-0.70^b	0.39^b
	(0.61)	(0.19)
5	-1.15^b	0.36^b
	(0.68)	(0.17)
6	-1.27^b	0.38^b
	(0.72)	(0.20)
7	-1.48^b	0.40^b
	(0.79)	(0.18)
8	-1.52^b	0.41^b
	(0.83)	(0.20)
9	-1.72^b	0.37^b
	(0.79)	(0.17)
10	-1.89^b	0.39^b
	(0.86)	(0.20)
11	-1.90^b	0.40^b
	(0.88)	(0.22)

[a]Column (1) is the regression coefficient a_1 for the regression in eq. (3). Column (2) is the regression coefficient b_1 for the regression in eq. (5). Standard errors are corrected for an MA($k-1$) error.
[b]Significantly less than one at the 95% confidence level.

component of order $k-1$. Using these standard errors, the hypotheses that the coefficients equal one are rejected in all cases. For the longer maturities investigated in Campbell and Shiller (1991), the point estimates increase and become insignificantly different than one.

The pattern of rejection in these regression tests provides information about the correlation between the risk premia ϕ_t^k and the yield spread. Applying arguments similar to those in Fama (1984), Evans and Lewis (1990) show how the coefficient estimates of a_1 and b_1 can be interpreted in terms of the variances and covariance between the risk premia and the yield spreads. These interpretations assume that nonoverlapping forecast errors are white noise, consistent with standard rational expectations, and that the time-varying risk premia are I(0)-stationary. Together, these assumptions imply that excess returns must also be I(0)-stationary.

3. Are excess returns stationary?

In this section, we examine whether excess returns are indeed I(0)-stationary. In particular, we will show how the regression tests above can be transformed into a cointegrating regression framework that provides information about long-term or low-frequency behavior of excess returns. Under standard rational expectations assumptions, this framework allows us to jointly test the hypothesis that interest rates contain unit roots while time-varying risk premia do not. In this section, we begin by developing the test and reporting the results. In the next section we consider the robustness of our results if interest rates are actually stationary.

3.1. A joint test of nonstationary interest rates and stationary time-varying risk premia

Empirical studies have found that interest rates appear to have disturbances with unit root components.[8] This observation has led to regression tests based upon differenced interest rate series, such as those in (3) and (5) above. We will first maintain the assumption that interest rates are subject to unit root disturbances and show how cointegrating regressions can provide information about whether excess returns contain unit root disturbances.

To develop the cointegrating regression, we use the definitions in (1) and (2) to rewrite the regression equation (3):[9]

$$(k-1)(R_{t+1}^{k-1} - R_t^k) = (R_t^k - R_t^1) - \phi_t^k + (k-1)(R_{t+1}^{k-1} - \mathrm{E}_t R_{t+1}^{k-1}). \quad (7)$$

[8]See, for example, Mishkin (1989). To confirm these results with our data, we tested the hypothesis of unit roots for interest rates using modified Dickey–Fuller tests and found we could not reject unit roots. Since these findings are similar to those in the literature, we do not report them to save space.

[9]Inspecting regression equation (3) and the definition of its residual in (4) makes clear that the following equation simply rewrites the regression equation with $a_1 = 1$.

After subtracting $(k-1)R_t^k$ from both sides and rearranging, eq. (7) may be rewritten as

$$(k-1)R_{t+1}^{k-1} = \alpha_0 + \alpha_1(k-1)F_t^{k-1,1} + u_{1,t+1}, \tag{8}$$

where α_0 equals a_0, the constant in the stationary regression (3), $\alpha_1 = 1$, $F_t^{k-1,1} = (kR_t^k - R_t^1)/(k-1)$, and $u_{1,t+1}$ is the long-rate regression residual from eq. (4). Within the framework of (1) and (2), this residual is the excess return comprising the risk premium on holding a k-period bond, ϕ_t^k, and the one-step-ahead forecast error on the $k-1$ bond, $R_{t+1}^{k-1} - \mathrm{E}_t R_{t+1}^{k-1}$. $F_t^{k-1,1}$ is the forward rate contracted at time t for a $(k-1)$-period bond to be bought at $t+1$. More generally, we define

$$F_t^{j,l} \equiv [(j+l)R_t^{j+l} - l\,R_t^l)]/j, \tag{9}$$

as the forward rate contracted at time t for a j-period bond to be bought at $t+l$.

We have formed (8) in terms of the levels of interest rates and forward rates to exploit a result from time-series analysis when variables have unit root disturbances. Specifically, as long as the residual is I(0)-stationary, a regression of an I(1)-nonstationary variable on another I(1) variable that shares its stochastic trend will provide an asymptotically consistent coefficient estimate. Such regressions are termed cointegrating regressions [see Campbell and Perron (1991) for a recent discussion].

This result means that, as long as $u_{1,t+1}$ is I(0)-stationary, the regression in (8) provides an estimate of α_1 that is asymptotically consistent and independent of $u_{1,t+1}$. By definition, the residual $u_{1,t+1}$ equals a risk premium plus a forecast error when α_1 equals one (i.e., the excess return). Under standard rational expectations, the one-step-ahead forecast error follows a white noise process, an I(0) process. Therefore, if risk premia are stationary, they cannot contaminate estimates of α_1. Thus, we can test a minimum requirement for stationary risk premia to explain the results of the regressions tests in table 1 by estimating α_1 in eq. (8) and testing whether it is equal to one. Since we have written eq. (8) in the form of I(1) variables, this regression provides coefficient estimates of α_1 that are asympotically consistent even though the yield spread may not be independent of the stationary risk premia.

3.2. Cointegrating regression results for one-period regressions

The first column of table 2 reports the OLS coefficient estimates of α_1 in eq. (8) for different (k-month) maturities of T-bills. As the column shows, all of the point estimates are less than one. While this regression provides asymptotically consistent estimates of α_1, OLS estimates are biased in finite samples. The appendix describes two methods we used that adjust for this bias.

Table 2

Cointegrating regressions of one-month-ahead spot rates on current forward rates.

$$(k-1)R_{t+1}^{k-1} = \alpha_0 + \alpha_1 (k-1)F_t^{k-1,1} + u_{1,t+1}, \quad (8)$$

where α_1 is OLS estimate of the regression, R_t^{k-1} is the yield on a $(k-1)$-month U.S. T-bill, $F_t^{k-1,1}$ is the forward rate on a contract bought at time t for a $(k-1)$-period bond at time $t+1$. H–P$_l$ are the Hansen and Phillips (1989) $\chi^2(1)$ Wald test statistics of the hypothesis that $\alpha_1 = 1$ assuming that $\text{cov}(u_{1,t}\, u_{1,t-r}) = 0$ for all $r > l$ for $l = 0$ and $l = 6$. S–W is the test that $\alpha_1 = 1$ using the $\chi^2(1)$ test statistic from Stock and Watson (1989). Marginal significance levels are in parentheses. Data are monthly U.S. T-bill rates from the CRSP for the period June 1964 to December 1988. The Monte Carlo p-values show the probability of observing the estimate of α_1 in column (1) when the true value of $\alpha_1 = 1$. The upper value assumes conditional homoscedasticity in the data generation process, the lower value allows for conditional heteroscedasticity. Details of the experiments are described in the appendix.

	(1)	(2)	(3)	(4)	(5)
Maturity k in months	α_1	H–P$_0$ (M.S.L.)	H–P$_6$ (M.S.L.)	S–W (M.S.L.)	Monte Carlo p-values
2	0.921	15.43	4.52	5.21	< 0.001
		(< 0.001)	(0.034)	(0.022)	0.003
3	0.929	32.51	9.35	3.51	< 0.001
		(< 0.001)	(0.002)	(0.061)	< 0.001
4	0.951	8.84	1.64	0.26	< 0.001
		(0.003)	(0.203)	(0.610)	0.015
5	0.956	3.82	0.01	0.11	0.009
		(0.051)	(0.999)	(0.740)	0.019
6	0.960	2.62	0.33	0.93	0.032
		(0.106)	(0.566)	(0.335)	0.084
7	0.967	0.09	0.05	0.03	0.420
		(0.764)	(0.823)	(0.862)	0.429
8	0.967	0.25	0.01	0.03	0.203
		(0.617)	(0.999)	(0.862)	0.295
9	0.962	2.88	0.15	0.06	0.029
		(0.090)	(0.699)	(0.806)	0.070
10	0.962	3.16	0.12	0.00	0.009
		(0.075)	(0.729)	(1.00)	0.044
11	0.968	0.01	0.03	0.06	0.537
		(0.999)	(0.862)	(0.0806)	0.538

The first method is from Hansen and Phillips (1990) who use the asymptotic variance–covariance matrix of the residuals to adjust the OLS coefficient for the finite-sample bias. The second and third columns of table 2 report Wald statistics for the null hypothesis that $\alpha_1 = 1$ using the Hansen and Phillips adjustment. When the risk premium, and therefore $u_{1,t+1}$, are serially correlated, a consistent estimator of the covariance matrix must incorporate autocorrelations of the error process. We calculated the Hansen–Phillips statistic for two extreme assumptions about the degree of autocorrelation. The statistics reported in column (2) are based on the assumption of no serial correlation in $u_{1,t+1}$. Those in column (3) allow for autocorrelation for up to six months. Inspection of the regression residuals revealed this to be a conservative over estimate of the degree of autocorrelation.

The results of these tests are reported in the table together with their marginal significance levels. For the statistics assuming no serial correlation, in column (2), the hypothesis that α_1 equals one is rejected with marginal significance levels less that 10% for six of the eleven maturities. The more conservative covariance estimates incorporating autocovariances up to six months tend to blow up the covariance matrix. As a result, the statistics in column (3) do not reject the hypothesis except at maturities of two and three months.

Column (4) of table 2 reports an alternative set of Wald tests for the hypothesis that $\alpha_1 = 1$, using the Stock and Watson (1989) procedure to adjust for the finite-sample bias in the OLS coefficients.[10] As the results indicate, the hypothesis that α_1 equals one is again rejected for the two- and three-month maturities. It should be noted that the tests reported in columns (2)–(4) are not independent across maturities.

The results in table 2 are based on returns calculated from the midpoints of the bid and ask T-bill rates rather than true quote rates, and so are potentially subject to an errors-in-variables bias. Since Stambaugh (1988) has shown that quotation errors can significantly affect the results of standard regression of holding returns on yield spreads, we re-estimated the cointegrating regressions in table 2 using the bid and ask rates to check the robustness of our results. These results, which are reported in the appendix table, are very similar to those in table 2. The appendix also shows that bid and ask rates are cointegrated one for one with each other. This evidence means that the use of the average bid–ask rate in table 2 introduces at most a stationary quotation error into the residual of the cointegrating equation (8), which cannot contaminate the estimates of α_1.[11] These findings indicate that the results in the columns (2)–(4) of table 2 are robust to effects of quotation errors.

[10]As the appendix describes in detail, the Stock and Watson method includes leads and lags of first differences of the right-hand-side variables as additional regressors in order to correct for the finite-sample bias in estimating α_1. The statistics shown uses two leads and lags. As in the Hansen–Phillips method, the test statistics also allow for serial correction in the residuals. Here we allowed for autocorrelation of up to two months. The results are not sensitive to this choice or the number of leads or lags we include.

[11]Our cointegrating regression differs in one other respect from the standard regressions studied by Stambaugh. He notes that regressions of holding returns on yield spreads typically introduce the same current rate on both sides of the regression, possibly biasing the coefficient estimates. This potential source of bias does not appear in eq. (8). The left-hand-side variable is the future realized rate, while the right-hand side is the forward rate and each of these variables are observed at different periods in time.

Column (5) of table 2 reports the results of some Monte Carlo experiments. These experiments were motivated by the fact that both the Hansen–Phillips and Stock–Watson procedures require a consistent estimate of the asymptotic variance–covariance matrix of the regression residuals. Since Chesher and Jewitt (1987) have shown that consistent estimators of this matrix may be biased in small samples, it is possible that the statistics in columns (2)–(4) may be affected by poor estimates of the variance–covariance matrix.

To investigate this issue, we conducted Monte Carlo experiments that incorporated the finite-sample properties in the data. For this purpose, we generated time series for the future interest rate based upon the actual data series. In order to examine the sensitivity of our experiments, we constructed the series in two ways that reflect different assumptions about conditional heteroscedasticity. We then estimated the cointegrating regression (8) repeatedly to produce the empirical distribution of the coefficient estimate α_1. Details of the experiments are provided in the appendix.

Table 2, column (5) reports the results of these Monte Carlo experiments. The numbers are the *p*-values for the hypothesis that the estimates of α_1 at each maturity are significantly different from one. The upper number is the *p*-value from the distribution of coefficients when the data are generated from a conditionally homoscedastic process. The lower number is the corresponding *p*-value based on a conditionally heteroscedastic process. As the column shows, the null hypothesis of $\alpha_1 = 1$ is rejected at the 5% marginal significance level for all maturities except for six, seven, eight, nine, and eleven months. At the 10% level, we would also reject for the six- and nine-month maturities.

Overall, the results in table 2 provide fairly strong evidence against the null hypothesis of $\alpha_1 = 1$. Although the asymptotic results suggest that the null can only be rejected for two- and three-month bonds, the small-sample Monte Carlo experiments – that do not depend upon estimates of the asymptotic variance–covariance matrix – indicate much stronger rejections of the null hypothesis across a wide range of maturities. Since a *minimal* condition for stationary risk premia with standard rational expectations is that α_1 is always equal to one, it is surprising to find rejections at any of the maturities.

3.3. *Cointegrating regressions for ex post short-rate regressions*

We will now use a similar cointegrating regression to examine the short-rate regression tests in (5). For this purpose, subtract $(1/k)R_t^1$ from both sides of the definition (1) as well as the forecast errors $(\mathrm{E}_t R_{t+i}^1 - R_{t+i}^1)/k$ for all $i = 1, \ldots, k-1$. This rearrangement yields the following form for the definition:

$$(1/k)\sum_{i=1}^{k-1} R_{t+i}^1 = (1/k)(kR_t^k - R_t^1) - (1/k)\theta_t^k + (1/k)\sum_{i=1}^{k-1}(R_{t+i}^1 - \mathrm{E}_t R_{t+i}^1). \tag{10}$$

Eq. (10) is the relationship behind the short-rate regression test in (5).[12]

We can now proceed by rewriting $(kR_t^k - R_t^1)$ as the sum of the one-month forward rates, $F_t^{1,i}$, using the definitions of the forward rates in (9). Substituting the result into (10), we obtain

$$(1/k)\sum_{i=1}^{k-1} R_{t+i}^1 = (1/k)\sum_{i=1}^{k-1} F_t^{1,i} - (1/k)\theta_t^k + (1/k)\sum_{i=1}^{k-1}(R_{t+i}^1 - \mathrm{E}_t R_{t+i}^1)$$

$$= (1/k)\sum_{i=1}^{k-1} F_t^{1,i} + u_{2,t+k-1}, \tag{11}$$

where $u_{2,t+k-1}$ is the residual in the regression equation (5) defined in (6) and where we have set the constant b_0 equal to zero for expositional simplicity.[13] In table 1, we examined this relationship for different maturities. With each incremental maturity, k, an additional future spot rate, R_{t+k-1}^1, and forward rate, $F_t^{1,k-1}$, were included in the summations shown in (11). To examine each component of the summation separately, we first note from (11) and (6) that $\theta_t^{k-1} = \sum_{i=1}^{k-2}(F_t^{1,i} - \mathrm{E}_t R_{t+i}^1)$ and define the difference between risk premia on holding bonds between adjacent maturities as $\Delta\theta_t^k \equiv \theta_t^k - \theta_t^{k-1}$. Note also that if the risk premia, θ_t^k and θ_t^{k-1}, are covariance-stationary variables, $\Delta\theta_t^k$ must be as well. Next, subtract the components, $(1/k)\sum_{i=1}^{k-2}\mathrm{E}_t R_{t+1}^1$, from each side of (11). Rearranging and leading the maturity one period, gives

$$R_{t+k}^1 = F_t^{1,k} - \Delta\theta_t^{k+1} + (R_{t+k}^1 - \mathrm{E}_t R_{t+k}^1). \tag{12}$$

We can now rewrite this equation in terms of the cointegrating regression,

$$R_{t+k}^1 = \beta_0 + \beta_1 F_t^{1,k} + w_{t+k}, \tag{13}$$

where $\beta_1 = 1$ and $w_{t+k} = -\Delta\theta_t^{k+1} + (R_{t+k}^1 - \mathrm{E}_t R_{t+k}^1) - \beta_0$.

We may apply the same intuition to the cointegrating regression in (13) as we did to the the regression in (8). Since forward rates contain unit roots, then the forward rate must be asymptotically independent of all stationary variables. Therefore, under standard rational expectations and the null hypothesis of stationary risk premia, the excess return w_{t+k} will be asymptotically independent of $F_t^{1,k}$ and we must find $\beta_1 = 1$.

Table 3 reports the results of estimating eq. (13). The left-hand margin gives the forecast horizon k and column (1) reports the coefficient estimate for β_1. As

[12]To see this clearly, add $[(k-1)/k]\, R_t^1$ to both sides of (5), impose $b_1 = 1$, and use the definition of the residual in (6).

[13]Incorporating b_0 will simply include a constant in the relationship that will be captured by the constant in our cointegrating regression estimated by β_0 below.

Table 3

Cointegrating regressions of k-month-ahead short rates on current forward rates.

$$R^1_{t+k} = \beta_0 + \beta_1 F^{1,k}_t + w_{t+k}, \quad (13)$$

where β_1 is the OLS estimate of the regression, R^1_{t+k} is the rate on a one-month T-bill at $t + k$, $F^{1,k}_t$ is the forward rate on a contract at time t for a one-month bill at time $t + k$. H–P_l denote the Hansen and Phillips (1989) $\chi^2(1)$ Wald test statistics of the hypothesis that $\beta_1 = 1$ assuming that $\text{cov}(w_t, w_{t-r}) = 0$ for all $r > l$ for $l = 3$ and $l = 9$. S–W is the test that $\beta_1 = 1$ using the $\chi^2(1)$ test statistic from Stock and Watson (1989). Marginal significance levels are in parentheses. Data are monthly U.S. T-bill rates from the CRSP for the period June 1964 to December 1988. The Monte Carlo p-values show the probability of observing the estimate of β_1 in column (1) when the true value of $\beta_1 = 1$. The upper value assumes conditional homoscedasticity in the data generation process, the lower value allows for conditional heteroscedasticity. Details of the experiments are described in the appendix.

	(1)	(2)	(3)	(4)	(5)
Horizon k in months	β_1	H–P_3 (M.S.L.)	H–P_9 (M.S.L.)	S–W (M.S.L.)	Monte Carlo p-values
2	0.835	13.66	16.10	10.46	< 0.000
		(< 0.001)	(< 0.001)	(0.001)	< 0.000
3	0.832	7.02	5.62	3.93	0.004
		(0.008)	(0.018)	(0.047)	0.008
4	0.771	7.42	4.11	3.44	0.004
		(0.006)	(0.043)	(0.064)	0.021
5	0.764	7.84	3.84	3.47	0.009
		(0.005)	(0.050)	(0.063)	0.029
6	0.757	6.88	2.17	3.25	0.021
		(0.009)	(0.141)	(0.071)	0.068
7	0.735	8.97	3.24	4.61	0.019
		(0.003)	(0.072)	(0.032)	0.057
8	0.696	11.99	4.98	6.57	0.003
		(0.001)	(0.026)	(0.010)	0.017
9	0.658	17.32	7.48	11.37	0.001
		(< 0.001)	(0.006)	(0.001)	0.031
10	0.694	11.01	4.20	5.86	0.018
		(< 0.001)	(0.040)	(0.015)	0.073

in table 2, the point estimates for the coefficient are all less than one. Columns (2) and (3) report the Hansen–Phillips statistic and marginal significance levels for the hypothesis that $\beta_1 = 1$. The statistics in columns (2) and (3) respectively allow for residual autocorrelations up to three and nine months in the calculation of the asymptotic variance–covariance matrix.[14] As the results show, the

[14]Estimate allowing for autocovariances from zero to six months as in table 2 gave similar results. In table 3, we allow for longer lags to account for the overlap in the forecast errors.

hypothesis is strongly rejected at the 5% marginal significance level for 15 of the 18 cases. In addition, the hypothesis is rejected at the 10% level for all cases except for the six-month-ahead forecast in column (3).

Column (4) of table 3 reports the Wald statistics and marginal significance levels using the Stock–Watson method for the same hypothesis. In this case, we reject $\beta_1 = 1$ at the 10% level for all horizons and at the 5% level for most horizons.

As before, we re-estimated the cointegrating regressions in the table using the bid and ask rates to examine the potential impact of quotation errors. The appendix table shows these results to be very similar to those reported in table 3. We also conducted Monte Carlo experiments (described in the appendix) to insure that our results were not unduly affected by poor estimates of the asymptotic variance–covariance matrix used by both the Hansen–Phillips and Stock–Watson tests. Column (5) of table 3 reports the two sets of Monte Carlo p-values from the null hypothesis of $\beta_1 = 1$. The upper values assume conditional homoscedasticity in the data generation process, while the lower allow for conditional heteroscedasticity. As the results show, the null hypothesis is rejected at the 5% level in 15 of the 18 reported cases and at the 10% level in all cases.

Overall, the results in table 3 provide even stronger evidence against the null hypothesis than those in table 2. They imply that the results of the regression test (5) cannot be interpreted solely in terms of the relationship between covariance stationary risk premia and the yield spread. Some component of the regression residual in (5) is correlated with the permanent component of the forward rate. Since forward rates contain unit roots, this means that the sum of the risk premium and the forecast error must contain a unit root as well.

Tables 2 and 3 also show that the point estimates of the cointegrating regression coefficients α_1 and β_1 are relatively close to one. Thus, while our results indicate that a number of excess returns contain a unit root component, this component may be small empirically. As we will show below in section 5, a small unit root component sufficient to generate results like those in tables 2 and 3 would be very hard to empirically detect with standard techniques. Thus, there is no necessary inconsistency between our cointegration results and the standard observation that excess returns appear stationary.

4. Are the cointegration results robust?

Our cointegrating tests are joint tests of the hypothesis that excess returns are stationary and interest rates contain unit roots. Although unit root tests cannot be rejected in interest rates, these tests cannot distinguish between roots equal to one or very close to one in finite samples. Since the asymptotic distribution of the estimators above are based upon the assumption that interest rates have unit roots, it is possible that our inferences are incorrect because interest rates are in fact borderline stationary.

To address this issue, we consider the following thought experiment. Suppose that interest rates had roots close to, but strictly less than, one. In this case, serial correlation in the risk premia and/or correlation between the risk premia and the yield spread could produce estimates of α_1 and β_1 in the cointegrating regressions that deviate from one due to standard omitted variables problems. If this were so, how likely are we to observed our estimates of α_1 and β_1?

To evaluate this possibility we conducted Monte Carlo experiments where the data was generated from the following processes:

$$R^1_{t+1} = 0.998\, R^1_t + v_{t+1} - 0.065 v_t + 0.057 v_{t-1} - 0.175 v_{t-2}$$
$$- 0.113 v_{t-4}, \qquad \sigma^2_v = 0.621, \tag{14a}$$

$$F^{1,1}_t = \mathrm{E}_t R^1_{t+1} + \phi^2_t, \tag{14b}$$

$$\phi^2_t = -0.291 + \rho\phi^2_{t-1} + e_t, \qquad \sigma^2_e = 0.092. \tag{14c}$$

(14a) is a stationary ARMA(1,4) process estimated from data on the one-month T-bill. (14b) defines the forward rate as the sum of the expected one-month rate [implied by the ARMA process in (14a)] and the risk premia ϕ^2_t. The dynamics of the risk premia are given by (14c). We parameterized this last equation to be consistent with the observed process for excess returns, $F^{1,1}_t - R^1_{t+1}$, assuming that R^1_{t+1} follows the process in (14a) and $\mathrm{corr}(v_t, e_t) = 0$.

We used this data generation process to examine the regression (8) with k equal to two:

$$R^1_{t+1} = \alpha_0 + \alpha_1 F^{1,1}_t + u_{1,t+1}. \tag{15}$$

Standard regression theory suggests that the relationship between the risk premia ϕ_t and the estimates of α_1 will depend upon the degree of serial correlation ρ and $\mathrm{corr}(v_t, e_t)$. We therefore generated the empirical distribution for the estimates of α_1 based upon different assumptions about each of these parameters. Specifically, for a given ρ and $\mathrm{corr}(v_t, e_t)$, we generated a sequence of normal shocks $\{v_t, e_t\}$ equal in length to our data sample. With these shocks we constructed series on R^1_{t+1} and $F^{1,1}_t$ with (14), and used these generated series to run the regression in (15). This entire procedure was repeated 1000 times to form the empirical distributions.

Table 4 reports the lower percentiles (1%, 5%, and 10%) of the empirical distribution using different values of ρ ranging from 0 to 0.9. This range includes $\rho = 0.29$, estimated from the actual data. We can use the empirical distribution of α_1 to consider whether values of ρ consistent with those observed in the data are likely to produce our cointegrating results.

Table 4

Monte Carlo distribution of cointegrating regressions assuming stationary regressors.

We assume that the data on the short rate R_t^1 and forward rate $F_t^{1,1}$ are generated by

$$R_{t+1}^1 = 0.998\, R_t^1 + v_{t+1} - 0.065 v_t + 0.057 v_{t-1} - 0.175 v_{t-2} - 0.113 v_{t-4}, \quad \sigma_v^2 = 0.621, \tag{14a}$$

$$F_t^{1,1} = \mathrm{E}_t R_{t+1}^1 + \phi_t^2, \tag{14b}$$

$$\phi_t^2 = -0.291 + \rho \phi_{t-1}^2 + e_t, \quad \sigma_e^2 = 0.092. \tag{14c}$$

Using this data generation process, the table below reports the empirical distribution of the estimates of α_1 from

$$R_{t+1}^1 = \alpha_0 + \alpha_1 F_t^{1,1} + u_{1,t+1}. \tag{15}$$

The percentiles in the left-hand margin refer to the left-hand tail of the empirical distributions of α_1. All distributions are based on the results of 1000 experiments. The corresponding estimate of α_1 in table 2 is $\alpha_1 = 0.921$.

$\rho =$	0.0	0.1	0.3	0.5	0.7	0.9
			(A) $\mathrm{corr}(v_t, e_t) = 0$			
1%	1.001	0.999	0.993	0.974	0.953	0.781
5%	1.008	1.006	1.000	0.991	0.970	0.860
10%	1.011	1.010	1.005	0.999	0.979	0.886
			(B) $\mathrm{corr}(v_t, e_t) = -0.3$			
1%	1.003	1.001	0.996	0.983	0.959	0.836
5%	1.004	1.009	1.004	0.999	0.978	0.891
10%	1.012	1.012	1.009	1.004	0.989	0.918
			(C) $\mathrm{corr}(v_t, e_t) = 0.3$			
1%	1.002	0.997	0.993	0.976	0.944	0.780
5%	1.009	1.005	0.999	0.988	0.964	0.830
10%	1.010	1.008	1.004	0.995	0.973	0.855

Panel A of table 4 reports the results based upon a benchmark case assuming no correlation between the risk premium and the interest rate, i.e., $\mathrm{corr}(v_t, e_t) = 0$. The probability of observing an estimate for α_1 equal to 0.921, as found in table 2, is less than 1% for autocorrelation coefficients ρ ranging from 0 to 0.7. Only when ρ reaches 0.9 is the probability of obtaining our estimate greater than 10%. However, this value for ρ is inconsistent with the empirically observed estimate of 0.29. Panels B and C give the critical values for the lower percentiles of the empirical distributions when the correlation between the risk premium and the interest rates are -0.3 and 0.3, respectively. These results differ little from those in panel A. The probability of observing our estimate for α_1 is greater than 1% only when we assume an autocorrelation coefficient equal to 0.9.

Overall, these results suggest that we are unlikely to observe cointegrating coefficients as low as those reported in tables 2 and 3, even in the case where interest rates are borderline stationary, unless the degree of autocorrelation in the risk premia is almost as high as the degree of autocorrelation in interest rates themselves. However, the autocorrelations in excess returns are not nearly this high.[15] Thus, it appears unlikely that we would find the cointegrating coefficient estimates we do in tables 2 and 3 if interest rates and risk premia were jointly stationary.

5. Excess returns with nonstationary risk premia

To this point, our evidence suggests that excess returns appear to be much more persistent than previously thought. Under the assumption that interest rates contain unit roots, the cointegrating regression tests indicate that excess returns also have unit roots. Moreover, even if we allow for the possibility that interest rates are borderline stationary, our results indicate that excess returns must also be borderline stationary. These results are quite surprising since the autocorrelations in excess returns are empirically well below one.

In this section, we reconcile our cointegrating regression tests with the low autocorrelations in excess returns. We begin by describing the link between the permanent disturbance to excess returns and the permanent disturbance to the interest rate. Using this link, we demonstrate that plausible relationships between the stationary and nonstationary components of excess returns can generate both the cointegrating regression estimates we find and autocorrelation coefficients in excess holding returns well below one. For the present, we shall continue to assume that nonoverlapping forecast errors are white noise as standard rational expectations implies. Consequently, we shall focus on the dynamics of the risk premia as an explanation for the behavior of the excess returns. In section 6 we will relax the standard rational expectations assumption in order to provide an alternative explanation for the behavior excess returns.

5.1. The cointegrating relationship between risk premia and interest rates

The results above provide information about the cointegrating relationship between the risk premium and the interest rates. To see why, reconsider the cointegrating regression

$$R^1_{t+k} = \beta_0 + \beta_1 F^{1,k}_t + w_{t+k}\,. \tag{13}$$

[15] Heston (1991), for example, finds that the autocorrelation of excess holding returns are well below one.

This equation estimates the cointegrating variables, $\beta' \equiv (1, -\beta_1)$, for the vector, $X_t' \equiv (R_{t+k}^1, F_t^{1,k})$, such that $\beta' X_t$ is I(0)-stationary. Therefore, by construction, the excess return can be written as a stationary component and a unit root component that depends upon the unit root in forward rates,

$$R_{t+k}^1 - F_t^{1,k} = (\beta_1 - 1) F_t^{1,k} + \mathrm{I}(0)\ \text{terms}. \tag{16}$$

Because $R_{t+k}^1 - \mathrm{E}_t R_{t+k}^1$ is I(0)-stationary under standard rational expectations, the identity in (12) implies that $R_{t+k}^1 - F_t^{1,k} = -\Delta\theta_t^{k+1} + \mathrm{I}(0)$ terms. Combining this expression with (16) gives

$$-\Delta\theta_t^{k+1} = (\beta_1 - 1) F_t^{1,k} + \mathrm{I}(0)\ \text{terms}. \tag{17}$$

Therefore, since forward rates and spot rates are cointegrated, the risk premium $\Delta\theta_t^{k+1}$ must also be cointegrated with these rates according to the divergence between β_1 and one.

If β_1 is close to one, as our results suggest, and if the variability in the I(0)-stationary component of the risk premium is large, then the unit root component in the risk premia may be difficult to detect with univariate time series tests of the excess returns. In our bivariate setting, however, the low-frequency relationship between spot and forward interest rates is sufficiently informative to detect the unit root components.

5.2. *Monte Carlo evidence*

We conducted a series of Monte Carlo experiments to examine whether a small unit root component in the risk premium can reconcile our cointegration results with the conventional wisdom that excess returns are stationary. The experiments were based upon the following processes:

$$\begin{aligned} R_t^1 &= R_{0,t} + R_{1,t}, \quad R_{0,t} = 0.792 R_{0,t-1} + v_t, \quad \sigma_v^2 = 0.482, \\ & \qquad R_{1,t} = R_{1,t-1} + u_t, \quad \sigma_u^2 = 0.219, \end{aligned} \tag{18a}$$

$$F_t^{1,k} = \mathrm{E}_t R_{t+k}^1 + \Delta\theta_t^{k+1}, \tag{18b}$$

$$\begin{aligned} \Delta\theta_t^{k+1} &= \phi_{0,t} + \phi_{1,t}, \quad \phi_{0,t} = -0.291 + e_t, \quad \sigma_e^2 = 0.092, \\ & \qquad \phi_{1,t} = \phi_{1,t-1} + \sqrt{\lambda}(\sigma_e/\sigma_u) u_t. \end{aligned} \tag{18c}$$

Eq. (18a) describes an I(1) process for the short rate that comprises the sum of a random walk and an AR(1) process. This specification implies that the first difference of the short rate follows the ARMA(1, 1) process we estimate in the data.[16] (18b) shows how forecasts of future short rates [derived from (18a)] are combined with the risk premia to form forward rates. (18c) specifies an I(1) process for the risk premia comprising the sum of a random walk $\phi_{1,t}$ and stationary component $\phi_{0,t}$. (For notational simplicity we have suppressed the k superscript on these components.) The unit root component of the risk premium depends upon the permanent disturbance in the interest rates according to λ. This variable parameterizes the ratio of the variance of the I(1) component to the I(0) component in the risk premium, e.g., $\lambda = (\mathrm{var}(\Delta\phi_1)/\mathrm{var}(\phi_0))$. We will examine the effects of changing λ upon the cointegrating coefficients as well as the first-order autocorrelation in excess returns.

In each of our experiments we first generated a sequence of normal shocks $\{e_t, u_t, v_t\}$ equal in length to our data sample. Given a value for λ, we then constructed series for R^1_{t+k} and $F^{1,k}_t$ in (18) from these shocks. We then used these constructed series to both run the cointegrating regression

$$R^1_{t+k} = \beta_0 + \beta_1 F^{1,k}_t + w_{t+k}, \tag{13}$$

and to estimate the autocorrelations in excess returns $R^1_{t+k} - F^{1,k}_t$. These steps were repeated 10,000 times in each experiment.

Table 5 reports the results of the Monte Carlo experiments for different levels of λ and for two different horizons, $k = 1$ and 6 months. Panel A illustrates the empirical distribution of the estimates of β_1 for the values of λ ranging from 0 to 0.10. The first three rows report the 5, 50, and 95 percentiles of the empirical distributions. The fourth row shows the p-values of the estimates we found in tables 2 and 3 with respect to each of these distributions. When $\lambda = 0$ (i.e., when the risk premium process does not contain a unit root), the probability is less than 3% that we would find estimates as low as we do in tables 2 and 3; i.e., when $k = 6$, the p-value is 2.6%, and when $k = 1$, the p-value is less than 0.001. However, the estimates become quite likely for values of $\lambda(= \mathrm{var}(\Delta\phi_1)/\mathrm{var}(\phi_0))$ between 0.05 and 0.10. Thus, we conclude that our estimates are consistent with risk premium processes where only 5% to 10% of the variance comes from a unit root component relative to the stationary component.[17]

[16]The estimated process is $\Delta R^1_t = 0.792\ \Delta R^1_{t-1} + v_t - 0.862\ v_{t-1},\ \sigma^2_v = 0.651$.

[17]These are probably conservative estimates of how large λ has to be to explain our results. Our specification for the risk premia process in (18c) implies that all of the serial correlation in excess returns arise from the unit root component. Serial correlation in the stationary component $\phi_{0,t}$ would add to the persistence of the risk premia, which would push the cointegrating coefficient further below one (as in table 4). The process in (18c) forces the variance of the unit root component (i.e., λ) to be higher than the truth in order to explain the divergence of the cointegrating coefficient, β_1, from one.

Table 5

Monte Carlo distribution of cointegrating regressions with nonstationary risk premia.[a]

We assume that the data on the short rate R_t^1 and forward rate $F_t^{1,k}$ are generated by

$$R_t^1 = R_{0,t} + R_{1,t}, \qquad R_{0,t} = 0.792 R_{0,t-1} + v_t, \quad \sigma_v^2 = 0.482, \tag{18a}$$
$$R_{1,t} = R_{1,t-1} + u_t, \qquad \sigma_u^2 = 0.219,$$

$$F_t^{1,k} = \mathrm{E}_t R_{t+k}^1 + \Delta\theta_t^{k+1}, \tag{18b}$$

$$\Delta\theta_t^{k+1} = \phi_{0,t} + \phi_{1,t}, \qquad \phi_{0,t} = -0.291 + e_t, \qquad \sigma_e^2 = 0.092, \tag{18c}$$
$$\phi_{1,t} = \phi_{1,t-1} + \sqrt{\lambda}(\sigma_e/\sigma_u)u_t.$$

Using this data generation process, the table reports the empirical distribution of the estimates of β_1 from

$$R_{t+k}^1 = \beta_0 + \beta_1 F_t^{1,k} + w_{t+k}, \tag{13}$$

and the autocorrelations ρ_i, in excess returns $R_{t+k}^1 - F_t^{1,k}$. The left-hand margin shows the percentiles of the empirical distributions. All distribution are based on the results of 10,000 experiments.

	$k = 1$			$k = 6$		
$\lambda =$	0.0	0.05	0.10	0.0	0.05	0.10
			(A) Marginal distribution of β_1			
5%	0.991	0.916	0.888	0.792	0.737	0.717
50%	1.004	0.930	0.903	0.940	0.871	0.846
95%	1.021	0.950	0.923	1.009	0.935	0.906
p-value	0.000	0.127	0.936	0.026	0.074	0.105
			(B) Marginal distribution of ρ_k for $R_{t+k}^1 - F_t^{1,k}$			
5%	− 0.101	− 0.019	0.024	− 0.212	− 0.193	− 0.177
50%	− 0.003	0.118	0.199	− 0.027	0.007	0.033
95%	0.092	0.343	0.503	0.160	0.206	0.255
			(C) Conditional distributions of ρ_k for $R_{t+k}^1 - F_t^{1,k}$ [conditioned on $\beta_1 \leq \beta^-$]			
5%	− 0.098	− 0.007	0.056	− 0.245	− 0.242	− 0.239
50%	− 0.008	0.119	0.194	− 0.066	− 0.066	− 0.050
95%	0.089	0.271	0.414	0.133	0.113	0.131
			(D) Conditional distributions of ρ_k for $R_{t+k}^1 - F_t^{1,k}$ [conditioned on $\beta_1 \geq \beta^+$]			
5%	− 0.096	− 0.088	− 0.049	− 0.204	− 0.251	− 0.216
50%	− 0.004	0.024	0.056	− 0.030	0.054	− 0.039
95%	0.093	0.121	0.175	0.133	0.128	0.148

[a]The p-values are calculated using the estimates of $\beta_1 = 0.921$ ($k = 1$) and $\beta_1 = 0.757$ ($k = 6$) from table 3. β^- and β^+ denote the 5% and 95% critical values for β_1 calculated from the empirical distributions. ρ_i is the ith autocorrelation coefficient.

This evidence shows that a small unit root component in the risk premia can generate our cointegrating regression results. However, it does not explain why standard unit root tests on excess returns would not pick up this component, nor why the autocorrelations in excess returns are quite small. To address this issue, panel B of table 5 provides the 5, 50, and 95 percentiles of the distribution of ρ_k, the kth-order autocorrelation of excess returns, for each of the levels of λ.[18] As the panel shows, the upper tail of the distribution of first-order autocorrelation is about 0.5 for $\lambda = 0.10$ and 0.3 for $\lambda = 0.05$. (The probability of observing both $\rho_1 = 0.29$ and $\rho_6 = -0.07$ estimated in the actual data are greater than 10% in each case.) Thus, the autocorrelation coefficients are typically well below one, even though a unit root component is incorporated into the returns by construction! This evidence indicates that there is a less that 5% probability that we would observe ρ_k greater than 0.5 even though the unit root component in the risk premia is sufficiently large to generate our cointegrating regression results.

Examining the joint distribution of the cointegrating regression coefficient, β_1, and the autocorrelation coefficient, ρ_k, provides even stronger evidence for these basic findings. Panels C and D in table 5, respectively, report the percentiles of the distribution of ρ_k conditioned on the cointegrating coefficients being in the lower and upper tails of their distribution. For the $k = 1$ case, β_1 coefficients in the lower 5% of the distribution generate estimates of ρ_1 that are slightly higher in panel C than for the corresponding estimates when β_1 is in the upper tail in panel D. Moreover, the probability of observing the autocorrelation coefficients we observe still remains high even when we condition upon unrealistically low values of β_1. We therefore conclude that unit root components in risk premia consistent with our estimates of β_1 in the cointegrating regressions would be very unlikely to generate autocorrelation coefficients in excess returns even as high as 0.5.

Overall, these results indicate that if the sample variance of the unit root component in excess returns is 5% to 10% of the variance of the stationary disturbances, then we would be very likely to find: (a) cointegrating regression results consistent with our estimates and (b) low serial correlation in excess returns. Thus, there is no inconsistency between our cointegrating results and standard test results indicating that excess returns appear I(0)-stationary.

6. Excess returns with shifts in the interest rate process

The results in the previous section show that a small unit root in excess returns can simultaneously explain our cointegration results and the observed

[18]Note that excess returns at a six-month horizon contain overlapping forecast errors that induce an MA(5) term. Taking the sixth-order autocorrelation avoids picking up the effects of the overlap.

autocorrelations in excess returns. Since excess returns are comprised of a risk premia and forecasts error, under standard rational expectations, the unit root in excess returns must be attributable to the risk premia. From this perspective, our results highlight the need for future research examining the source of the I(1)-nonstationary component in the risk premia.

There exists an alternative explanation for the presence of permanent shocks in excess returns, however. In this section, we will show how rational expectations of shifts in the process of interest rates can produce systematic forecast errors. It is therefore quite possible for excess returns to contain small unit roots even though the risk premia are I(0)-stationary. We should stress that this discussion is only meant to be suggestive. Our aim is to demonstrate that occasional shifts in the interest rate process can make excess returns appear to contain permanent shocks. The question of whether permanent shocks to the risk premia or shifts in the interest rate process, or both, contribute to the unit root in excess returns is left for future research.

Rational forecasts of process shifts. The effects of rationally expected shifts in the interest rate process may be illustrated with an example. Suppose the market believes at time t that the policy process generating interest rates may change by period $t+j$ with probability $\pi_{t,j}$. Let interest rate realizations from the current process be denoted $R_{C,t}$ and realizations from the alternative process be written $R_{A,t}$, where we suppress the superscript k for maturity to provide expositional clarity. In this case, the market's assessed forecast at time t of the rate at $t+j$ is

$$\mathrm{E}_t R_{t+j} = (1-\pi_{t,j})\mathrm{E}_t R_{C,t+j} + \pi_{t,j}\mathrm{E}_t R_{A,t+j}, \tag{19}$$

where $\mathrm{E}_t R_{C,t+j}$ and $\mathrm{E}_t R_{A,t+j}$ are the expected values conditional upon time t information of time $t+j$ rates that are realizations from the 'C' process the 'A' process, respectively.

Now suppose that *ex post* the interest rate process continues to be driven by the current process 'C'.[19] In this case, the observed rates would be uncorrelated with the forecasts that were conditioned upon the current process 'C', i.e., $\mathrm{E}\{R_{C,t+j} - \mathrm{E}_t R_{C,t+j}\} = 0$. However, the actual market forecast error may be correlated with current information. For example, if the forecasting horizon is j periods ahead, the market's error in forecasting when viewed *ex post* is the difference between the actual realized rate and the forecast given in eq. (19):

$$\begin{aligned}\varepsilon_{t+j} &\equiv R_{t+j} - \mathrm{E}_t R_{t+j} \\ &= (R_{C,t+j} - \mathrm{E}_t R_{C,t+j}) + \pi_{t,j}(\mathrm{E}_t R_{C,t+j} - \mathrm{E}_t R_{A,t+j}) \\ &= (R_{C,t+j} - \mathrm{E}_t R_{C,t+j}) + \delta_{t,j}\mathrm{E}_t R_{C,t+j},\end{aligned} \tag{20}$$

[19]For the following analysis, this condition is more restrictive than we need. As will be discussed below, we require only that any switches in the process occur infrequently within the sample.

where $\delta_{t,j} \equiv \pi_{t,j}(1 - \mu_{t,j})$ and $\mathrm{E}_t R_{A,t+j} \equiv \mu_{t,j} \mathrm{E}_t R_{C,t+j}$. Although the first term on the right is uncorrelated with current information, the second term depends upon the efficient forecast conditional upon the current process according to a parameter that reflects both the probability of a switch, $\pi_{t,j}$, and the difference between forecasts, $(1 - \mu_{t,j})$.[20]

Eq. (20) provides a simple way to express the implications of process switching. When there is no anticipated switch, δ is always equal to zero and we have the standard rational expectations condition on the forecast errors since the mean of $(R_{C,t+j} - \mathrm{E}_t R_{C,t+j})$ is zero. But if a switch is anticipated, δ_t varies over time and measures the direction and magnitude of the switch. For example, if people anticipate that rates will shift to a process with a higher conditional expectation, then $\mathrm{E}_t R_{A,t+j} > \mathrm{E}_t R_{C,t+j}$ so that $\mu_{t,j}$ will be greater than one. If this shift does not occur in the sample, then $\delta_{t,j}$ will be negative, on average, so that the sample mean of the forecast errors will also be negative.[21]

The implications of process switching for the behavior of excess returns are easily seen. Combining the identity in (12), $\mathrm{E}_t R^1_{t+k} \equiv F^{1,k}_t - \Delta\theta^{k+1}_t$, with the link between the true expected interest rate and the expectation conditional upon the current process, $\mathrm{E}_t R^1_{t+k} = (1 - \delta_{t,k}) \mathrm{E}_t R^1_{C,t+k}$, we have that $\mathrm{E}_t R^1_{C,t+k} = [F^{1,k}_t - \Delta\theta^{k+1}_t]/(1 - \delta_{t,k})$. When future rates are realized from the current process, we may use this expression to write the excess returns as

$$R^1_{C,t+k} - F^{1,k}_t = [\delta_{t,k}/(1 - \delta_{t,k})] F^{1,k}_t - (1 - \delta_{t,k})^{-1} \Delta\theta^{k+1}_t + R^1_{C,t+k} - \mathrm{E}_t R^1_{C,t+k}. \tag{21}$$

Eq. (21) shows that excess returns will share the same unit root process as the forward rate during periods where $\delta_{t,k}$ persistently deviates from zero. Thus, anticipated switches in the interest rate process can generate the appearance of unit roots in excess returns which in turn will cause the sample cointegrating coefficient on the forward rate to deviate from one. Notice also that the coefficient on the forward rate depends upon the magnitude of the probability of a switch over k periods, $\pi_{t,k}$, through the term $\delta_{t,k}$. Insofar as the probability of

[20] As discussed in Lewis (1989), a similar relationship arises if the market believes that a policy shift may have recently taken place.

[21] If switches occur within the sample, but market participants believe rates may revert back to the old process [as in Lewis (1991)] or they are learning about the new process [as in Lewis (1989)], then forecast errors would continue to have the same basic form as (20). In this case, we would have

$$\varepsilon_{t+j} = (R_{A,t+j} - \mathrm{E}_t R_{A,t+j}) + (1 - \pi_{t,j})[1 - (\mu_{t,j})^{-1}] \mathrm{E}_t R_{A,t+j}.$$

With interest rate realizations arising from process 'A', the first component would be white noise, while systematic forecast errors would arise from the second term. If the probability of a shift back to process 'C' is not zero, then forecast errors with incorporate a component that depends upon the expected alternative process.

a switch increases with k, (21) suggests that we should expect to see cointegrating coefficients 'pushed' further below one as the forecast horizon lengthens. Interestingly, this pattern is precisely what we observe in tables 2 and 3.

It is worth stressing that this discussion does not rule out the possibility that the risk premium contains a unit root. As (21) shows, permanent shocks to the risk premia can also induce unit roots in excess returns through the second term $(1-\delta_{t,k})^{-1}\Delta\theta_t^{k+1}$. If we redefine $\phi_{1,t}$ in our Monte Carlo experiments to represent the unit root components in $[\delta_{t,k}/(1-\delta_{t,k})]F_t^{1,k}-(1-\delta_{t,k})^{-1}\times \Delta\theta_t^{k+1}$, the results in table 5 can be reinterpreted as showing the combined effects of process shifts and/or permanent shocks to the risk premia.

In summary, if agents anticipate shifts in the interest rate process, then excess returns may appear to contain a unit root component whether or not the risk premia are subject to permanent shocks. This component will trend with the permanent component of forward rates, biasing the cointegrating coefficient estimates on forward rates away from one.

7. Reinterpreting standard regressions

The results above provide evidence of small but statistically significant permanent disturbances in excess bond returns. These findings raise the question of how these disturbances contribute to the standard findings of regression tests in table 1. In this section, we return to the results in table 1 and consider how they are affected by the presence of permanent shocks to excess returns.

For this analysis we will use a decomposition of interest rates and risk premia similar to that introduced in the Monte Carlo experiments described in section 5. Based upon the discussion in section 6, we now recognize that the unit root in the predictable component in excess returns, ϕ_{1t}, can come from either expectational errors or risk premia or both. We represent the process for the short rate and this predictable excess return as

$$\begin{aligned} R_t^1 &= R_{0,t}+R_{1,t}, \qquad R_{0,t}\sim I(0), \qquad R_{1,t}=R_{1,t-1}+u_t,\\ \phi_t &= \phi_{0,t}+\phi_{1,t}, \qquad \phi_{0,t}\sim I(0), \qquad \phi_{1,t}=\phi_{1,t-1}+\lambda u_t. \end{aligned} \tag{22}$$

Our analysis below allows for correlation between I(0) components of R_t^1 and ϕ_t, but not between these components and u_t.[22]

[22]Allowing for this correlation significantly complicates the analysis, but the basic tendencies of the estimators will continue to hold. Notice also that we could incorporate the effects of process shifts if we defined $\phi_{1,t}$ as the unit root component in $[\delta_{t,k}/(1-\delta_{t,k})]F_t^{1,k}-(1-\delta_{t,k})^{-1}\,\Delta\theta_t^{k+1}$ and assumed that all realizations came from the current process. Allowing for actual shifts in the process would add further complications to the analysis that we leave for further work.

In order to examine the regression coefficients across equations involving different maturity bonds, we need to place some minimal restrictions on the structure of predictable excess returns across maturities. For this purpose, we assume that the unit root components of these returns have a single-factor structure:

$$\phi_{1,t}^{k} = \gamma(k)\phi_{1,t}. \tag{23}$$

Eq. (23) does not place any restriction on the total number of factors governing the risk premia since the stationary components may have a multiple-factor structure.[23]

7.1. *Implications for long-rate regression tests*

We can use the structure in (22) and (23) to evaluate the effects of permanent disturbances upon the regression tests reported in table 1. For this purpose, it will be useful to define the yield spread as $\nabla R_t^k \equiv R_t^k - R_t^1$, so that the first regression test of the change in the long rate on the yield spread in eq. (3) can be written as

$$(k-1)(R_{t+1}^{k-1} - R_t^k) = a_0 + a_1 \nabla R_t^k + u_{1,t+1}. \tag{3'}$$

Using eq. (1) and (2) and standard regression theory, the estimate of a_1 is

$$a_T = 1 - \frac{\text{cov}_T(\phi_t^k, \nabla R_t^k)}{\text{var}_T(\nabla R_t^k)}, \tag{24}$$

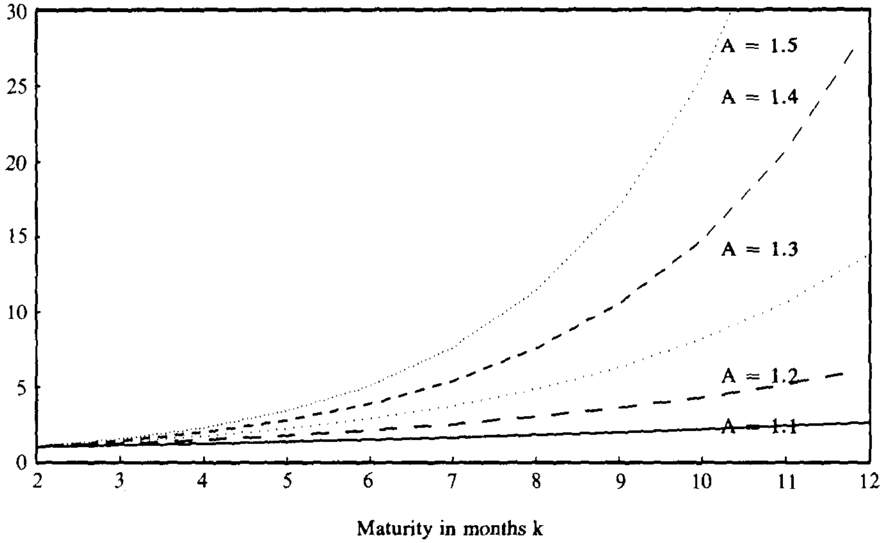
where $\text{var}_T(\cdot)$ and $\text{cov}_T(\cdot)$ denote the sample variance and covariance for sample size T.

To relate these sample moments to the estimates of a_1 when the risk premium contains a unit root component, we note that the spread can also be written as $\nabla R_t^k = \nabla R_{0,t}^k + \nabla R_{1,t}^k$, where the subscripts 0 and 1 refer to the I(0) and I(1) components, respectively. Using this decomposition, the appendix shows that the sample estimate of a_1 can be written as

$$\begin{aligned} a_T - 1 = (\hat{a}_T - 1)&\frac{\text{var}_T(\nabla R_{0,t}^k)}{\text{var}_t(\nabla R_{1,t}^k) + \text{var}_T(\nabla R_{0,t}^k)} \\ &- \frac{k\gamma(k)\Gamma(k)}{\Gamma(k)^2 + k^2 \text{var}_T(\nabla R_{0,t}^k)\text{var}_T(\phi_{1,t})^{-1}}, \end{aligned} \tag{25}$$

[23]In this respect (23) is consistent with Stambaugh's (1988) finding that a single-factor structure is rejected across maturities of one-period holding returns. The only assumption in (23) is that interest rates across maturities are cointegrated, consistent with our findings above.

Factor structures for the I(1) component of the risk premia, $\gamma(k) = A^k$



Limiting value of the a_T

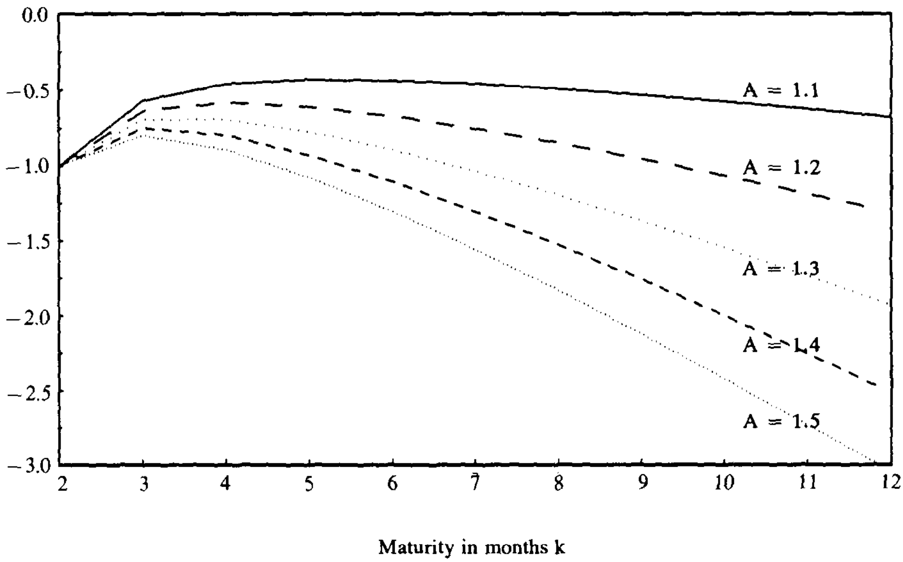


Fig. 1

where $\Gamma(k) \equiv \sum_{i=0}^{k-2} \gamma(k-i)$ and $\hat{a}_T$ is the estimate of a_1 that would be obtained if the risk premium were stationary (based on T observations).

In addition to the conventional factors that push the estimate of a_1 away from one when the risk premia are stationary (identified by $\hat{a}_T - 1$), eq. (25) shows that a_T will depend upon the factor structure $\gamma(k)$ and the sample covariances and variances of the spreads and the I(1) component of the risk premia. As the sample size increases, the variance of the unit root in the yield spread will explode and the first term on the right-hand side will vanish. Similarly, the variance of the unit root in the risk premium, $\text{var}_T(\phi_{1,t})$, will become arbitrarily large, and the last term will collapse to the ratio of factor structure parameters, $k\gamma(k)/\Gamma(k)$. Thus the limiting value of a_T reveals nothing about the correlation between the stationary components of the risk premia and the yield spread.

Fig. 1 plots the limiting value of a_T [equal to $1 - k\gamma(k)/\Gamma(k)$] for different maturity regressions and different factor structures. The upper panel of the figure plots examples of different factor structures $\gamma(k) = A^k$. The lower panel shows the corresponding limiting coefficients across maturities. For all the factor structures, the coefficients tend to decline as the maturity increases. Interestingly, the estimates of a_1 in table 1 show a similar pattern.

7.2. Implications for short-rate regression tests

In the case of the second standard regression,

$$(1/k) \sum_{i=0}^{k-1} R_{t+i}^1 - [(k-1)/k] R_t^1 = b_0 + b_1 \nabla R_t^k + u_{2,t+k}, \tag{5}$$

the framework in (1) and (2) implies that the estimate of b_1 is given by

$$b_T = 1 - \frac{1}{k} \sum_{i=0}^{k-2} \frac{\text{cov}_T(\text{E}_t \phi_{t+i}^{k-i}, \nabla R_t^k)}{\text{var}_T(\nabla R_t^k)}. \tag{26}$$

When the predictable component of excess returns follows an I(1) process, the appendix shows that

$$b_T - 1 = (\hat{b}_T - 1) \frac{\text{var}_T(\nabla R_{0,t}^k)}{\text{var}_T(\nabla R_{1,t}^k) + \text{var}_T(\nabla R_{0,t}^k)}$$

$$- \frac{\Gamma(k)^2}{\Gamma(k)^2 + k^2 \text{var}_T(\nabla R_{0,t}^k) \text{var}_T(\phi_{1,t})^{-1}}, \tag{27}$$

where $\hat{b}_T$ is the sample estimate of b_1 if the risk premium is stationary.

Inspection of (27) reveals that as the sample size increases and the variance of the unit root components rises, the first term and the second term on the right-hand side go to zero and to one, respectively. Thus, the estimate of b_1 goes to zero regardless of the factor structure. Interestingly, this result also appears consistent with the evidence in table 1, where the point estimates of b_1 are all relatively low and close to zero. Furthermore, there is no apparent pattern to the coefficients across maturities.

8. Concluding remarks

In this paper, we used a cointegrating regression framework to test whether excess returns of U.S. T-bills are stationary. Surprisingly, we rejected this hypothesis for bonds of two- to eleven-month maturities. The cointegrating regression analysis suggests that a component of excess returns is subject to permanent shocks. Monte Carlo experiments show that a permanent disturbance with variance equal to 5% to 10% of the stationary disturbance in excess returns can explain our results. A component of this size would induce both cointegrating regression estimates compatible with our findings and low serial correlation in excess returns. We also demonstrated that this unit root component would tend to make some standard regression tests reject at longer horizons, but not others. This implication is also consistent with the empirical evidence.

The cointegrating results presented in this paper can be interpreted in two ways. First, time-varying risk premia could contain unit roots, in contradiction to conventional wisdom. Second, errors in forecasting interest rates could be systematically biased over some periods in the post-war sample reflecting rationally anticipated shifts in interest rates processes. It is the task of future research to sort out these explanations.

Appendix 1: Cointegrating methods

In this appendix, we describe the methods used in section 3 to adjust for the small sample bias in the cointegrating regressions (8) and (13). For more detailed and thorough discussions, see Hansen and Phillips (1989) and Stock and Watson (1989). We will focus upon eq. (8) although the same analysis applies to (13) as well.

For notational simplicity, note that eq. (8) may be written as

$$y_t = \gamma x_t + u_{1t} \tag{A.1}$$

$$\Delta x_t = u_{2t}, \tag{A.2}$$

where u_{1t} and u_{2t} are stationary, $y_t \equiv R_t^{k-1}, x_t \equiv F_t^{k-1,1}$, and the constant term is omitted for simplicity. We are interested in testing $\gamma = 1$. Since x_t is

Table A.1

Cointegrating regressions with bid and ask rates.[a]

This table examines the cointegrating regressions

$$(k-1)R^{ak-1}_{t+1} = \alpha_0 + \alpha_1(k-1)F^{bk-1,1}_t + u_{1,t+1}, \quad (8)$$

$$R^{a1}_{t+k} = \beta_0 + \beta_1 F^{b1,k}_t + w_{t+k}, \quad (13)$$

$$F^{a1,k}_t = \gamma_0 + \gamma_1 F^{b1,k}_t + \xi_t,$$

where the 'a' and 'b' superscripts denote ask and bid rates, R^k_t is the yield on a k-month U.S. T-bill and $F^{j,l}_t$ is the forward rate contracted at time t for a j period bond to be bought at $t+l$. The table below reports Wald tests for the hypotheses that $\alpha_1 = 1$, $\beta_1 = 1$, and $\gamma_1 = 1$, using the Stock and Watson procedure to adjust for small-sample bias in the OLS coefficient estimates. Marginal significance levels are in parentheses. Data are monthly U.S. T-bill rates from CRSP for the period June 1964 to December 1988.

	(1)	(2)	(3)
Maturity k in months	H_0: $\alpha_1 = 1$ (M.S.L.)	H_0: $\beta_1 = 1$ (M.S.L)	H_0: $\gamma_1 = 1$ (M.S.L)
2	5.949	11.242	0.721
	(0.015)	(<0.001)	(0.396)
3	12.588	4.176	3.665
	(<0.001)	(0.041)	(0.056)
4	3.179	3.386	2.129
	(0.074)	(0.066)	(0.144)
5	1.709	2.300	1.757
	(0.191)	(0.129)	(0.185)
6	1.394	4.172	2.680
	(0.238)	(0.041)	(0.102)
7	0.171	5.331	2.475
	(0.679)	(0.021)	(0.116)
8	1.301	4.965	3.300
	(0.254)	(0.026)	(0.103)
9	3.158	6.465	1.103
	(0.053)	(0.011)	(0.294)
10	1.243	6.465	3.101
	(0.265)	(0.047)	(0.078)
11	0.984	6.011	2.680
	(0.321)	(0.014)	(0.102)

[a]The statistics in columns (1) and (2) use the same Stock–Watson procedure as tables 2 and 3. For the statistics in column (3), we include three leads and lags of $\Delta F^{b1,k}_t$ on the right-hand side, and allow for autocovariances up to six lags in the variance–covariance matrix.

endogenous, it is likely that $\mathrm{cov}(x_t, u_{1t}) \neq 0$. In this case, γ will be biased in any finite sample. Therefore, test statistics on γ must take account of this bias. This bias arises even though the estimate of γ remains consistent. We use two methods to adjust for the bias. Below, we give the steps for estimating (A.1) and (A.2) using each of these methods.

Hansen and Phillips method

(1) Estimate (A.1) and (A.2) to get the estimates of u_{1t} and u_{2t} and the OLS estimate of γ. For future reference, define the vector of residual estimates as $u_t \equiv [u_{1t}, u_{2t}]'$.

(2) Calculate

$$y_t^+ = Y_t - \Omega_{12}\Omega_{22}^{-1}\Delta x_t,$$

where

$$\Omega \equiv [\Omega_{ij}] = T^{-1}\sum_{t=1}^{T} u_t u_t' + T^{-1}\sum_{r=1}^{l} \omega_{rl} \sum_{t=r+1}^{T} (u_t u_{t-r}' + u_{t-r} u_t'),$$

i.e., Ω is the Newey–West estimator of the variance–covariance matrix with l lags of autocovariances and weights ω.

(3) Calculate the 'bias-adjusted' estimate of γ as

$$\gamma^+ = \left[\sum_{t=1}^{T} y_t^+ x_t - T[I, -\Omega_{12}\Omega_{22}^{-1}][\Lambda_{21}, \Lambda_{22}]'\right]\left[\sum_{t=1}^{T} x_t^2\right]^{-1},$$

where

$$\Lambda \equiv [\Lambda_{ij}] = T^{-1}\left[\sum_{t=1}^{T} u_t u_t' + \sum_{r=1}^{l} \omega_{rl} \sum_{t=r+1}^{T} (u_t u_{t-r}')\right].$$

(4) Calculate a modified Wald statistic, known as the G-statistic, to test $\gamma^+ = 1$. This G-statistic is

$$G_l = (\gamma^+ - 1)^2 \left[\Omega_{11.2} \otimes \left(\sum_{t=1}^{T} x_t^2\right)^{-1}\right]^{-1} \sim \chi^2_{(1)},$$

where

$$\Omega_{11.2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}.$$

Note that subscript l refers to the number of lags included in the estimators Ω and Λ.

Stock and Watson procedure

Rewrite eq. (A.1) according to

$$y_t = \gamma_0 + \gamma x_t + \beta(L)\Delta x_t + u_{1t}, \tag{A.3}$$

where $\beta(L)$ is a polynomial in the lag operator, L, i.e., $\beta(L) = (L^n + L^{n-1} + \dots + L + 1 + L^{-1} + \dots + L^{-n+1} + L^{-n})$. The idea to rewriting (A.1) in this form is to include as many of leads and lags of Δx_t on the right-hand side of the equation to make u_{1t} independent of x. This implies that the asymptotic distribution of the OLS estimator of γ is normal. Intuitively, including the leads and lags of Δx_t on the right-hand side gets rid of the simultaneous equation bias problem. Note that since u_{1t} will be serially correlated in general, the Wald test of $\gamma = 1$ from (A.3) should also use the Newey–West estimator for the covariance matrix. The results reported in the text are not sensitive to these choices, however.

Appendix 2: Description of Monte Carlo experiments in tables 2 and 3

The Monte Carlo p-values reported in tables 2 and 3 we calculated as follows:

(1) We estimated the Stock–Watson version of eqs. (8) and (13), saving the residuals and the coefficient estimates. As discussed in Stock and Watson (1989), the residuals from this equation are independent of the entire sequence of the right-hand-side variable, and so can be treated as strictly exogenous. In the case of (13), the residuals contain overlapping forecast errors, and so it is necessary to remove the induced serial correlation before sampling. For this purpose we estimated AR(6) models for the residuals and used the estimated errors in the sampling procedure described in (2) below.

(2) In the experiments that assumed conditional homoscedasticity, we then drew randomly from the distribution of residuals in the data. In the experiments that allowed for conditional heteroscedasticity we first estimated an ARCH process with residuals.[24] We then scaled the residuals by the predictions of the ARCH model and drew randomly from the scaled distribution. Finally, we rescaled the distribution of residuals using predictions from the ARCH process and used these to generate the time series process of rates.

[24]In estimating the ARCH process, we used a simple rule of thumb to be consistent across maturities. We first estimated an ARCH process of order six and checked to make sure that the implied variances were positive. In the few instances where negative variances were encountered, we reduced the order of the ARCH process until variances were always positive.

(3) In table 2 we took the coefficient estimates for α_0 (and the Stock–Watson coefficients), set $\alpha_1 = 1$, and generated a time series for R_{t+1}^{k-1} equal to the number of observations, 294. In table 3 we took the coefficient estimates for β_0 (and the Stock–Watson coefficients), set $\beta_1 = 1$, and generated a time series for R_{t+k}^1. In this instance we used the estimates of our AR(6) model to generate a new set of serially correlated residuals (due to the forecast overlap) from which we then generated a new time series for R_{t+k}^1.

(4) Using the generated time series for rates on the left-hand side, we estimated the cointegrating regressions (8) and (13) and saved the estimates of α_1 and β_1.

(5) We then repeated steps (2) through (4) 1000 times to from empirical distributions for the estimates.

(6) The p-values are calculated by comparing the empirical distributions against the estimates in column (1) of each table.

Appendix 3: Derivations of the regression coefficients in section 7

To calculate the expression in (25), we first note that (24) can be rewritten as

$$a_T = 1 - \gamma(k)\frac{\operatorname{cov}_T(\phi_{1,t}, \nabla R_{1,t}^k)}{\operatorname{var}_T(\nabla R_t^k)} - \frac{\operatorname{cov}_T(\phi_t^k - \gamma(k)\phi_{1,t}, \nabla R_{0,t}^k)}{\operatorname{var}_T(\nabla R_t^k)}. \tag{A.4}$$

Eq. (A.4) makes use of the fact that the I(1) and I(0) components of the risk premia and yield spread are uncorrelated with one another. Next we use the framework in (1) and (2), together with the assumption about the factor structure in (23), to write the I(1) component of the yield spread as

$$\nabla R_{1,t}^k = (1/k)\sum_{i=0}^{k-2} \gamma(k-i)\phi_{1,t}. \tag{A.5}$$

If we substitute this expression in the second term on the right of (A.4), and note that $\operatorname{var}_T(\nabla R_t^k) = \operatorname{var}_T(\nabla R_{1,t}^k) + \operatorname{var}_T(\nabla R_{0,t}^k)$, we obtain

$$a_T = 1 - \frac{\dfrac{\gamma(k)}{k}\displaystyle\sum_{i=0}^{k-2} \gamma(k-i)\operatorname{var}_T(\phi_{1,t})}{\operatorname{var}_T\left(\dfrac{1}{k}\displaystyle\sum_{i=0}^{k-2}\gamma(k-i)\phi_{1,t}\right) + \operatorname{var}_T(\nabla R_{0,t}^k)}$$
$$- \frac{\operatorname{cov}_T(\phi_t^k - \gamma(k)\phi_{1,t}, \nabla R_{0,t}^k)}{\operatorname{var}_T(\nabla R_{1,t}^k) + \operatorname{var}_T(\nabla R_{0,t}^k)}. \tag{A.6}$$

Finally, we note that when the risk premia is I(0)-stationary, the estimate of the regression coefficient a_1 is

$$\hat{a}_T = 1 - \frac{\mathrm{cov}_T(\phi_t^k - \gamma(k)\phi_{1,t}, \nabla R_{0,t}^k)}{\mathrm{var}_T(\nabla R_{0,t}^k)}. \tag{A.7}$$

Combining (A.7) with (A.6) and simplifying gives the expression in (25).

To obtain the expression in (27), we substitute for ϕ_{t+i}^{k-i} in (26) using (23),

$$b_T = 1 - \frac{\frac{1}{k}\sum_{i=0}^{k-2} \mathrm{cov}_T(\gamma(k-i)\phi_{1,t}, \nabla R_{1,t}^k)}{\mathrm{var}_T(\nabla R_t^k)}$$

$$- \frac{\frac{1}{k}\sum_{i=0}^{k-2} \mathrm{cov}_T(\mathrm{E}_t(\phi_t^k - \gamma(k-i)\phi_{1,t}, \nabla R_{1,t}^k)}{\mathrm{var}_T(\nabla R_t^k)}. \tag{A.8}$$

As before, we have made use of fact that the I(1) and I(0) components of the risk premia and yield spread are uncorrelated with one another. Also, we note that when the risk premia is I(0)-stationary, the estimate of the regression coefficient b_1 is

$$\hat{b}_T = 1 - \frac{\frac{1}{k}\sum_{i=0}^{k-2} \mathrm{cov}_T(\mathrm{E}_t(\phi_t^k - \gamma(k-i)\phi_{1,t}, \nabla R_{0,t}^k)}{\mathrm{var}_T(\nabla R_{0,t}^k)}. \tag{A.9}$$

Substituting (A.5) in (A.8) and combining the result with (A.9) gives, after some simplification, the expression in (27).

References

Backus, D., A. Gregory, and S. Zin, 1989, Risk premiums in the term structure: Evidence from artificial economies, Journal of Monetary Economics 24, 371–399.

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

Campbell, J.Y. and P. Perron, 1991, Pitfalls and opportunities: What macroeconomists should know about unit roots, in: O.J. Blanchard and S. Fischer, eds., NBER macroeconomics annual (MIT Press, Cambridge, MA) 141–217.

Campbell, J.Y. and R.J. Shiller, 1987, Cointegration and test of present value models, Journal of Political Economy 95, 1062–1088.
Campbell, J.Y. and R.J. Shiller, 1991, Yield spreads and interest rate movements: A bird's eye view, Review of Economic Studies 58, 495–514.
Chesher, A. and I. Jewitt, 1987, The bias of a heteroskedasticity consistent covariance matrix estimator, Econometrica 55, 1217–1222.
Cox, J.C., J.E. Ingersoll, and S.A. Ross, 1981, A reexamination of traditional hypotheses about the term structure of interest rates, Journal of Finance 36, 769–799.
Cox, J.C., J.E. Ingersoll, and S.A. Ross, 1985, A theory of the term structure of interest rates, Econometrica 53, 385–408.
Evans, M.D.D. and K.K. Lewis, 1990, Do risk premia explain it all?, NBER working paper no. 3451, Oct.
Fama, E.F., 1984a, The information in the term structure, Journal of Financial Economics 13, 509–528.
Fama, E.F., 1984b, Term premiums in bond returns, Journal of Financial Economics 13, 529–546.
Hansen, B. and P. Phillips, 1990, Statistical inference in instrumental variables regression with I(1) processes, Review of Economic Studies 57, 99–125.
Heston, S.L., 1990, Testing continuous time models of the term structure, Mimeo. (Yale School of Organization and Management, New Haven, CT).
Lewis, K.K., 1989, Changing beliefs and systematic forecast errors, American Economic Review 79, 621–636.
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.
Mishkin, F., 1988, The information in the term structure: Some further results, Journal of Applied Econometrics 3, 307–314.
Mishkin, F., 1989, Is the Fisher effect for real?, Mimeo. (Columbia University, New York, NY).
Shiller, R.J., 1987, The term structure of interest rates, NBER working paper no. 2341.
Shiller, R.J., J.Y. Campbell, and K.L. Schoenholtz, 1983, Forward rates and future policy: Interpreting the term structure of interest rates, Brookings Papers on Economic Activity, 173–217.
Stambaugh, R.F., 1988, The information in the forward rates: Implications for models of the term structure, Journal of Financial Economics 21, 41–69.
Stock, J. and M. Watson, 1989, A simple MLE of cointegrating vectors in higher order integrated systems, NBER technical working paper no. 83.
Vasicek, O.A., 1977, An equilibrium characterization of the term structure, Journal of Financial Economics 5, 177–188.