ELSEVIER Journal of Financial Economics 44 (1997) 309–348

JOURNAL OF Financial ECONOMICS

On biases in tests of the expectations hypothesis of the term structure of interest rates

Geert Bekaert[a,*], Robert J. Hodrick[b], David A. Marshall[c]

[a] *Graduate School of Business, Stanford University, Stanford, CA 94305, USA, National Bureau of Economic Research, Cambridge, MA 02137, USA*

[b] *Columbia University, New York, NY 10027, USA, National Bureau of Economic Research, Cambridge, MA 02137, USA*

[c] *Federal Reserve Bank of Chicago, Chicago, IL 60604, USA*

Received January 1996; received in revised form October 1996

Abstract

We document extreme bias and dispersion in the small-sample distributions of four standard regression-based tests of the expectations hypothesis of the term structure of interest rates. The biases arise because of the extreme persistence in short interest rates. We derive approximate analytic expressions for the biases under a simple first-order autoregressive data generating process for the short rate. We then conduct Monte Carlo experiments based on a bias-adjusted first-order autoregressive process for the short rate and for a more realistic bias-adjusted VAR-GARCH model incorporating the short rate and three term spreads. Conducting inference with the small-sample distributions of test statistics rather than with their asymptotic distributions provides a more consistent rejection of the expectations hypothesis. Plausible sources of measurement error in short and long yields do not salvage the expectations hypothesis.

Keywords: Interest rates; Expectations hypothesis; Small sample bias; Vector autoregression; Conditional heteroskedasticity

JEL classification: C15; E43; G12

* Corresponding author

We are grateful for the comments of John Campbell (the referee), whose advice substantially improved the paper, as well as Tim Bollerslev, Andrea Buraschi, Bill Schwert (the editor), and seminar participants at Carnegie-Mellon University, Dartmouth College, INSEAD, Georgetown University, New York University, Northwestern University, Ohio State University, Washington University, the University of Amsterdam, the University of California at Santa Cruz, the University of Chicago, the University of Southern California, the 1995 NBER Summer Institute, and the 1996 European Finance Association Meetings in Oslo. We thank Rob Bliss for providing us with his data on bond yields. Geert Bekaert acknowledges financial support from the National Science Foundation and the Financial Research Initiative at Stanford University. We also thank Linda Bethel and Ann Babb for their help with typing the manuscript.

PII S0304-405X(97)00007-X

1. Introduction

The expectations hypothesis is probably the oldest and most studied theory of the term structure of interest rates (see Fisher, 1896; Lutz, 1940). Although the modern finance literature has developed more sophisticated models of the term structure, the empirical evidence against the basic expectations hypothesis is far from conclusive and offers several interesting puzzles.

For example, Campbell and Shiller (1991) find different results with US data depending on the regression specification and the maturity of the bonds. In brief, the change in the US long-term interest rate does not behave as predicted by the theory. The regressions indicate that actual long-term rates move in the opposite direction from that predicted by the theory. The predictions of the expectations hypothesis for long rates are rejected very strongly at the short end of the term structure and quite comfortably at the long end using the traditional asymptotic distribution theory. On the other hand, the regressions indicate that future short-term rates move in the direction predicted by the expectations hypothesis. The theory is still rejected at the short end of the term structure, but this empirical specification does not reject the expectations hypothesis at the long end of the term structure. Campbell and Shiller (1991, p. 505) note that these two sets of results produce an apparent paradox:

> The slope of the term structure almost always gives a forecast in the wrong direction for the short-term change in the yield on the longer bond, but gives a forecast in the right direction for long-term changes in short rates.

Campbell and Shiller use two regression tests involving term spreads and two specification tests from vector autoregressions in changes in short rates and term spreads. A fifth specification test uses forward rates as in Fama (1984). An earlier version of this paper (Bekaert et al., 1996) reports results for this fifth specification test. Because both the analytical biases and the Monte Carlo simulations for this test are similar to those of the second Campbell-Shiller specification test, we focus on the two regression tests and the two VAR statistics.[1]

The purpose of this paper is to reexamine the econometric methodology underlying these different specification tests. Our main contribution is to

[1]The regression test based on forward rates has delivered rejections of the expectations hypothesis at the short end of the maturity spectrum using US data (see Fama, 1984; Fama and Bliss, 1987; Stambaugh, 1988), but Jorion and Mishkin (1991) found no evidence against the expectations hypothesis using longer maturities and data from the United States, the United Kingdom, Germany, and Switzerland. Hardouvelis (1994) also found less evidence against the expectations hypothesis with international data using the Campbell–Shiller (1991) specification tests.

demonstrate that *all* regression-based tests of the expectations hypothesis are severely biased in small samples. We show that the high persistence of short-term interest rates induces extreme bias and extreme dispersion into the small-sample distributions of the test statistics. Intuitively, the well-known downward bias in estimating autocorrelations (see Marriott and Pope, 1954; Kendall, 1954) translates into a large upward bias in the slope coefficients of standard tests of the expectation hypothesis because the dependent variables depend on future short rates and the regressors depend negatively on current short rates. This bias is an example of the biases in regressions on persistent, predetermined, but not necessarily exogenous regressors studied by Stambaugh (1986), Mankiw and Shapiro (1986), and Elliott and Stock (1994).

The bias makes it important to use well-designed Monte Carlo simulations to derive the small-sample distributions of test statistics. Evaluating the econometric analysis of the expectations hypothesis using the small-sample distributions may strengthen rejections (because of the positive bias) or weaken rejections (because of the increased dispersion). While Campbell and Shiller (1991) and others have used Monte Carlo methods to assess the validity of their asymptotic distribution theory and have reported that the asymptotic theory is not to be trusted, the typical Monte Carlo experiment has not adjusted for the small-sample bias in the coefficients that are estimated to form the data-generating process.

The organization of the paper is as follows. Section 2 examines some empirical evidence on the expectations hypothesis using the four statistical tests mentioned above. Section 3 analytically derives first-order approximations to the small-sample biases for these specification tests assuming a first-order autoregressive model for the short rate. Section 4 examines Monte Carlo evidence on the four tests under the first-order autoregression data-generating process. Section 5 considers Monte Carlo simulations for a more realistic data-generating process for the short rate and the term spreads. We model the short rate as part of a vector autoregression that includes three spreads between longer maturity rates and the short rate and that has conditionally heteroskedastic innovations. Our conditional volatility model combines a factor GARCH structure with a square-root process. In Section 6 we discuss whether measurement error explains departures from the expectations hypothesis. The last section provides some concluding remarks.

2. An update of empirical evidence on the expectations hypothesis

We follow Campbell and Shiller (1991) in defining the expectations hypothesis of the term structure as the requirement that continuously compounded long-term interest rates (the yields on long-term pure discount bonds) be weighted averages of expected future values of continuously compounded short interest rates, possibly with an additive time-invariant term premium. Formally,

$$r(t, n) = \frac{1}{n} \sum_{i=0}^{n-1} \mathrm{E}_t\, r(t + i) + \kappa_n, \tag{1}$$

where $r(t, n)$ denotes the continuously compounded annualized yield on a bond with n periods to maturity at time t, $r(t)$ denotes the one-period short rate, and κ_n is a constant term premium.

A number of tests of Eq. (1) have been proposed in the literature. First, as noted by Campbell and Shiller (1991), Eq. (1) implies that a maturity-specific multiple of the term spread, $r(t, n) - r(t)$, predicts future changes in the long bond yield. In particular, the slope coefficient, α_1, should equal unity in the following regression:

$$r(t + 1, n - 1) - r(t, n) = \alpha_0 + \alpha_1 \frac{1}{(n - 1)} [r(t, n) - r(t)] + \varepsilon(t + 1). \tag{2}$$

Second, Eq. (1) implies that the current term spread should forecast a weighted average of future changes in short interest rates. Campbell and Shiller (1991) note that the slope coefficient, δ_1, should equal unity in the following regression:

$$\sum_{i=1}^{n-1} \left[1 - \frac{i}{n}\right] [r(t + i) - r(t + i - 1)] = \delta_0 + \delta_1 [r(t, n) - r(t)] + \varepsilon(t + n - 1). \tag{3}$$

The second and third columns of Table 1 report evidence on Eqs. (2) and (3) for the period May 1952 to December 1995, a total of 524 monthly observations. The data from May 1952 to February 1991 are from McCulloch and Kwon (1993), and the data from March 1991 to December 1995 are from Robert Bliss, as discussed in Bliss (1994). These former data are computed using the cubic-spline procedure of McCulloch (1975) and McCulloch and Kwon (1993). Bliss (1994) does not adjust the yields for tax effects, while the McCulloch-Kwon (1993) data are tax-adjusted. After the late 1980s, the tax adjustments are extremely small or zero. In Table 1, the short rate has a 1-month maturity.

Notice that the estimated slope coefficients for Eq. (2) are significantly below unity for all maturities, and the point estimates are negative. Furthermore, the point estimates become more negative as yields of longer-term bonds are used to form the dependent variable and the term spread. In contrast, the point estimates of the slope coefficients in Eq. (3) are all positive, and as the horizon

Table 1
Estimates and asymptotic standard errors

The sample period is May 1952 to December 1995 (524 observations) for Eq. (2) and the VARs, with commensurately fewer observations at the ends of the samples for Eq. (3). The column labeled Eq. (2) reports the slope coefficients from regressions of the change in the yield on an n-period bond on $[1/(n-1)]$ times the term spread between the n-period yield and the short rate. The column labeled Eq. (3) reports the slope coefficients from regressions of the weighted average of changes in future short rates on the term spread. The last two columns report statistics based on a fourth-order bivariate VAR in the change in the short rate and the n-period term spread. The two statistics are the correlation between the theoretical spread that satisfies the expectations hypothesis and the actual spread and the ratio of the standard deviation of the theoretical spread to the standard deviation of the actual spread. Asymptotic standard errors are in parentheses, and the p-values associated with the hypothesis that the statistics equal one are in square brackets.

n (months)	Eq. (2)	Eq. (3)	Correlation	Standard dev. ratio
12	−0.819	0.329	0.759	0.404
	(0.620)	(0.174)	(0.236)	(0.129)
	[0.003]	[0.0001]	[0.307]	[0.00004]
36	−1.673	0.440	0.913	0.373
	(1.255)	(0.271)	(0.164)	(0.245)
	[0.033]	[0.039]	[0.596]	[0.010]
60	−2.320	0.569	0.960	0.407
	(1.479)	(0.324)	(0.080)	(0.272)
	[0.025]	[0.183]	[0.617]	[0.029]

increases, the estimated slope coefficients become insignificantly different from unity.[2]

In addition to the single-equation regression tests in Eqs. (2) and (3), Campbell and Shiller derive tests of Eq. (1) based on a bivariate vector autoregression (VAR) in the change in the short rate, $\Delta r(t) \equiv r(t) - r(t-1)$, and the term spread, $s(t,n) \equiv r(t,n) - r(t)$. To understand the VAR statistics, let A denote the first-order companion form of the VAR parameter matrix and let $w(t)$ denote the corresponding companion-form vector of regressors, $w(t) = [\Delta r(t), s(t,n), \Delta r(t-1), s(t-1,n), \ldots, \Delta r(t-j), s(t-j,n)]'$, where the lag length of the VAR is $j+1$. From Eq. (1), the term spread that satisfies the expectations hypothesis is

[2]At this point in the paper, we are defining the concept of statistical significance relative to the asymptotic distributions of the OLS estimators with conditionally heteroskedastic standard errors as in Newey and West (1987). As we shall see in Sections 3 and 5 below, the small-sample distributions of these estimators differ substantially from their asymptotic counterparts, which creates severe distortions in the sizes of the tests. Inference with the asymptotic distribution is consequently untrustworthy.

$\sum_{i=1}^{n-1} [1 - (i/n)]\mathrm{E}_t[\Delta r(t+i)]$. Campbell and Shiller note that the VAR produces expected changes in the short rate from which one can obtain an expression for the time-varying part of 'theoretical spread', denoted $s'(t, n)$:

$$s'(t, n) \equiv \sum_{i=1}^{n-1} (1 - i/n)\, e1' A^i\, w(t), \tag{4}$$

where $e1$ is an indicator vector with one in the first row and zeros everywhere else.

The statistics proposed by Campbell and Shiller as tests of the expectations hypothesis are the correlation between $s'(t, n)$ and $s(t, n)$ and the ratio of the standard deviation of $s'(t, n)$ to the standard deviation of $s(t, n)$. Both statistics are functions of the coefficients of the VAR and the covariance matrix of the VAR innovations. Under the expectations hypothesis, both should equal one. The columns of Table 1 labeled 'Correlation' and 'Standard Dev. Ratio' report these statistics for our sample period. Notice that the asymptotic distribution theory underlying the correlation statistic does not provide evidence against the null hypothesis, whereas the ratio of the standard deviations provides quite strong evidence against the expectations hypothesis. For the longer maturities, the correlation statistic is quite close to its theoretical value of unity, whereas the ratio of the standard deviations is always less than 0.5.

3. Analytical approximations of the biases in tests of the expectation hypothesis

In this section, we derive first-order approximations to the small-sample biases for the four specification tests of the expectations hypothesis under the following AR(1) model for the short rate:

$$r(t+1) = \mu + \rho r(t) + v(t+1). \tag{5}$$

We choose the AR(1) for analytical tractability and to clearly illustrate the link between the small-sample biases and the degree of persistence in the short-rate process. Although a highly serially correlated AR(1) model is a reasonable approximation of the monthly short-rate data, the implications of the AR(1) model for term spreads are somewhat counterfactual.[3] Therefore, we also present results in Section 5 for an alternative data-generating process based on a conditionally heteroskedastic VAR.

[3]An unpublished paper by Tauchen (1985) analogously uses an AR(1) process for the spot exchange rate to investigate the bias in regression tests of the expectations hypothesis that the forward premium is an unbiased predictor of the ex post rate of depreciation.

Table 2 reports the ordinary least squares (OLS) estimates of μ, ρ, and σ (the standard deviation of $v(t+1)$) in Eq. (5), denoted by $\hat{\mu}$, $\hat{\rho}$, and, $\hat{\sigma}$, respectively. These estimates are reported for the sample corresponding to 524 observations in Table 1. We also report bias-adjusted values of μ, ρ, and σ. Kendall (1954) shows that, to a first-order approximation,

$$\mathrm{E}(\hat{\rho}) - \rho = -\frac{1+3\rho}{T} + \mathrm{O}\left(\frac{1}{T^2}\right), \tag{6}$$

where T denotes the sample size. The bias adjustment in Table 2 unwinds the bias in Eq. (6) such that the 'bias-adjusted ρ' is $(\hat{\rho} + (1/T))/(1 - (3/T)) = 0.9840$. Thus, the bias equals -0.0075 for the 524 observations and the estimated ρ. The bias-adjusted μ and bias-adjusted σ modify $\hat{\mu}$ and $\hat{\sigma}$ to insure that the unconditional mean and standard deviation of $r(t)$ remain unchanged by the bias adjustment in ρ.

We now derive first-order approximations for the four specification tests of the expectations hypothesis. (The proofs of the propositions are given in the appendix.)

Proposition 1. Under Eqs. (1) *and* (5), *the expected value of the slope coefficient of the first specification test in Eq.* (2) *is*

$$\mathrm{E}(\hat{\alpha}_1) = 1 + \frac{-n(1-\rho^{n-1})}{n(1-\rho)-(1-\rho^n)}\,[\mathrm{E}(\hat{\rho}) - \rho]\,. \tag{7}$$

Denote the coefficient multiplying the bias in $\hat{\rho}$ in Eq. (7) by $\phi(\rho, n)$ and the bias in $\hat{\rho}$ by θ_1. Note that both $\phi(\rho, n)$ and θ_1 are negative. Therefore, the

Table 2
A first-order autoregressive model for the short interest rate

The sample is monthly data from May 1952 to December 1995 (524 observations). The OLS estimates are given with a hat, and 'bias-adj.' denotes the values of the parameters after adjusting for small-sample bias. For ρ, the bias adjustment is $\rho = (\hat{\rho} + (1/T))/(1-(3/T))$. For μ, the bias adjustment is $\mu = \hat{\mu}(1-\rho)/(1-\hat{\rho})$. For σ, the bias adjustment is $\sigma = \hat{\sigma}\,[(1-\rho^2)/(1-\hat{\rho}^2)]^{0.5}$.

$r(t) = \mu + \rho r(t-1) + \sigma u(t)$

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

$\hat{\rho}$	0.9765
Bias-adj. ρ	0.9840
$\hat{\mu}$	0.1285
Bias-adj. μ	0.0875
$\hat{\sigma}$	0.6115
Bias-adj. σ	0.5055

estimate of the slope coefficient $\hat{\alpha}_1$ is biased upward. Note also that $\phi(\rho, 2) = -2/(1-\rho)$ and $\phi(\rho, \infty) = -1/(1-\rho)$. While the bias decreases as the horizon n increases, the bias is still substantial for large values of n, especially if ρ is near unity. For example, at the five-year maturity ($n = 60$), $\phi(0.9840, 60) = -108.93$, which combined with the estimated value of the bias in $\hat{\rho}$ of -0.0075 produces an expected value of $\hat{\alpha}_1$ of 1.817, a value substantially larger than the asymptotic value of unity. Analytical values for the expected values of the slope coefficients for other values of n are reported below in column 2 of panel A in Table 3.

When n is large, data on bonds of slightly different maturities are often unavailable. As a result, researchers modify Eq. (2) by using a constant maturity in the regressand, $r(t+1, n) - r(t, n)$. While this approximation may seem relatively innocuous, it introduces an approximation error into the regression in addition to the small-sample bias. Under our data-generating process in Eq. (5), we can analytically determine the size of this approximation error.

Proposition 2. Under Eqs. (1) *and* (5), *the expected value of the slope coefficient of the first specification test using the approximation for the regressand described above is*

$$\mathrm{E}(\hat{\alpha}_1) = 1 + c(\rho, n) + \phi_2(\rho, n)\theta_1 , \tag{8}$$

where

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

and

$$\phi_2(\rho, n) = \frac{-(n-1)(1-\rho^n)}{n(1-\rho) - (1-\rho^n)} . \tag{9}$$

The bias term $\phi_2(\rho, n)$ is quite similar to $\phi(\rho, n)$, derived above in Proposition 1. The approximation error term $c(\rho, n)$ is positive, can differ substantially from zero, and most importantly, does not change as the sample size increases. For $n = 2$, $c(\rho, n) = \rho$. While the approximation error becomes smaller for large values of n, it is still substantial for maturities often used in empirical work. Even with maturities as high as five years ($n = 60$), $c(0.9840, 60) = 0.722$, which makes the expected value of $\hat{\alpha}_1$ equal to 2.534 for the 524 observations, rather than the population value of unity. The analytical biases in this specification test are reported in column 2 of panel B in Table 3.

The second specification test, Eq. (3), is also subject to severe small-sample bias.

Table 3

Monte Carlo distributions of the slope coefficients and VAR-based statistics

The Monte Carlo evidence is based on 5000 replications. The data-generating process is an AR(1) model for the short rate:

$$r(t+1) = 0.0875 + 0.9840r(t) + 0.5055u(t+1),$$

where $u(t)$ is a N(0, 1) variable and the parameters are the bias-adjusted parameters from Table 2. The sample size is 524, and the horizon is n months. The column labeled 'Analytical Estimate' contains the expected values of the distributions predicted by the analytical derivations in Section 3. The columns labeled Mean, σ, 1%, 5%, and 10% are the sample mean, the standard deviation, and the respective quantiles of the empirical distributions. The panels correspond to five different tests. Eq. (2) reports the slope coefficients from regressions of the change in the yield on an n-period bond on $[1/(n-1)]$ times the term spread between the n-period yield and the short rate. In panel B, the same regression is run but the $(n-1)$-period yield at time $t+1$ is approximated by the n-period yield at time $t+1$. Eq. (3) reports the slope coefficients from regressions of the weighted average of changes in future short rates on the term spread. Panels D and E report statistics based on a first-order bivariate VAR in the change in the short rate and the n-period term spread. The two statistics are the correlation between the theoretical spread that satisfies the expectations hypothesis and the actual spread and the ratio of the standard deviation of the theoretical spread to the standard deviation of the actual spread.

n	Analytical estimate	Mean	σ	1%	5%	10%
Panel A: Eq. (2)						
12	1.918	2.018	1.355	− 0.069	0.265	0.531
36	1.864	1.958	1.276	− 0.006	0.308	0.559
60	1.817	1.906	1.206	0.049	0.346	0.583
Panel B: Eq. (2) (with approximation error)						
12	2.844	2.942	1.345	0.872	1.203	1.467
36	2.678	2.771	1.266	0.821	1.133	1.382
60	2.534	2.622	1.198	0.777	1.072	1.307
Panel C: Eq. (3)						
12	1.469	1.506	0.648	0.454	0.646	0.781
36	1.463	1.469	0.545	0.448	0.690	0.820
60	1.461	1.440	0.464	0.477	0.717	0.857
Panel D: VAR statistics (order = 1) correlation coefficient						
12	1.0	0.998	0.004	0.981	0.991	0.994
36	1.0	0.9998	0.0004	0.998	0.9991	0.9994
60	1.0	0.9999	0.0001	0.9994	0.9997	0.9998
Panel E: VAR statistics (order = 1) standard deviation ratio						
12	1.225	1.475	0.616	0.456	0.639	0.773
36	1.225	1.355	0.457	0.480	0.665	0.792
60	1.225	1.273	0.357	0.509	0.692	0.812

Proposition 3. Under Eqs. (1) *and* (5), *the expected value of the slope coefficient of the second specification test in Eq.* (3) *is*

$$\mathrm{E}[\hat{\delta}_1] = 1 + \frac{-(1-\rho)}{n(1-\rho) - (1-\rho^n)} \sum_{j=1}^{n-1} \theta_j , \tag{10}$$

where θ_j *denotes the small-sample bias in the OLS estimate of* ρ^j.

Using results from Kendall (1954, Eq. (20)), we derive the following first-order approximation to θ_j:

$$\theta_j \approx -\left[\frac{(1+\rho)}{(1-\rho)}(1-\rho^j) + 2j\rho^j\right]\left[\frac{1}{T-(n-1)}\right]. \tag{11}$$

For $n = 2$, the bias is $-\theta_1/(1-\rho)$, which is one-half of the small-sample bias in Eq. (2). For longer horizons, additional bias terms must be added. In column 2 of panel C in Table 3, we report biases for this specification. Interestingly, they are substantially smaller than for specification (2). Of course, this specification implies the loss of $n-1$ observations relative to the first specification, which will be reflected in the Monte Carlo results in the next section. Finally, specification (3) implies that the error term $\varepsilon(t+n-1)$ is a moving average process of order $n-1$. Several studies (Hodrick, 1992; Richardson and Stock, 1989) have noted the poor small-sample properties of test statistics based on kernel estimators of the asymptotic variance of the OLS estimators in analogous situations.[4]

We now turn to biases in the two VAR-based statistics under the maintained assumption that the short rate is generated by Eq. (5). We derive analytical approximations of the biases for a first-order VAR. In the Monte Carlo experiments below, we show results for both a first-order and a fourth-order VAR, the latter being the order used in Campbell and Shiller and in Table 1. To simplify notation, let $\eta(\rho, n) \equiv (1/n)(1-\rho^n)/(1-\rho) - 1$. Under the expectations hypothesis and the data-generating process of Eq. (5), the term spread is $s(t, n) = \eta(\rho, n) r(t)$. The first-order VAR can then be written:

$$\begin{pmatrix} \Delta r(t+1) \\ \eta(\rho,n) r(t+1) \end{pmatrix} = A_0 + A \begin{pmatrix} \Delta r(t) \\ \eta(\rho, n) r(t) \end{pmatrix} + e(t+1). \tag{12}$$

Let $\hat{A}$ denote the OLS estimator of A. It follows from Eqs. (5) and (12) that

$$\operatorname{plim}(\hat{A}) = \begin{bmatrix} 0 & \dfrac{\rho-1}{\eta(\rho,n)} \\ 0 & \rho \end{bmatrix}. \tag{13}$$

[4] To conserve space, we do not report results for standard errors or t-statistics. The interested reader is referred to Bekaert, Hodrick, and Marshall (1996) for evidence that the t-statistics for Eq. (3) are much more severely biased than those for Eq. (2).

We can write the theoretical spread as follows:

$$s'(t, n) = x_n \Delta r(t) + y_n s(t, n), \tag{14}$$

where x_n and y_n are the implied coefficients from Eq. (4) and are functions of the estimated A parameters. When x_n and y_n are evaluated using $\operatorname{plim}(\hat{A})$, $x_n = 0$ and $y_n = 1$. Hence, in an infinite sample the correlation of the theoretical spread and the actual spread would be one, and the ratio of their standard deviations would also equal one. In finite samples, however, x_n and y_n are biased. The biases in x_n and y_n are functions of the bias matrix of the VAR coefficients, which we denote E[B]. While both of the VAR-based statistics could inherit a bias from the biases in x_n and y_n, it turns out that only the standard deviation ratio is biased.

Proposition 4. To a first-order approximation, the correlation of the theoretical spread $s'(t, n)$ and the actual spread $s(t, n)$ is unbiased in small samples, but the ratio of the standard deviation of the theoretical spread to the standard deviation of the actual spread is biased:

$$\mathrm{E}\left\{\left[\frac{\operatorname{var}(s'(t, n))}{\operatorname{var}(s(t, n))}\right]^{1/2}\right\} \approx 1 + \frac{(1-\rho)}{\eta(\rho, n)}(x_n \text{ bias}) + (y_n \text{ bias}). \tag{15}$$

To a first-order approximation, the biases in x_n and y_n are

$$(x_n \text{ bias}, y_n \text{ bias}) \approx \frac{e1'}{n}\sum_{j=1}^{n-1}(n-j) \times\left\{A^j + \sum_{i=1}^{2}\sum_{k=1}^{2}\left[\sum_{l=0}^{j-1} A^l J_{i,k} A^{j-l-1}\right]\mathrm{E}[B_{i,k}]\right\} - (0, 1), \tag{16}$$

where $J_{i,k}$ is the indicator matrix with one in the i–k position and zeros everywhere else, and $\mathrm{E}[B_{i,k}]$ *denotes the bias in the* (i, k)th *element of* $\hat{A}$.

To implement Proposition 4, we need to determine the bias matrix E[B] of the VAR coefficients. It is extremely difficult to derive a first-order approximation to this bias matrix when A_0 (the constant term in the VAR) is unknown. For this reason, we present a first-order approximation to E[B] for the simpler case in which the unconditional means of the variables in the VAR are known. (Without loss of generality, these means can then be set equal to zero, since the VAR can be estimated with 'de-meaned' data.) It is likely that this approximation understates the magnitude of the bias. In the univariate case of Eq. (5), Kendall (1954) shows that the first-order approximation to the bias in $\hat{\rho}$ is

$-(1+3\rho)/T$ when μ is unknown, but only $-2\rho/T$ when μ is known. In the next section, we will compare these first-order approximations to the average values of our Monte Carlo estimates.

The VAR coefficients are nonlinear functions of the following random variables:

$$Q \equiv \frac{\sum_{t=1}^{T} v(t+1)r(t)}{\sum_{t=1}^{T} r(t)^2} \tag{17}$$

and

$$V \equiv \frac{\sum_{t=1}^{T} v(t+1)r(t-1)}{\sum_{t=1}^{T} r(t-1)^2}. \tag{18}$$

The OLS estimator for the VAR parameters for a sample of size T can be written as

$$\hat{A} = \text{plim}(\hat{A}) + B \equiv \text{plim}(\hat{A}) + \begin{bmatrix} \dfrac{Q(Q+\rho)-V}{1-(Q+\rho)^2} & \dfrac{Q+V}{\eta(\rho,n)(1+Q+\rho)} \\ \eta(\rho,n)\dfrac{Q(Q+\rho)-V}{1-(Q+\rho)^2} & \dfrac{Q+V}{1+Q+\rho} \end{bmatrix}. \tag{19}$$

The bias matrix $\mathrm{E}[B]$ is a nonlinear function of the moments of Q and V:

$\bar{Q} \equiv \mathrm{E}[Q]$, $\bar{V} \equiv \mathrm{E}[V]$, $\mathrm{var}[Q]$, and $\mathrm{cov}[Q, V]$.

Proposition 5. To a second-order approximation, the elements of the bias matrix $\mathrm{E}[B]$ *in Eq.* (19) *are given by*

$$\begin{aligned} \mathrm{E}(B_{11}) = {} & \frac{\bar{Q}^2+\rho\bar{Q}-\bar{V}}{1-(\bar{Q}+\rho)^2} + \frac{1}{1-(\bar{Q}+\rho)^2}\left[1+\frac{2(2\bar{Q}+\rho)(\bar{Q}+\rho)}{1-(\bar{Q}+\rho)^2}\right. \\ & \left. + \frac{(\bar{Q}^2+\rho\bar{Q}-\bar{V})(1+3(\bar{Q}+\rho)^2)}{(1-(\bar{Q}+\rho)^2)^2}\right]\mathrm{var}[Q] \\ & - \frac{2(\bar{Q}+\rho)}{(1-(\bar{Q}+\rho)^2)^2}\,\mathrm{cov}[Q,V] \end{aligned} \tag{20}$$

$$\begin{aligned} \mathrm{E}(B_{22}) = {} & \frac{\bar{Q}+\bar{V}}{1+\bar{Q}+\rho} - \frac{1}{(1+\bar{Q}+\rho)^2}\left[1-\frac{(\bar{Q}+\bar{V})}{(1+\bar{Q}+\rho)}\right]\mathrm{var}(Q) \\ & - \frac{1}{(1+\bar{Q}+\rho)^2}\,\mathrm{cov}(Q,V), \end{aligned} \tag{21}$$

$\mathrm{E}(B_{21}) = \eta(\rho,n)\,\mathrm{E}(B_{11})$, *and* $\mathrm{E}(B_{12}) = \mathrm{E}(B_{22})/\eta(\rho,n)$.

We report analytical values for the biases in the VAR statistics in column 2 of panels E and F in Table 3. The moments of Q and V are derived using 200,000 Monte Carlo experiments. We use Monte Carlo approximations to the moments of Q and V because analytic approximations to var$[Q]$ and cov$[Q, V]$ are extremely difficult to derive. Kendall (1954) shows that $\mathrm{E}[Q] \approx -2\rho/T$, and $\mathrm{E}[V] \approx -2\rho^2/T$. Our Monte Carlo estimates of these means are extremely close to the Kendall approximations.

4. Monte carlo results for an AR(1) data-generating process

Kendall (1954) notes that derivations of first-order bias may be of doubtful validity for values of ρ near unity. This intuition has been confirmed by numerous papers on unit roots (see Stock, 1994, for a recent survey of the unit root literature). Kendall also notes that the distribution of $\hat{\rho}$ is highly skewed, so that use of the expected value as a criterion of bias is itself open to question. Moreover, the derivations of the biases in the VAR statistics involve some approximations. For all of these reasons, it is important to examine Monte Carlo evidence regarding the slope coefficients in the regression tests and the VAR statistics. In the interest of brevity, we report only the mean, the standard deviation, and the left-tail behavior of the small-sample distributions for the slope coefficients. Table 3 reports results for the 524-observation case coinciding with our sample.

The results of the Monte Carlo experiments in Table 3 support the accuracy of the theoretical bias calculations for the specifications corresponding to Eqs. (2) and (3). For Eq. (2), the theoretical estimates are all 95% of the mean values of the Monte Carlo experiments. For Eq. (2) with the approximation error, the theoretical estimates are all 97% of the means of the Monte Carlo experiments. For Eq. (3), the theoretical estimates are between 97.5% and 101.5% of the means of the Monte Carlo experiments.

The Monte Carlo analysis confirms the pitfalls in using the approximation $n \approx n - 1$ for the long-term bond without recognizing that the specification is biased. In fact, at $n = 60$, the coefficient must only be less than 1.072 to have a 5% rejection in a one-tailed test.

The two VAR statistics (panels D and E of Table 3) have different small-sample properties. The correlation statistic has virtually no bias as is predicted by the analytical results. The distribution of the statistic is also very tight. Of course, this statistic may have low power. The theoretical estimate of the ratio of the standard deviations performs somewhat less well. For $n = 12$, the value of 1.225 is only 83% of the mean value of the Monte Carlo experiments. The underestimate was anticipated since the theoretical estimate assumes that the means of the variables in the VAR are known. For larger values of n, the approximation is more accurate. For $n = 60$, the analytical estimate is 96% of the mean value of the Monte Carlo experiments.

5. Analysis with an alternative data-generating process

5.1. The VAR-GARCH model

The data-generating process of Section 4 is unrealistic along two important dimensions. First, short rates are conditionally heteroskedastic (see Gray, 1996, for a recent discussion). Incorporating heteroskedasticity in the data-generating process typically implies a highly leptokurtic unconditional distribution for short rates. Because quite large samples are required for conventional asymptotic distribution theory to work well with leptokurtic distributions (Bollerslev et al., 1992), this feature of the data may worsen the bias and increase the dispersion of the small-sample distributions.

Second, the autoregressive data-generating process implies that the term spread is a fixed multiple of the short rate. Consequently, the spread is perfectly correlated with the short rate and inherits its persistence.[5] The analysis in Stambaugh (1986) suggests that our results may be sensitive to this stochastic singularity in our data-generating process. Stambaugh shows that the bias in the slope coefficient of a regression of a dependent variable (such as the change in the long rate) on an AR(1), predetermined regressor (such as the term spread) depends on the persistence of the regressor and the correlation between the innovations of the original regression and the innovations in the AR(1) process for the regressor. With an AR(1) data-generating process for the short rate, this latter correlation is exactly minus one, which exaggerates the biases compared to a model with imperfect correlation between spreads and the short rate. Similarly, by counterfactually making the persistence of the term spread as large as the persistence of the short rate, the autoregressive data-generating process may further exaggerate the biases.

The purpose of this section is to determine the robustness of our analytical results to a more realistic data-generating process. We begin by modeling the short rate and the term spreads using a second-order vector autoregressive process with conditionally heteroskedastic innovations. Since the expectations hypothesis implies that term spreads are the optimal predictors of future short rate changes, we introduce three term spreads in the VAR to serve as information variables forecasting future short rates, in addition to the lagged short rate.

[5] We thank the referee for pointing out the potential importance of these implications of the AR(1) process.

At this point, we do not impose the expectations hypothesis (see below for further discussion). Formally, let $y(t) = [r(t),\ s(t,12),\ s(t,36),\ s(t,60)]'$, and assume that a pth order VAR adequately describes the four series:

$$y(t) = \mu + \sum_{j=1}^{p} C_j\, y(t-j) + \varepsilon(t)\,. \tag{22}$$

A second-order VAR is chosen by the Schwarz (1978) selection criterion. Furthermore, we model the innovation vector as a factor structure with the innovations of the short rate and the 60-month term spread as the factors. That is,

$$\varepsilon(t) = Fe(t) \tag{23}$$

with

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 \\ f_{21} & 1 & 0 & f_{24} \\ f_{31} & 0 & 1 & f_{34} \\ f_{41} & 0 & 0 & 1 \end{bmatrix}. \tag{24}$$

In Eq. (23) the vector $e(t)$ represents the idiosyncratic innovations. Hence, $\mathrm{E}[e(t)\,e(t)' \mid \mathrm{I}(t-1)] = H(t)$, and $H(t)$ is a diagonal matrix. As a result, the conditional covariance matrix of the innovations $\varepsilon(t)$, denoted by $\Omega(t)$, can be written as $\Omega(t) = FH(t)F'$. We assume that each diagonal element in $H(t)$ follows a GARCH(1,1) process (see Bollerslev, 1986) augmented with a square-root process as in the univariate model of Gray (1996). The square-root process helps to accommodate the dramatic shift in short-rate volatility during the monetary targeting period of 1979–1982:

$$h(t,i) = \gamma_i \sqrt{r(t-1)} + \alpha_i\, e(t-1,i)^2 + \beta_i\, h(t-1,i), \quad i = 1,2,3,4. \tag{25}$$

In this model, the conditional variances of the 12-month and 36-month term spreads have three components: a component linear in the conditional variance of the short rate, a component linear in the conditional variance of the long-term spread, and an idiosyncratic component. The model nests a one-factor model in which the short rate is the only factor driving the conditional volatility of $y(t)$. However, the one-factor model is firmly rejected in favor of the two-factor model using a likelihood ratio test. Moreover, the idiosyncratic components of the residuals of the spreads from the one-factor model remain very highly correlated. The two-factor model still implies that higher variability of the short rate leads to higher variability of all term spreads. Compared to other multivariate GARCH models (see, for example, Kroner and Ng, 1995), the model is very parsimonious with only 17 parameters. This parsimony is achieved by restricting the covariance matrix so that its depends only on the conditional variances of the short rate and the long-term spread. The model is therefore similar to the factor GARCH models of Bekaert and Harvey (1997) and Engle et al. (1990).

To estimate the model in Eqs. (22)–(25), we first exploit the block-diagonality of the information matrix to obtain estimates of the VAR parameters, μ, C_1, and C_2, using ordinary least squares. We then estimate the multivariate GARCH model from the VAR residuals, using quasi-maximum likelihood. Hence, we assume normal innovations to construct the likelihood function although the true distribution of the innovations may not be normal. White (1982) and Bollerslev and Woolridge (1992) show that the resulting estimator is consistent and asymptotically normal.

One problem with the above approach is that the estimated VAR parameters are likely to be biased in small samples. Therefore, prior to generating the residuals, we correct for bias in the VAR parameters using the following Monte Carlo methodology. We estimate an unconditional covariance matrix for the innovations based on the OLS point estimates. We generate new data for the short rate and the spreads by assuming that the innovations of the VAR are normally distributed using the estimated covariance matrix. In each experiment we generate 626 values of each variable using the unconditional means as starting values, and we discard the first 100 values before rerunning the second-order VAR with 524 observations on the dependent variables. The differences between the OLS point estimates and the averages of the OLS point estimates for 200,000 experiments are then added to the original OLS point estimates to get the bias-adjusted values. The GARCH estimation uses the resulting residuals.

We checked the validity of our bias-adjustment procedure by estimating an additional 200,000 VARs using the bias-adjusted parameters as the data-generating process. The means of these Monte Carlo experiments reproduce the original OLS estimates up to the third decimal place.

The final step of the data-generating process involves obtaining data that satisfy the expectations hypothesis by construction. As noted above, the expectations hypothesis imposes restrictions on the VAR parameters. Let A denote the first-order companion form of the (bias-corrected) VAR parameter matrix, and let $z(t)$ be the stacked eight-by-one companion form for $y(t)$ and $y(t-1)$. Then, the 'theoretical spreads' that satisfy the expectations hypothesis are given by

$$s'(t, n) = \frac{1}{n} \sum_{i=0}^{n-1} e1' A^i z(t) - r(t), \tag{26}$$

where $e1$ is the eight-by-one indicator vector defined above. We use the bias-corrected VAR parameters from the unconstrained system to construct theoretical spreads as in Eq. (26).[6]

One can easily calculate the time series properties of the theoretical spreads for the sample of data used in the estimation. The autocorrelations of the theoretical spreads match the autocorrelations of the actual term spreads much better than the term spreads implied by the autoregressive data-generating process. For example, the first-order autocorrelations of the 12-, 36-, and 60-month theoretical spreads versus those of the actual spreads are 0.58 versus 0.69, 0.83 versus 0.84, and 0.88 versus 0.88, respectively. Of course, we know from Table 1 that the expectations hypothesis is not supported by the data. Hence, there must be dimensions along which the time series properties of the theoretical spreads do not match those of the actual spreads. It turns out that the standard deviations of the theoretical spreads are smaller than those of the actual spreads. The standard deviations of the 12-, 36-, and 60- month theoretical spreads are 57%, 63%, and 66% of the standard deviations of the respective actual spreads. The correlations of the theoretical spreads with the short rate are also more negative than the respective correlations between the actual spreads and the short rate. The correlations of the 12-, 36-, and 60-month theoretical spreads with the short rate are -0.52, -0.68, and -0.74 versus the respective correlations for the actual spreads of 0.08, -0.16, and -0.23. Of course, the correlations of the theoretical spreads and the short rate are less negative than the -1 implied by the AR(1) data-generating process.

5.2. Estimation results

Tables 4 and 5 contain the estimation results for the VAR-GARCH model. Panel A in Table 4 reports the parameter estimates of the second-order VAR and the bias-corrected counterparts. The eigenvalues of the resulting companion form of the original parameter matrix and the bias-corrected parameter matrix are reported in panel B of Table 4. Notice that, as with the AR(1) case, bias adjustment increases the persistence of the series since the maximal eigenvalue increases, but none of the eigenvalues have moduli larger than one. Hence, the bias-adjusted VAR system is stationary. Although the term spreads are not

[6] This approach is similar to that of Campbell and Shiller (1991), who use bivariate VARs in the change in the short rate and the term spread. As an alternative data-generating process, we considered estimation of the four-variable VAR subject to the restrictions of the expectations hypothesis. Although it is straightforward to write down the restrictions, as in Melino (1983), our attempts to estimate the VAR subject to these constraints failed. Furthermore, heteroskedasticity in short rates makes it harder to construct an arbitrage-free economy in which the expectations hypothesis holds. This is further discussed in Bekaert et al. (1995).

Table 4
A four-variable VAR-GARCH data-generating process

The sample period is May 1952 to December 1995. Panel A reports the parameter estimates of a second-order VAR for $y(t) = [r(t), s(t,12), s(t,36), s(t,60)]'$, where $r(t)$ is the short rate at time t and $s(t, n)$ represents the n-month term spread at time t. Heteroskedasticity-consistent standard errors are in parentheses. The bias-adjusted coefficients in panel A are derived from a Monte Carlo experiment in which 200,000 VARs are estimated using the original OLS point estimates as the data-generating process. New data are constructed assuming that the innovations of the VAR are normally distributed using the estimated covariance matrix. For each experiment, 626 values of the $y(t)$ are generated using the unconditional means as starting values, and the first 100 values are discarded before running the VAR with 524 observations. The differences between the original point estimates and the averages of the 200,000 experiments are then added to the point estimates to get the bias-adjusted values. The last line reports the adjusted R^2. Panel B reports the eigenvalues of the companion forms for the original estimates and the bias-adjusted estimates. Panel C reports the l-test for serial correlation (Cumby and Huizinga, 1992) using the first four autocorrelations of the residuals (l(4)), and the Ljung-Box (1978) test applied to the squared residuals using four autocorrelations. Both test statistics are distributed as $\chi^2(4)$. Ku stands for excess kurtosis, Sk stands for skewness and BJ is the Bera-Jarque (1982) normality test. The p-values in brackets are based on the relevant asymptotic χ^2 distributions.

Panel A: VAR parameter estimates

Coef.	Short Rate	Bias Adj.	12-Mo. Spread	Bias Adj.	36-Mo. Spread	Bias Adj.	60-Mo. Spread	Bias Adj.
Const.	0.003	−0.068	0.083	0.090	0.075	0.091	0.091	0.109
	(0.069)		(0.050)		(0.058)		(0.060)	
$r(t-1)$	1.274	1.284	0.053	0.052	−0.083	−0.085	−0.178	−0.180
	(0.118)		(0.085)		(0.087)		(0.093)	
$r(t-2)$	−0.304	−0.302	−0.032	−0.034	0.108	0.105	0.206	0.203
	(0.119)		(0.086)		(0.086)		(0.092)	
$s(t-1, 12)$	0.519	0.520	0.245	0.252	−0.471	−0.471	−0.611	−0.609
	(0.324)		(0.192)		(0.240)		(0.258)	
$s(t-2, 12)$	−0.085	−0.094	0.157	0.170	0.083	0.089	0.143	0.151
	(0.262)		(0.163)		(0.193)		(0.207)	
$s(t-1, 36)$	0.032	0.033	0.016	0.017	0.378	0.388	−0.053	−0.051
	(0.595)		(0.414)		(0.489)		(0.513)	
$s(t-2, 36)$	−0.631	−0.643	0.510	0.518	0.697	0.726	0.388	0.410
	(0.442)		(0.327)		(0.366)		(0.381)	
$s(t-1, 60)$	0.108	0.110	0.291	0.291	0.543	0.540	1.013	1.017
	(0.462)		(0.327)		(0.368)		(0.388)	
$s(t-2, 60)$	0.274	0.289	−0.545	−0.560	−0.539	−0.568	−0.207	−0.230
	(0.374)		(0.278)		(0.299)		(0.312)	
Adj. R^2	0.962		0.530		0.748		0.806	

Table 4. (continued)

Panel B: eigenvalues

Not bias-adjusted	Bias-adjusted
− 0.369	− 0.373
− 0.240	− 0.248
0.067 − 0.071i	0.064 − 0.066i
0.067 + 0.071i	0.064 + 0.066i
0.657	0.674
0.812	0.827
0.931	0.937
0.986	0.997

Panel C: diagnostics for the VAR residuals

	l(4)	Q2(4)	Ku	Sk	BJ
$r(t)$	3.660	91.11	10.98	− 1.110	2742.0
	[0.454]	[0.000]	[0.000]	[0.000]	[0.000]
$s(t, 12)$	3.149	110.0	4.865	0.537	541.9
	[0.533]	[0.000]	[0.000]	[0.000]	[0.000]
$s(t, 36)$	2.732	125.5	5.247	0.255	606.6
	[0.604]	[0.000]	[0.000]	[0.017]	[0.000]
$s(t, 60)$	3.598	108.0	6.434	0.329	913.3
	[0.463]	[0.000]	[0.000]	[0.002]	[0.000]

individually statistically significant predictors of the future short rate, a joint test reveals significant predictive power. A Wald test on the six term spread coefficients in the short-rate equation yields a value of 34.05, well above the critical value for a 1% significance level. The diagnostic tests in panel C reveal that the residuals are serially uncorrelated but display significant heteroskedasticity and deviations from normal distributions. We do not perform a Monte Carlo analysis of the properties of the VAR because its purpose is to serve as a more realistic data-generating process than the AR(1) model. Undoubtedly, statistics such as the Schwarz (1978) criterion, the Wald test for predictability of the short rate, and the Wald test of the expectations hypothesis are all subject to small-sample bias.

Panel A of Table 5 reports the parameter estimates from the GARCH model. The coefficients of the square-root processes are all significant, indicating that conditional variances of both the short rate and the spreads increase with the level of the short rate. The GARCH parameters display the persistence that is typical for the conditional variance processes of financial data. Not surprisingly, the residuals of the term spreads have statistically negative factor loadings with

Table 5

A four-variable VAR-GARCH data-generating process

The sample period is May 1952 to December 1995. Panel A reports parameter estimates of a multivariate GARCH model in the bias-corrected residuals of the VAR in Table 4. The innovation vector of the VAR follows a factor structure with the short rate and the 60-month term spread as the factors:

$$\varepsilon(t) = F e(t),$$

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 \\ f_{21} & 1 & 0 & f_{24} \\ f_{31} & 0 & 1 & f_{34} \\ f_{41} & 0 & 0 & 1 \end{bmatrix}.$$

The vector $e(t)$ represents the idiosyncratic innovations, with $\mathrm{E}[e(t)\, e(t)' \,|\, I(t-1)] = H(t)$ and $H(t)$ a diagonal matrix. The conditional covariance matrix of the innovations $\varepsilon(t)$ is $\Omega(t)$ and equals $F H(t) F'$. Each diagonal element in $H(t)$ follows:

$$h(t, i) = \gamma_i \sqrt{r(t-1)} + \alpha_i e(t-1, i)^2 + \beta_i h(t-1, i) \quad i = 1, 2, 3, 4.$$

Estimation is by quasi-maximum likelihood, and White (1980) standard errors are in parentheses. Panel B reports diagnostic tests on the standardized idiosyncratic shocks, $z(t,i) = e(t,i)/H(t,ii)^{0.5}$ $(i = 1,2,3,4)$, where i (ii) denotes the ith (iith) element in the vector (matrix). The second column tests whether the first four autocorrelations of $z(t, i)$ are zero, the third column tests whether the first four autocorrelations of $z(t, i)^2 - 1$ are zero as in Bekaert and Harvey (1997). In the fourth column, the first test statistic examines whether the short rate residual is uncorrelated with the three idiosyncratic shocks $e(t, i)$ $(i = 2, 3, 4)$ and the test statistic for the 60-month term spread examines whether $e(t, 4)$, the idiosyncratic shock of the 60-month term spread, is uncorrelated with $e(t, 2)$ and $e(t, 3)$, the idiosyncratic shocks of the other term spreads. P-values based on the relevant asymptotic χ^2 distributions are reported in brackets. In panel C, normality tests are applied to the idiosyncratic innovations scaled by their conditional volatilities, $z(t, i)$, for $i = 1, 2, 3, 4$. Ku stands for excess kurtosis, Sk stands for skewness, and BJ is the Bera–Jarque (1982) normality test. The asymptotic p-values are in brackets.

Panel A: volatility parameter estimates

	γ_i	α_i	β_i	f_{i1}	f_{i4}
$r(t)$	0.0067	0.322	0.659	1	0
	(0.0016)	(0.015)	(0.015)		
$s(t, 12)$	0.0011	0.141	0.800	− 0.435	0.839
	(0.0002)	(0.018)	(0.018)	(0.046)	(0.002)
$s(t, 36)$	0.0003	0.122	0.795	− 0.611	1.039
	(0.00004)	(0.021)	(0.021)	(0.042)	(0.0002)
$s(t, 60)$	0.0007	0.139	0.857	− 0.690	1
	(0.0003)	(0.003)	(0.003)	(0.036)	

Table 5. (continued)

Panel B: Diagnostics for the standardized idiosyncratic shocks

	Serial correlation of standardized idiosyncratic shocks	Serial correlation of squared standardized idiosyncratic shocks	Orthogonality of idiosyncratic shocks
$r(t)$	7.415 [0.116]	3.284 [0.511]	4.053 [0.256]
$s(t, 12)$	10.721 [0.029]	2.988 [0.560]	–
$s(t, 36)$	11.756 [0.019]	3.390 [0.495]	–
$s(t, 60)$	5.003 [0.287]	4.100 [0.393]	3.584 [0.167]

Panel C: properties of the scaled idiosyncratic shocks

	Ku	Sk	BJ
$r(t)$	2.708 [0.000]	− 0.582 [0.000]	189.7 [0.000]
$s(t, 12)$	1.331 [0.000]	0.505 [0.000]	61.00 [0.000]
$s(t, 36)$	0.736 [0.0006]	0.087 [0.419]	12.49 [0.002]
$s(t, 60)$	0.875 [0.000]	0.217 [0.043]	20.81 [0.000]

respect to the short-rate residual, whereas the 12- and 36-month term spreads have positive factor loadings with respect to the idiosyncratic shock of the 60-month term spread.

Panel B provides several specification tests. Define the ith idiosyncratic shock of the VAR scaled by its conditional volatility as $z(t, i)$. If the model is correctly specified, $z(t, i)$ and $z(t, i)^2 - 1$ should be serially uncorrelated. Generalized method of moments (Hansen, 1982) tests of these hypotheses using four correlations are reported in the first and second columns. Bekaert and Harvey (1997) describe these tests and discuss their small-sample properties in more detail. Given their results, there does not appear to be any remaining serial correlation in $z(t, i)^2 - 1$, for all i and $z(t, 1)$ and $z(t, 4)$ but there is some weak evidence of remaining serial correlation in $z(t, 2)$ and $z(t, 3)$. The results are not sensitive to the presence or absence of an Andrews (1991) serial correlation correction for the weighting matrix. We also test whether the idiosyncratic shocks are actually uncorrelated as the model implies. The first test shows that the short-rate shock is uncorrelated with the three other idiosyncratic shocks. The second test shows that the idiosyncratic shocks for the 12- and 36-month term spreads are uncorrelated with the idiosyncratic shock for the 60-month term spread equation. This is the main dimension along which a one-factor model fails. In summary, the diagnostic tests do not reveal strong evidence against our model.

Panel C reports some distributional properties for the scaled idiosyncratic shocks, that is $e(t,i)$, the ith idiosyncratic shock, divided by its conditional standard deviation, the square root of the ith element on the diagonal of $H(t)$. Compared to the original residuals (see panel B in Table 4), the distributions of these shocks seem much closer to a normal distribution. Nevertheless, substantial and statistically significant nonnormalities remain, confirming the need for quasi-maximum likelihood estimation.

5.3. *Monte Carlo analysis*

To generate observations on short rates and long rates, we use a bootstrap approach. Panel C from Table 5 shows that the idiosyncratic shocks from the VAR-GARCH model, scaled by their conditional volatilities, are not normally distributed. Hence, we draw from these scaled shocks with replacement, and we use the VAR-GARCH model of Tables 4 and 5 to construct series of short rates and long rates. Eq. (26) is then employed to construct long rates that satisfy the expectations hypothesis. Finally, we perform the statistical tests in Table 1 for 5000 independent replications.

The distributions of the specification test statistics under this alternative data-generating process are summarized in Table 6. As before, we report the mean, the standard deviation, and the left-tail behavior of the small-sample distributions of slope coefficients.

As with the AR(1) model for the short interest rate, the slope coefficients implied by our more realistic data-generating process have substantial upward bias. For example, the mean of the estimated slope coefficients in Eq. (2) is 2.259 for the 60-month horizon, as compared to the population value of unity.

Unlike the distributions reported in Table 3, both the bias and the dispersion in the slope coefficient estimators increase with maturity in Table 6. The increased dispersion of the slope coefficients has a dramatic effect on the left-tail critical values of the distributions for Eq. (2). Consider the 60-month horizon and Eq. (2) with approximation error. The 1% critical value in panel B of Table 3 is 0.777, whereas the comparable value in Table 6 is -1.429.

In general, the AR(1) model overstates both the bias and the dispersion for the shorter maturities relative to the more realistic VAR-GARCH model, but the AR(1) model understates the dispersion for the longer maturities. At all maturities, the results with the VAR-GARCH model have a small-sample standard

Table 6
Monte Carlo distributions of the slope coefficients, and VAR-based statistics using the VAR-GARCH model as the data-generating process

The Monte Carlo evidence is based on 5000 replications. The data-generating process is a four-variable VAR-GARCH model based on the parameters reported in Tables 4 and 5. The sample period is May 1952 to December 1995. A second-order VAR is estimated for $y(t) = [r(t), s(t, 12), s(t, 36), s(t, 60)]$ where $r(t)$ is the short rate at time t and $s(t, n)$ represents the n-month term spread at time t. The VAR parameters, reported in Table 4, are bias-adjusted and the corresponding residuals are used in a multivariate GARCH estimation. The innovation vector of the VAR follows a factor structure with the short rate and the 60-month term spread as the factors:

$$\varepsilon(t) = F e(t),$$

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 \\ f_{21} & 1 & 0 & f_{24} \\ f_{31} & 0 & 1 & f_{34} \\ f_{41} & 0 & 0 & 1 \end{bmatrix}.$$

The vector $e(t)$ represents the idiosyncratic innovations, with $\mathrm{E}[e(t)e(t)' \mid I(t-1)] = H(t)$ and $H(t)$ a diagonal matrix. The conditional variance–covariance matrix of the innovations $\varepsilon(t)$ is $\Omega(t)$ and equals $FH(t)F'$. Each diagonal element in $H(t)$ follows:

$$h(t, i) = \gamma_i \sqrt{r(t-1)} + \alpha_i e(t-1, i)^2 + \beta_i h(t-1, i) \quad i = 1,2,3,4.$$

The parameters for the GARCH model are reported in Table 5.

To generate Monte Carlo series on $y(t)$, we bootstrap from the idiosyncratic shocks, scaled by their conditional volatilities, and use the VAR-GARCH model to generate a sample of 524 observations on $y(t)$. We then use the bias-adjusted VAR parameter matrix to create term spreads that satisfy the expectations hypothesis.

The columns labeled Mean, σ, 1%, 5%, and 10% are the sample mean, the standard deviation, and the respective quantiles of the empirical distributions. The panels correspond to five different tests. Eq. (2) reports the slope coefficients from regressions of the change in the yield on an n-period bond on $[1/(n-1)]$ times the term spread between the n-period yield and the short rate. In panel B, the same regression is run but the $(n-1)$-period yield at time $t+1$ is approximated by the n-period yield at time $t+1$. Eq. (3) reports the slope coefficients from regressions of the weighted average of changes in future short rates on the term spread. Panels D and F report statistics based on a first-order bivariate VAR in the change in the short rate and the n-period term spread. The two statistics are the correlation between the theoretical spread that satisfies the expectations hypothesis and the actual spread and the ratio of the standard deviation of the theoretical spread to the standard deviation of the actual spread. Panels E and G report the same statistics for a fourth-order bivariate VAR.

n	Mean	σ	1%	5%	10%
Panel A: Eq. (2)					
12	1.318	1.094	−1.032	−0.258	0.111
36	1.868	1.816	−1.614	−0.690	−0.193
60	2.259	2.231	−1.883	−0.814	−0.241

Table 6. (continued)

n	Mean	σ	1%	5%	10%
Panel B: Eq. (2) (with approximation error)					
12	1.647	1.082	−0.541	0.131	0.483
36	2.257	1.785	−1.197	−0.239	0.253
60	2.635	2.198	−1.429	−0.364	0.199
Panel C: Eq. (3)					
12	1.134	0.363	0.280	0.621	0.754
36	1.313	0.437	0.271	0.613	0.781
60	1.432	0.456	0.258	0.670	0.849
Panel D: VAR Statistics (Order = 1) Correlation Coefficient					
12	0.988	0.089	0.897	0.971	0.984
36	0.992	0.103	0.974	0.993	0.996
60	0.994	0.095	0.986	0.996	0.998
Panel E: VAR Statistics (Order = 4) Correlation Coefficient					
12	0.940	0.112	0.421	0.807	0.883
36	0.972	0.102	0.629	0.928	0.960
60	0.981	0.098	0.775	0.956	0.977
Panel F: VAR Statistics (Order = 1) Standard Deviation Ratio					
12	1.091	0.232	0.577	0.818	0.880
36	1.112	0.226	0.484	0.790	0.879
60	1.119	0.215	0.491	0.792	0.887
Panel G: VAR Statistics (Order = 4) Standard Deviation Ratio					
12	1.189	0.337	0.590	0.751	0.840
36	1.213	0.336	0.406	0.645	0.791
60	1.215	0.315	0.383	0.649	0.806

deviation of the slope coefficient estimator for Eq. (2) that is approximately equal to its small-sample mean. As a result, negative estimates can occur relatively frequently, in spite of the substantial positive bias. In particular, negative estimates of this coefficient for the 36- and 60-month maturities occur more than 10% of the time. The Monte Carlo analysis confirms that the approximation $n \approx n-1$ in Eq. (2) exacerbates the bias in the slope coefficient estimator.

The slope coefficient estimates of Eq. (3) are less biased than those of Eq. (2), and they also display less dispersion. The means and the standard deviations of the estimates from the VAR-GARCH model for the 12- and 36-month horizons are less than those of the AR(1) model, and the critical values of the VAR-GARCH model are lower than those of the AR(1) model.

It is apparent that the small-sample distributions from the Monte Carlo experiments differ substantially from the asymptotic distributions. In particular, the distributions of the slope coefficients are positively biased and asymmetric. Figs. 1 and 2 compare the empirical distributions from our Monte Carlo simulations of the VAR-GARCH model to the respective asymptotic distributions of the slope coefficients corresponding to Eq. (2) without approximation error and Eq. (3) for the 60-month horizon. The asymptotic distributions are normally distributed with means of unity and standard errors from Newey and

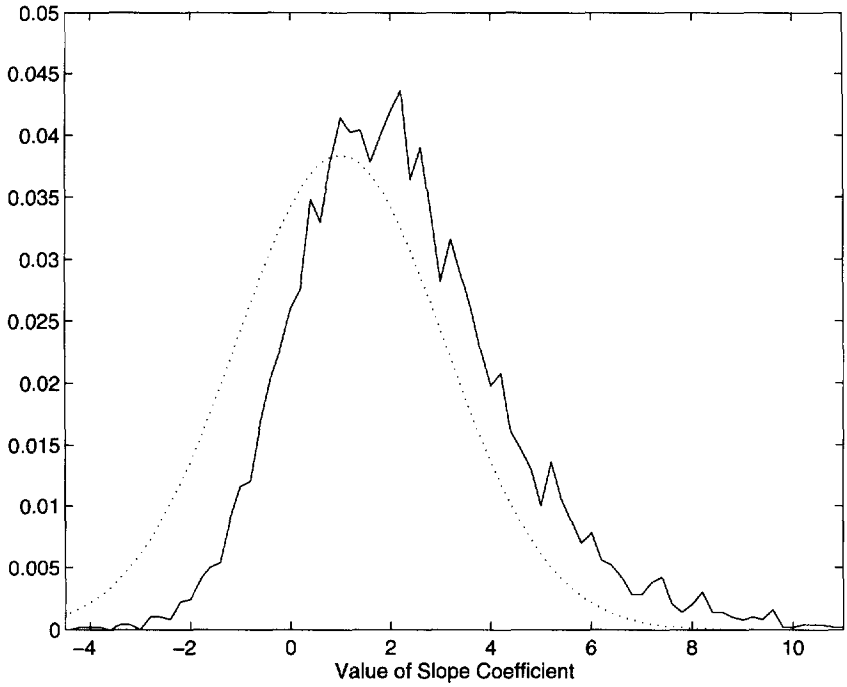

Fig. 1. Monte Carlo distribution vs. asymptotic distribution of the first Campbell–Shiller (1991) specification statistic, Eq. (2). The solid line displays the Monte Carlo density function (scaled histogram) for the OLS estimate of α_1, the slope coefficient in Eq. (2), when the maturity equals 60 months, the sample size equals 524 months, and the data are generated by the VAR-GARCH process given in Tables 4 and 5. The Monte Carlo evidence is based on 5000 replications. The dotted line displays the density implied by the asymptotic approximation $\sqrt{T}\,[\hat{\alpha}_1 - 1] \sim \text{normal}\,(0, T\sigma_\alpha^2)$. The asymptotic standard error σ_α was set equal to 2.08198, the average Newey-West standard error over the 5000 Monte Carlo replications. In computing σ_α, a single Newey-West lag was used.

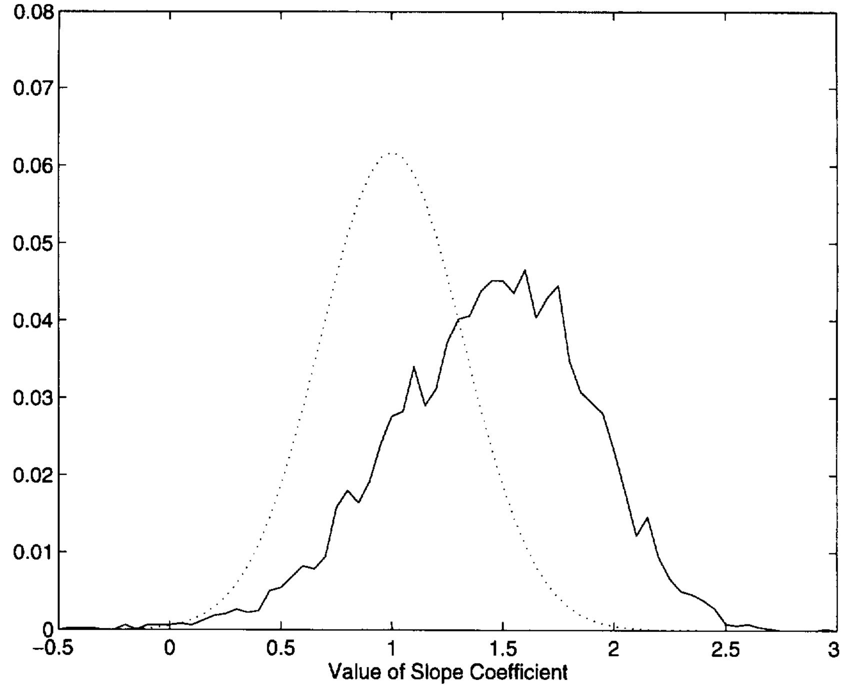

Fig. 2. Monte Carlo distribution vs. asymptotic distribution of the second Campbell–Shiller (1991) specification statistic, Eq. (3). The solid line displays the Monte Carlo density function (scaled histogram) for the OLS estimate of δ_1, the slope coefficient in Eq. (3), when the maturity equals 60 months, the sample size equals 524 months, and the data are generated by the VAR-GARCH process given in Tables 4 and 5. The Monte Carlo evidence is based on 5000 replications. The dotted line displays the density implied by the asymptotic approximation $\sqrt{T}\,[\hat{\delta}_1 - 1] \sim$ normal $(0, T\sigma_{\delta}^2)$. The asymptotic standard error σ_δ was set equal to 0.322965, the average Newey-West standard error over the 5000 Monte Carlo replications. In computing σ_δ, 60 Newey-West lags were used (one more than the minimum number of lags needed to account for the overlap in constructing the dependent variable in (3) from monthly data).

West (1987). The standard errors are the sample means of the Newey-West estimators from 5000 Monte Carlo experiments. The bias and skewness of the distributions are quite apparent.

To provide some perspective on how large a sample is needed to overcome the small-sample biases, Figs. 3 and 4 contain the empirical distributions of the same coefficients as Figs. 1 and 2 but for three sample sizes: 524, 2000, and 20,000 observations. The biases are still apparent even with 2000 monthly observations (166.67 years). Even with 20,000 observations (1667 years), the dispersion in the distribution is substantial.

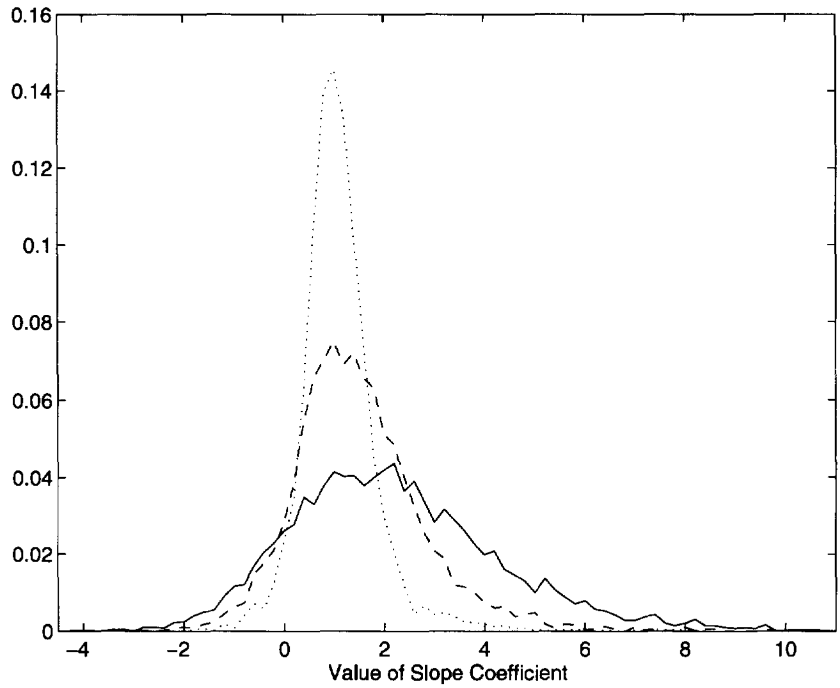

Fig. 3. Monte Carlo distributions of first Campbell–Shiller (1991) specification statistic, Eq. (2), as sample size increases. This figure displays three Monte Carlo density functions (scaled histograms) for the OLS estimate of α_1, the slope coefficient in Eq. (2), when the maturity equals 60 months and the data are generated by the VAR-GARCH process given in Tables 4 and 5. The solid line displays the density function with 524 monthly observations. (This plot is identical to the solid line in Fig. 1.) The dashed line is the corresponding density with 2000 observations, and the dotted line is the density with 20,000 observations. This Monte Carlo evidence is based on 5000 replications.

The two VAR statistics (panels D–G of Table 6) have somewhat different small-sample properties. We report VAR statistics based on both first-order and fourth-order VARs because the results in Table 3 are based on a first-order specification and the results in Table 1 are from a fourth-order specification. Recall that the correlation statistic is virtually unbiased under the AR(1) specification, with a very tight distribution. Under the more realistic data-generating process, this statistic displays a small but nontrivial bias, with somewhat more dispersion. The statistics based on the fourth-order specification are slightly more biased and dispersed than those based on the first-order specification. Evidently, the lack of bias and dispersion reported in Table 3 is due in part to the singularity in the vector process comprising the long and short rates. The standard deviation ratio for the shorter maturities is less biased and

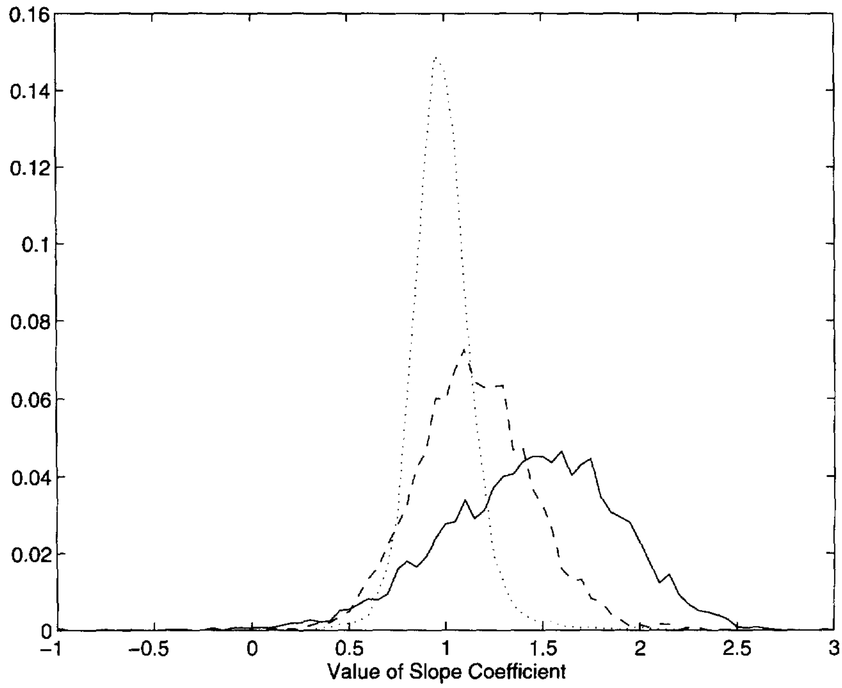

Fig. 4. Monte Carlo distributions of second Campbell–Shiller (1991) specification statistic, Eq. (3), as sample size increases. This figure displays three Monte Carlo density functions (scaled histograms) for the OLS estimate of δ_1, the slope coefficient in Eq. (3), when the maturity equals 60 months and the data are generated by the VAR-GARCH process given in Tables 4 and 5. The solid line displays the density function with 524 monthly observations. (This plot is identical to the solid line in Fig. 2.) The dashed line is the corresponding density with 2000 observations, and the dotted line is the density with 20,000 observations. This Monte Carlo evidence is based on 5000 replications.

dispersed than in the AR(1) case. However, the small-sample distribution of this statistic for the 60-month horizon is comparable to that implied by the AR(1) model.

5.4. Inference using the small-sample distributions

Consider how inferences about the expectations hypothesis using the small-sample distributions differ from the inferences drawn using the asymptotic theory. To be concrete, we focus the discussion on the 60-month horizon, but the results for the other two horizons are very similar. Based on asymptotic distribution theory, the results in Table 1 would indicate a moderately strong rejection of the expectations hypothesis with the specification in Eq. (2) ($\hat{\alpha}_1 = -2.320$, s.e. = 1.479, one-sided p-value = 1.24%). In contrast, the evidence against the expectations hypothesis is somewhat stronger with our small-sample distribution. In particular, we find that only 19 of our 5000 estimated slope coefficients are more negative than the sample value (a small-sample p-value of 0.004). The reason for this stronger rejection is the extreme positive bias in the small-sample distribution. The specification corresponding to Eq. (3) did not provide evidence as strong against the null hypothesis with asymptotic inference as did Eq. (2) ($\delta_1 = 0.569$, s.e. = 0.324, one-sided p-value = 9.17%). In contrast, we find that 96.6% of the 5000 point estimates are higher than 0.569. Again, the small-sample distribution indicates somewhat more evidence against the null hypothesis than the asymptotic distribution. The ratio of the standard deviations of the theoretical and actual spreads is slightly more than two asymptotic standard errors less than one (0.407, s.e. = 0.272, one-sided p-value = 1.46%). Our small-sample distribution implies a similar inference since slightly more than 1% of our simulations had values below 0.407. Finally, the correlation of the theoretical and actual spreads is less than one asymptotic standard error below one (0.960, s.e. = 0.080, one-sided p-value = 30.85%), providing little evidence against the expectations hypothesis. In contrast, our small-sample distribution assigns a one-sided p-value of 5% to this point estimate.

In summary, when our Monte Carlo distributions are used to evaluate the specification tests of the expectations hypothesis, the inference is uniformly less favorable to the null than with the asymptotic distribution. In particular, all four tests reject the hypothesis at (or near) the 5% marginal significance level. We conclude that the results of these four tests appear less paradoxical when viewed from the perspective of the small-sample distributions. They make a consistent case against the expectations hypothesis. The only remaining issue is why some statistics reject more strongly than others. We suspect that this simply reflects differences in the power of the tests.

6. The effect of additive noise on the small-sample distributions

On the basis of the results in the preceding section, we conclude that the expectations hypothesis cannot be saved by appealing to problems with asymptotic distribution theory. The issue at hand, then, is to provide a satisfactory explanation of these rejections. It is beyond the scope of this paper to attempt a comprehensive survey of possible reasons for the theory's poor performance with US data. However, we do want to address one possible explanation put forth by Campbell and Shiller (1991) and Hardouvelis (1994).

Suppose the true yield on an n-month zero-coupon bond, denoted $r^*(t, n)$, satisfies the expectations hypothesis, but observed yields $r(t, n)$ are contaminated by serially uncorrelated noise $\zeta(t,n)$ that is also uncorrelated with true short rates and long rates at all leads and lags:

$$r(t,n) = r^*(t,n) + \zeta(t,n). \tag{27}$$

The OLS estimators of the slope coefficients α_1 and δ_1 in Eqs. (2) and (3), respectively, will not converge to unity. Rather,

$$\text{plim}\,(\alpha_1) = 1 - b_n\,, \tag{28}$$

$$\text{plim}\,(\delta_1) = 1 - \frac{b_n}{n}, \tag{29}$$

where

$$b_n \equiv \frac{n\sigma_n^2 - (n+1)\,\sigma_{n,1} + \sigma_1^2}{\Sigma_n + \sigma_n^2 + \sigma_1^2 - 2\sigma_{n,1}}, \tag{30}$$

and $\Sigma_n \equiv \text{var}\,[r^*(t,n) - r^*(t,1)]$ (the variance of the true n-month term spread), $\sigma_n^2 \equiv \text{var}\,(\zeta(t,n))$, $\sigma_1^2 \equiv \text{var}\,(\zeta(t,1))$, and $\sigma_{n,1} \equiv \text{cov}(\zeta(t,n), \zeta(t,1))$. If the covariance terms $\sigma_{n,1}$ are small relative to the variances σ_n^2 and σ_1^2, then both plim(α_1) and plim(δ_1) will be below unity. Furthermore, plim(α_1) will tend to fall as n increases, while plim(δ_1) will tend to rise as n increases. The plim(α_1) does not unambiguously decrease with n, because Σ_n, σ_n^2, and $\sigma_{n,1}$ also change as n increases. Noise with sufficient variability could induce patterns like those displayed in Table 1, where both α_1 and δ_1 are below unity, but α_1 moves further away from unity as n increases while δ_1 appears to approach unity from below.

We conduct two exercises to see whether noise effects could be a plausible explanation for the patterns displayed in Table 1. In our first exercise, we interpret $\zeta(t,n)$ as pure measurement error, which we calibrate using alternative ways of estimating zero-coupon bond yields. In our second exercise, we simply assume that long yields depart from the expectations hypothesis due to serially uncorrelated disturbances. For example, these disturbances could be due to exogenous demand shocks of the sort described in the noise-trader literature, transient departures from rationality, or time-varying risk-premiums. We ask how big these disturbances must be to explain the patterns found in the data.

Consider first the measurement-error interpretation of $\zeta(t,n)$. One possible source of measurement error is the bid–ask spread in bond markets. However, the bid-ask spreads are quite small for US Treasury coupon bonds, typically three-to-five basis points and rarely exceeding ten basis points. A potential source of larger measurement error is the way in which the zero-coupon yields are constructed. The data we use are constructed by using a cubic spline to approximate the discount function $\exp(-nr(t, n))$. At each date t the parameters of the cubic spline (as a function of n) are chosen such that the approximate discount function minimizes the pricing errors, in the least-squares sense. For additional details see McCulloch (1975) and McCulloch and Kwon (1993). The data for $r(t, n)$ are constructed by evaluating the yield curve implied by the fitted spline for date t at maturity n. This procedure potentially introduces approximation error into the yield data.

To explore the size of this measurement error, we compare bond yield data constructed using the cubic-spline procedure to bond yield data constructed using an alternative procedure developed by Fama and Bliss (1987). At each date, the Fama–Bliss procedure starts with observed Treasury bill yields for the shortest maturities, and then constructs $r(t, n)$ for longer and longer maturities by selecting particular longer maturity bonds to add to the data being fitted. This procedure results in a piece-wise linear yield curve that exactly prices a subset of the bonds. Both sets of zero-coupon yield data were provided by Robert Bliss, and both are constructed from the same monthly bond price observations from 1970 to 1995. In Table 7, panel A, we display the standard deviations of the discrepancies between these two approximations for seven maturities. Panel B of Table 7 displays the correlation matrix of these discrepancies. Note that the discrepancies are rather minor for the long yields (standard deviations between six and nine basis points). However, the standard deviation of the discrepancies between the two approximation methods for the one-month yield is over 20 basis points.

We regard the covariance matrix of these approximation discrepancies, given in Table 7, as an upper bound on plausible measurement error in our term structure data. The covariance matrix in Table 7 may overstate the variability of measurement error for two reasons. First, it implicitly attributes all discrepancies between the two series to measurement error in the McCulloch data. In reality, both procedures are approximations to the true term structure. Second, the covariance matrix given in Table 7 uses data only for 1970–1995, thus excluding the less volatile years between 1952 and 1969. On the other hand, we ignore other sources of measurement error such as bid–ask spreads. To see if measurement errors of this magnitude can help explain the empirical results in Table 1, we simulate 5000 replications of the VAR-GARCH process described in Section 5 and construct long yields according to the expectations hypothesis, as described above. To these yields we then add independently and identically distributed (i.i.d.) Gaussian noise with a covariance matrix as in Table 7. We then compute each of the test statistics and derive the properties of the small-sample distributions, as we did for Table 6.

The results of this exercise are displayed in Table 8. As suggested by Eq. (28), the most dramatic effect is to shift the small-sample distribution of α_1 to the left. For example, the mean of this coefficient with a 60-month maturity is reduced

Table 7
Statistical properties of term structure approximation discrepancy

This table examines the size of measurement error in bond yields by comparing bond yield data constructed using the cubic-spline procedure of McCulloch (1975) with bond yield data constructed using the procedure of Fama and Bliss (1987). Robert Bliss generously provided us with two sets of monthly zero-coupon yield data constructed from the same bond price observations from 1970 to 1995. Panel A reports the standard deviations of the discrepancies between these two approximations for seven maturities. Panel B displays the correlation matrix of these discrepancies.

Panel A: standard deviations

Maturity (in months)	σ (in percentage points)
1	0.205
11	0.075
12	0.088
35	0.062
36	0.063
59	0.089
60	0.090

Panel B: correlation matrix

Maturity (in months)	1	11	12	35	36	59
11	0.0063					
12	– 0.0364	– 0.118				
35	– 0.0740	– 0.191	0.019			
36	0.0138	– 0.126	– 0.048	0.787		
59	– 0.0157	– 0.211	0.276	0.049	– 0.019	
60	– 0.0412	– 0.206	0.216	0.044	– 0.004	0.931

from 2.259 to 1.187. The cutoff for the 1% p-value is reduced from -1.883 to -2.571. The shifts in the small-sample distributions where the approximation $n \approx n-1$ is used (displayed in panel B of Table 8) are of comparable magnitude. While these changes in the distributions of α_1 move in the right direction, they are not sufficient to explain the results in Table 1. According to the distribution summarized in panel B of Table 8, the expectations hypothesis with this degree of measurement error is still rejected at a (one-sided) marginal significance level of 1% for the 36- and 60-month maturities, and at a 5% significance level for the 12-month maturity. Finally, as predicted by Eq. (29), the effect of measurement error on the distribution of δ_1 is much smaller than for α_1, attenuating as n gets larger. The VAR-based statistics, reported in panels D and E, are virtually unaffected by the introduction of measurement error. We only present results for the fourth-order specification because the distributions of the statistics were quite similar and the results in Table 1 are based on that order.

Table 8
Monte Carlo distribution of the slope coefficients, and VAR-based statistics using the VAR-GARCH model with measurement error as the data-generating process

The Monte Carlo evidence is based on 5000 replications. The data-generating process is the four-variable VAR-GARCH model with parameters from Tables 4 and 5. It is identical to the process described in Table 6 except for the addition of measurement error. That is, we add to the yields generated as in Table 6 independently and identically distributed Gaussian noise with a covariance matrix as in Table 7.

The horizon is n months. The columns labeled Mean, σ, 1%, 5%, and 10% are the sample mean, the standard deviation, and the respective quantiles of the empirical distributions. The panels correspond to five different tests. Eq. (2) reports the slope coefficients from regressions of the change in the yield on an n-period bond on $[1/(n-1)]$ times the term spread between the n-period yield and the short rate. In panel B, the same regression is run but the $(n-1)$-period yield at time $t+1$ is approximated by the n-period yield at time $t+1$. Eq. (3) reports the slope coefficients from regressions of the weighted average of changes in future short rates on the term spread. Panels D and E report statistics based on a fourth-order bivariate VAR in the change in the short rate and the n-period term spread. The two statistics are the correlation between the theoretical spread that satisfies the expectations hypothesis and the actual spread and the ratio of the standard deviation of the theoretical spread to the standard deviation of the actual spread.

n	Mean	σ	1%	5%	10%
Panel A: Eq. (2)					
12	0.528	0.946	− 1.347	− 0.759	− 0.506
36	1.328	1.673	− 1.869	− 1.009	− 0.521
60	1.187	2.068	− 2.571	− 1.584	− 1.069
Panel B: Eq. (2) (with approximation error)					
12	0.770	0.949	− 1.000	− 0.489	− 0.239
36	1.676	1.646	− 1.422	− 0.588	− 0.120
60	1.533	2.044	− 2.200	− 1.167	− 0.700
Panel C: eq. (3)					
12	1.042	0.296	0.366	0.642	0.750
36	1.265	0.396	0.316	0.640	0.787
60	1.376	0.418	0.293	0.679	0.842
Panel D: VAR statistics (order = 4) correlation coefficient					
12	0.948	0.096	0.562	0.843	0.903
36	0.974	0.090	0.732	0.940	0.965
60	0.982	0.088	0.824	0.966	0.977
Panel E: VAR statistics (order = 4) standard deviation ratio					
12	1.086	0.283	0.589	0.742	0.809
36	1.172	0.303	0.454	0.674	0.802
60	1.170	0.287	0.408	0.669	0.807

In our second exercise, we simply set $\sigma_1 = \sigma_{n,1} = 0$, and we add i.i.d. Gaussian noise with standard deviation σ_n to the long rates ($n > 1$) as in Eq. (27). It is now straightforward to see that the bias in δ_1 is bounded above by one, whereas the bias in α_1 is larger than one if $(n-1)\sigma_n^2 > \Sigma_n$. Since, empirically, Σ_n is decreasing in n, this condition is unambiguously more easily satisfied for large values of n. As with the first exercise, we conduct 5000 Monte Carlo experiments of length 524. For simplicity, we use the same value of σ_n for all maturities, setting σ_n equal to 25, 37.5, and 50 basis points. When $\sigma_n = 37.5$ basis points, none of the four test statistics reject the model at the 5% marginal significance level, according to a two-sided test.[7] Reducing the noise level makes it difficult to accommodate the estimated values of δ_1 and the standard-deviation ratio statistic. Increasing the noise level substantially above 37.5 basis points makes it easier to accommodate the estimated values of δ_1 and the two VAR-based statistics. However, this additional noise shifts the distribution of α_1 too far into the negative region.

We conclude from this second exercise that departures from the expectations hypothesis can be explained by noise in long rates with a standard deviation in the vicinity of 35 to 40 basis points. While we believe that this is more noise than could plausibly be ascribed to measurement error, we do not regard noise of this magnitude as an insurmountable challenge to economic modeling. In particular, time-varying risk premiums can exhibit the required variability. Since such premiums are likely to be persistent and correlated across maturities, our experiments should not be used to evaluate the effects of realistic time-varying risk premiums on the empirical evidence from the Campbell–Shiller regressions.

7. Conclusions

We explore the small-sample properties of four commonly used tests of the expectations hypothesis of the term structure of interest rates. We document that, even with what seems like a relatively large sample size of 524 monthly observations, the asymptotic distributions of most of these statistics are not to be trusted. Perhaps the most surprising result of the paper is the extreme positive bias in the slope coefficients of traditional single-equation regression tests. The problems arise because these statistics essentially estimate transformations of serial correlation coefficients. There are well-known downward biases in OLS estimates of autocorrelation coefficients for very persistent data and the negative transformation of the regression specification tests creates a positive bias in the slope coefficients. An exception to this pattern is the

[7]Detailed tabulation of the results of these exercises are available upon request.

Campbell–Shiller (1991) VAR-based correlation statistic, which, somewhat surprisingly, displays little bias and has a tight distribution around its probability limit.

When we evaluate the expectations hypothesis relative to the small-sample distributions of these statistics under the null hypothesis, derived either from an AR(1) or from a more realistic VAR-GARCH data-generating process, we find that the evidence against the expectations hypothesis is strengthened and appears to be less paradoxical than if either the asymptotic distributions or the Campbell and Shiller small-sample distributions are employed. We also document that measurement error in bond yields, calibrated from the differences in ways of generating zero-coupon yields from coupon bond data, is insufficient to save the expectations hypothesis.

As with other asset-pricing anomalies, there are essentially three alternative explanations of the data. First, we have ignored rational time-varying risk premiums. Second, we have not attempted to model behavioral characteristics that would cause investors to misprice bonds. Third, we have not attempted to correct small-sample problems due to regime shifts and the possibility that there are inconsistencies between the frequency distributions of what agents thought at a point in time and what actually materialized. Bekaert, Hodrick, and Marshall (1995) explore this alternative application.

The main message of the paper is that it is imperative that researchers use well-designed Monte Carlo experiments with bias-adjusted parameters to assess the significance of their test statistics. As computation costs have fallen, our ability to look beyond asymptotic distribution theory has improved. Unfortunately, what we see often does not look like what we derived theoretically, even in samples of over 43 years of monthly data.

Appendix

Proof of Proposition 1. Under the AR(1) data-generating process of Eq. (5), the regressor in Eq. (2) can be written as

$$\frac{1}{(n-1)}[r(t,n)-r(t)] = \frac{1}{(n-1)}\left[\frac{(1-\rho^n)}{n(1-\rho)}-1\right]r(t)\,, \tag{A.1}$$

and the dependent variable in Eq. (2) can be written as

$$\begin{aligned} r(t+1,n-1)-r(t,n) = {} & \frac{1}{(n-1)}\left[\frac{(1-\rho^n)}{n(1-\rho)}-1\right]r(t) \\ & + \frac{(1-\rho^{n-1})}{(n-1)(1-\rho)}\,v(t+1)\,. \end{aligned} \tag{A.2}$$

By the algebra of OLS, Eqs. (A.1) and (A.2) imply

$$\hat{\alpha}_1 = 1 + \frac{-n(1-\rho^{n-1})}{n(1-\rho)-(1-\rho^n)}\left[\frac{\mathrm{cov}_T(v(t+1), r(t))}{\mathrm{var}_T(r(t))}\right], \tag{A.3}$$

where cov_T and var_T denote sample second moments. The conclusion of Proposition 1 follows from (A.3), along with the observation that, under the assumed data-generating process (5),

$$\mathrm{E}(\hat{\rho}) - \rho = \left[\frac{\mathrm{cov}_T(v(t+1), r(t))}{\mathrm{var}_T(r(t))}\right], \tag{A.4}$$

Proof of Proposition 2. If constant-maturity bonds are used to construct the dependent variable in Eq. (2), the dependent variable can be written as

$$r(t+1, n) - r(t, n) = \frac{1}{n}(\rho^n - 1)\, r(t) + \frac{(1-\rho^n)}{n(1-\rho)}\, v(t+1). \tag{A.5}$$

The regressor is given in Eq. (A.1). The algebra of OLS implies Eqs. (8) and (9) in the text.

Proof of Proposition 3. From the AR(1) data-generating process, the regressor in Eq. (3) can be written as $\eta(\rho, n)\, r(t)$, where $\eta(\rho, n) \equiv (1/n)(1-\rho^n)(1-\rho) - 1$. Note that the dependent variable in Eq. (3) is the weighted average of future short rates minus the current short rate:

$$\sum_{i=1}^{n-1}\left[1 - \frac{i}{n}\right][r(t+i) - r(t+i-1)] = \frac{1}{n}[r(t) + r(t+1) + \cdots + r(t+n-1)] - r(t). \tag{A.6}$$

The slope coefficient in Eq. (3) can therefore be written as

$$\frac{1}{n\eta(\rho, n)}\sum_{j=0}^{n-1}\sum_{t=1}^{T-n+1}\frac{[r(t+j)\, r(t)]}{r(t)^2} - \frac{1}{\eta(\rho, n)} = 1 + \frac{1}{n\eta(\rho, n)}\sum_{j=1}^{n-1}\theta_j \tag{A.7}$$

because each of the bivariate OLS slope coefficient terms in Eq. (A.6) can be written as the true jth autocorrelation plus a bias term:

$$\sum_{t=1}^{T-n+1}\frac{r(t+j)\, r(t)}{r(t)^2} = \rho^j + \theta_j. \tag{A.8}$$

Eq. (10) follows from combining terms in ρ^j to get $\eta(\rho, n)$ and rewriting the coefficient multiplying the sum of the bias terms.

Proof of Proposition 4. We write the theoretical spread as

$$s'(t, n) = x_n \, \Delta r(t) + y_n s(t, n) \tag{A.9}$$

with the x_n and y_n coefficients evaluated according to Eq. (4) in the text using the estimated value of A from the VAR. The sample variance of $s'(t, n)$ will involve the sample variances of $\Delta r(t)$ and $s(t, n)$ as well as their covariance:

$$\mathrm{var}_T(\Delta r(t)) \approx 2(1 - \rho - Q)\,\mathrm{var}_T(r(t)), \tag{A.10}$$

$$\mathrm{var}_T(s(t,n)) \approx \eta(\rho, n)^2 \mathrm{var}_T(r(t)), \tag{A.11}$$

$$\mathrm{cov}_T(s(t, n), \Delta r(t)) \approx \eta(\rho, n)(1 - \rho - Q)\,\mathrm{var}_T(r(t)), \tag{A.12}$$

where the approximation involves $\mathrm{var}_T(r_{t-1}) \approx \mathrm{var}_T(r_t)$. Eqs. (A.10) and (A.12) employ the definition of Q in Eq. (17) in the text. Then, the sample variance of the theoretical spread is

$$\mathrm{var}_T(s'(t, n)) = \{x_n^2\, 2(1 - \rho - Q) + y_n^2 \eta(\rho, n)^2 \\ + 2x_n y_n \,\eta(\rho, n)(1 - \rho - Q)\}\,\mathrm{var}_T(r(t)). \tag{A.13}$$

The ratio of the sample standard deviation of the theoretical spread to the sample standard deviation of the actual spread and the sample correlation of the theoretical spread and the actual spread can be formed from Eqs. (A.12) and (A.13). These statistics are functions of x_n, y_n, and Q. We evaluate the expected value of these functions by taking the expected value of a first-order Taylor's series expansion of the functions around the plims of x_n, y_n, and Q, which are zero, one, and zero, respectively. The result is given in Proposition 4. Finally, the biases in the coefficients x_n and y_n as a function of the biases in the A parameters are found using the results of Graham (1981, Chapter 4, Section 4.6, pp. 67–68).

Proof of Proposition 5. First, partition the VAR coefficient matrix A as

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}. \tag{A.14}$$

The estimated coefficients are given by

$$\begin{bmatrix} \hat{a}_{i1} \\ \hat{a}_{i2} \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{T} \Delta r(t)^2 & \eta(\rho,n) \sum_{t=1}^{T} \Delta r(t) r(t) \\ \eta(\rho,n) \sum_{t=1}^{T} \Delta r(t) r(t) & \eta(\rho,n)^2 \sum_{t=1}^{T} r(t)^2 \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} \sum_{t=1}^{T} \Delta r(t) y^i(t+1) \\ \sum_{t=1}^{T} \eta(\rho,n) r(t) y^i(t+1) \end{bmatrix}. \tag{A.15}$$

where $y^1(t+1) = \Delta r(t+1)$, $y^2(t+1) = \eta(\rho,n) r(t+1)$, and $\eta(\rho,n)$ is given above in Proposition 3. The determinant of the matrix to be inverted, D, provides the denominator of the VAR coefficients. In deriving the results, we used the following assumption:

$$\sum_{t=1}^{T} r(t)^2 = \sum_{t=1}^{T} r(t-1)^2. \tag{A.16}$$

The following results are also useful:

$$\frac{\sum_{t=1}^{T} \Delta r(t)^2}{\sum_{t=1}^{T} r(t)^2} = 2(1 - Q - \rho), \tag{A.17}$$

$$\frac{\sum_{t=1}^{T} \Delta r(t) r(t)}{\sum_{t=1}^{T} r(t)^2} = 1 - (Q + \rho), \tag{A.18}$$

Eqs. (A.10)–(A.12) imply that

$$\frac{D}{[\sum r(t)^2]^2} = \eta(\rho,n)^2 [1 - (Q+\rho)^2]. \tag{A.19}$$

With V defined as in Eq. (18) in the text, using Eq. (5) as the data-generating process to evaluate $y^i(t)$, and using Eqs. (A.14)–(A.19), Eq. (19) in the text follows from dividing the numerator and denominator of the expression in (A.15) by $[\sum_t^T r(t)^2]^2$. The biases in the VAR parameters in Proposition 4 are found by taking the expected values of a second-order Taylor's series expansions of Eq. (19) around the unconditional mean values of Q and V.

References

Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance estimation. Econometrica 59, 817–858.

Bekaert, G., Harvey, C., 1997. Emerging equity market volatility. Journal of Financial Economics 43, 29–77.

Bekaert, G., Hodrick, R.J., Marshall, D.A., 1995. Peso problem explanations for term structure anomalies. Working paper, Stanford University, Stanford, CA.

Bekaert, G., Hodrick, R.J., Marshall, D.A., 1996. On biases in tests of expectations hypothesis of the term structure of interest rates. NBER technical working paper, No. 191.

Bera, A., Jarque, C., 1982. Model specification tests: a simultaneous approach. Journal of Econometrics 20, 59–82.

Bliss, R., 1994. Testing term structure estimation methods. Working paper, Federal Reserve Bank, Atlanta, GA.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics 72, 307–327.

Bollerslev, T., Chou, R.Y., Kroner, K.F., 1992. ARCH modelling in finance: a review of the theory and empirical evidence. Journal of Econometrics 52, 5–60.

Bollerslev, T., Woolridge, J., 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. Econometric Reviews 11, 143–172.

Campbell, J.Y., Shiller, R.J., 1991. Yield spreads and interest rate movements: a bird's eye view. Review of Economic Studies 58, 495–514.

Cumby, R.E., Huizinga, J., 1992. Testing the autocorrelation structure of disturbances in ordinary least squares and instrumental variables regressions, Econometrica 60, 185–196.

Elliott, G., Stock, J.H., 1994. Inference in time series regression when the order of integration of a regressor is unknown, Econometric Theory 10, 672–700.

Engle, R.F., Ng, V.M., Rothschild, M., 1990. Asset pricing with a factor-ARCH covariance structure: empirical estimations for treasury bills. Journal of Econometrics 45, 213–237.

Fama, E.F., 1984. The information in the term structure of interest rates, Journal of Financial Economics 13, 509–528.

Fama, E.F., Bliss, R., 1987. The information in long-maturity rates. American Economic Review 77, 680–692.

Fisher, I., 1896. Appreciation and interest. Publications of the American Economic Association 11, 23–29, 91–92.

Graham, A., 1981. Kronecker products and matrix calculus: with applications. John Wiley, New York, NY.

Gray, S.F., 1996. Modeling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Economics 42, 27–62.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Hardouvelis, G.A., 1994. The term structure spread and future changes in long and short rates in the G7 countries: is there a puzzle? Journal of Monetary Economics 33, 255–284.

Hodrick, R.J., 1992. Dividend yields and expected stock returns: alternative procedures for inference and measurement. Review of Financial Studies 5, 357–386.

Jorion, P., Mishkin, F., 1991. A multicountry comparison of term-structure forecasts at long horizons. Journal of Financial Economics 29, 59–80.

Kendall, M.G., 1954. Note on bias in the estimation of autocorrelation, Biometrika 41, 403–404.

Kroner K.F., Ng, V.M., 1995. Modelling the time-varying component of asset returns. Barclays BZW, San Francisco, CA.

Ljung, G.M., Box, G.E.P., 1978. On a measure of lack of fit in time series models. Biometrika 67, 297–303.

Lutz, F.A., 1940. The structure of interest rates, Quarterly Journal of Economics 55, 36–63.
Mankiw, N.G., Shapiro, M., 1986. Do we reject too often: small sample properties of tests of rational expectations models. Economic Letters 20, 139–145.
Marriott, F.H.C., Pope, J.A., 1954. Bias in the estimation of autocorrelations. Biometrika 41, 390–402.
McCulloch, J.H., 1975. The tax-adjusted yield curve. Journal of Finance 30, 811–830.
McCulloch, J.H., Kwon, H.C., 1993. U.S. term structure data, 1947–1991. Ohio State working paper # 93-6, Ohio State University, Columbus, OH.
Melino, A., 1983. Essays on estimation and inference in linear rational expectations models. Ph.D. Dissertation, Harvard University, Cambridge, MA.
Newey, W., West, K., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica 55, 703–708.
Richardson, M., Stock, J.H., 1989. Drawing inferences from statistics based on multi-year asset returns. Journal of Financial Economics 25, 323–348.
Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6, 461–464.
Stambaugh, R., 1986. Bias in regression with lagged stochastic regressors. C.R.S.P. working paper No. 156, University of Chicago, Chicago, IL.
Stambaugh, R., 1988. The information in the forward rates: implications for models of the term structure. Journal of Financial Economics 21, 41–69.
Stock, J.H., 1994. Unit roots, structural breaks and trends. In: Engle, R.F., McFadden, D.L., (Eds.), Handbook of Econometrics, vol. IV. North-Holland, Amsterdam, pp. 2739–2841.
Tauchen, G., 1985. The small-sample bias of tests for a risk premium in forward exchange rates. Working paper, Duke University.
White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.
White, H., 1982. Maximum likelihood estimation of misspecified models, Econometrica 50, 1–25.