

Working Paper Series

On Biases in Tests of the Expectations Hypothesis of the Term Structure of Interest Rates

Geert Bekaert, Robert J. Hodrick and
David Marshall

Working Papers Series
Issues in Financial Regulation
Research Department
Federal Reserve Bank of Chicago
January 1996 (WP-96-3)

FEDERAL RESERVE BANK
OF CHICAGO

**On Biases in Tests of the Expectations Hypothesis
of the Term Structure of Interest Rates**

Geert Bekaert

Stanford University and NBER

Robert J. Hodrick

Northwestern University and NBER

David A. Marshall

Federal Reserve Bank of Chicago

January 24, 1996

The views expressed in this paper are strictly those of the authors. They do not necessarily represent the position of the Federal Reserve Bank of Chicago, or the Federal Reserve System. We are grateful for the comments of 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 and the 1995 NBER Summer Institute. Geert Bekaert acknowledges financial support from an NSF grant and the Financial Services Research Initiative at Stanford University. We also thank Linda Bethel and Ann Babb for their help with typing the manuscript.

**On Biases in Tests of the Expectations Hypothesis
of the Term Structure of Interest Rates**

Abstract

We document extreme bias and dispersion in the small sample distributions of five standard regression tests of the expectations hypothesis of the term structure of interest rates. These biases derive from the extreme persistence in short interest rates. We derive approximate analytic expressions for these biases, and we characterize the small-sample distributions of these test statistics under a simple first-order autoregressive data generating process for the short rate. The biases are also present when the short rate is modeled with a more realistic regime-switching process. The differences between the small-sample distributions of test statistics and the asymptotic distributions partially reconcile the different inferences drawn when alternative tests are used to evaluate the expectations hypothesis. In general, the test statistics reject the expectations hypothesis more strongly and uniformly when they are evaluated using the small-sample distributions, as compared to the asymptotic distributions.

The expectations hypothesis is probably the oldest and most studied theory of the term structure of interest rates (see Fisher (1896) and 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 U.S. data depending on the regression specification and the maturity of the bonds. Briefly, the change in the U.S. long-term interest rate does not behave as predicted by the theory. 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, 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:

[T]he 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 (1991) use two regression tests and two specification tests from a vector autoregression. A fifth specification test is that of Fama (1984). Although this specification has delivered rejections of the expectations hypothesis at the short end of the maturity spectrum using U.S. data (see Fama (1984), Fama and Bliss (1987) and Stambaugh (1988)), 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 using international data and the Campbell-Shiller specification tests.

The purpose of this paper is to reexamine the econometric methodology underlying these different

specification tests. The main contribution of the paper is to demonstrate that *all* regression 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) and 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, pre-determined, 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 often 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 one reproduces some empirical evidence on the expectation hypothesis, using five statistical tests that have appeared in the literature. Section two derives the small sample biases for these specifications analytically assuming a first-order autoregressive model for the short rate. Section three examines Monte Carlo evidence on both the slope coefficients and their t-statistics under the first-order autoregression data generating process. The fourth section considers Monte Carlo simulations for a more realistic data generating process for the short rate (a regime-switching model). The last section provides some concluding remarks.

1. A Review of Empirical Evidence on the Expectations Theory of the Term Structure

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

$$r(t,n) = \frac{1}{n} \sum_{i=0}^{n-1} E_t r(t+i) + c_n, \quad (1)$$

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 c_n is the term premium.

A number of tests of equation (1) have been proposed in the literature. First, as noted by Campbell and Shiller (1991), equation (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:

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

Second, equation (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:

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

A third test of equation (1), implemented by Fama (1984), Fama and Bliss (1987), and Jorion and Mishkin (1991), uses forward rates implicit in the term structure. Define $f(t,n-1)$ to be the one-period forward rate at time t for an investment $n-1$ periods in the future:

$$f(t,n-1) \equiv n r(t,n) - (n-1) r(t,n-1). \quad (4)$$

The expectations hypothesis implies that the forward premium, $f(t,n-1) - r(t)$, is the expected change in the short rate, $E_t[r(t+n-1) - r(t)]$. Consequently, the slope coefficient, γ_1 , should equal unity in the regression of the ex post change in the short rate on the forward premium:

$$r(t+n-1) - r(t) = \gamma_0 + \gamma_1 [f(t,n-1) - r(t)] + \varepsilon(t+n-1). \quad (5)$$

The second and third columns of Table 1 reproduce some of the evidence on equations (2) and (3) from Campbell and Shiller (1991). These authors use McCulloch's (1990) monthly data on zero-coupon bond yields, from 1952:1 through 1987:2. In Table 1, the short rate has a one-month maturity. The fourth column displays evidence on equation (5), using the same data set. (The other columns will be discussed below.)

In the estimates of equation (2) reported in Table 1, the slope coefficient is significantly below unity for all maturities, and the point estimate is almost always 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. However, the specification tests in equations (3) and (5) provide weaker evidence against the expectations theory than was found with equation (2). The estimated slope coefficients are almost always positive, and, for the longer maturities, they are insignificantly different from unity.¹

In addition to the single-equation regression tests in equations (2), (3), and (5), Campbell and Shiller (1991) derive tests of equation (1) that use a bivariate vector autoregression (VAR). The VAR uses the change in the short rate, $\Delta r(t)$, and the term spread, $s(t,n) \equiv r(t,n) - r(t)$. To understand the VAR

¹ At this point in the paper, we are defining the concept of statistical significance relative to the asymptotic distribution of the OLS estimators. As we shall see in sections 2 and 3, below, the small-sample distributions of these estimators differ substantially from their asymptotic counterparts, so the asymptotic standard errors should be interpreted with care.

statistics, let A denote the first-order, companion-form of the VAR parameter matrix. From equation (1), the term spread is $s(t,n) = \sum_{i=1}^{n-1} [1-(i/n)] E_t \Delta r(t+i)$. By evaluating the expected changes in the short rate using the VAR, one can obtain an expression for the term spread that Campbell and Shiller (1991) refer to as the "theoretical spread", denoted $s'(t,n)$:

$$s'(t,n) \equiv \sum_{i=1}^{n-1} (1-i/n) e_1' A^i \begin{pmatrix} \Delta r(t) \\ s(t,n) \end{pmatrix}, \quad (6)$$

where e_1 is an indicator vector with one in the first row and zeros everywhere else.

The statistics proposed by Campbell and Shiller (1991) as tests of the expectations hypothesis are (i) the correlation between $s'(t,n)$ and the actual term spread, $s(t,n)$, and (ii) 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 from Campbell and Shiller (1991). It is of interest that the ratio of the standard deviations provides stronger evidence against the expectations hypothesis than the correlation statistic. For the longer maturities, the correlation statistic is close to its theoretical value of unity.

2. Analytical Approximations to the Biases in Regression Tests of the Expectation Hypothesis

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

$$r(t+1) = \mu + \rho r(t) + \varepsilon(t+1). \quad (7)$$

Although we choose the AR(1) for analytical tractability, for monthly data a highly serially correlated AR(1) model is actually a reasonable approximation of the data. In Table 2, we report the OLS estimates of μ , ρ , and σ (the standard deviation of $\varepsilon(t+1)$), which we denote $\hat{\mu}$, $\hat{\rho}$, and $\hat{\sigma}$, respectively. These estimates are reported for the sample corresponding to the data for Table 1.

Table 2 also reports bias-adjusted values of μ , ρ , and σ . Kendall (1954) shows that, to a first-

order approximation,

$$E(\hat{\rho}) - \rho = -\frac{1 + 3\rho}{T} + O\left(\frac{1}{T^2}\right), \quad (8)$$

where T denotes the sample size. The bias adjustment unwinds the bias in equation (8) such that the "bias-adjusted ρ " is $(\hat{\rho} + (1/T))/(1 - (3/T))$, which makes the bias equal to -0.0094 for the 421 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 five specification tests of the expectations hypothesis.

Proposition 1

Under equations (1) and (7), the expected value of the slope coefficient of the first specification test in equation (2) is

$$E(\hat{\alpha}_1) = 1 + \frac{-n(1 - \rho^{n-1})}{n(1 - \rho) - (1 - \rho^n)} [E(\hat{\rho}) - \rho]. \quad (9)$$

(The proofs of the Propositions are given in the Appendix.)

Denote the coefficient multiplying the bias in $\hat{\rho}$ in equation (9) by $A(\rho, n)$ and the bias in $\hat{\rho}$ by θ_1 . Note that $A(\rho, n)$ is negative and θ_1 is negative as well. Therefore, the estimate of the slope coefficient $\hat{\alpha}_1$ is biased upward. Note that $A(\rho, 2) = -2/(1 - \rho)$ and $A(\rho, \infty) = -1/(1 - \rho)$. While the bias decreases as the horizon n increases, the bias is still substantial for large n , especially if ρ is near unity. For example, at the ten-year maturity ($n = 120$), $A(0.9865, 120) = -117.9$, which combined with the estimated value of the bias in $\hat{\rho}$ of -0.0094, produces an expected value of α_1 of 2.109, more than double the asymptotic value of unity. Analytical values for the expected values of the slope coefficients for other values of n will be 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 use the following modification of equation (2), in which the regressand is constructed from yields on constant maturity bonds:

$$r(t+1,n) - r(t,n) = \alpha_0 + \alpha_1 \frac{1}{(n-1)} [r(t,n) - r(t)] + \varepsilon(t+1). \quad (10)$$

While the 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 equation (7), we can analytically determine the size of this approximation error.

Proposition 2

Under equations (1) and (7), the expected value of the slope coefficient of the first specification test using the approximation as in equation (10) is

$$E(\hat{\alpha}_1) = 1 + c(\rho, n) + A_2(\rho, n)\theta_1 \quad (11)$$

where

$$c(\rho, n) = \frac{-n(1-\rho)\rho^n + \rho(1-\rho^n)}{n(1-\rho) - (1-\rho^n)} \quad (12)$$

$$A_2(\rho, n) = \frac{-(n-1)(1-\rho^n)}{n(1-\rho) - (1-\rho^n)}$$

The bias term $A_2(\rho, n)$ is quite similar to $A(\rho, n)$, derived above in Proposition 1. The approximation error term $c(\rho, n)$, which is not zero asymptotically as the sample size increases, is positive and can differ substantially from zero. For $n = 2$, $c(\rho, n) = \rho$. While the approximation error becomes smaller for large n , it is still substantial for maturities often used in empirical work. Even with maturities as high as ten years ($n=120$), $c(0.9865, 120) = 0.584$, and the mean of the OLS estimator of α_1 in the Campbell-Shiller sample is 2.687, 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, equation (3), is also subject to severe small sample bias.

Proposition 3

Under equations (1) and (7), the expected value of the slope coefficient of the second specification test in equation (3) is

$$E[\hat{\delta}_1] = 1 + \frac{-(1-\rho)}{n(1-\rho) - (1-\rho^n)} \sum_{j=1}^{n-1} \theta_j, \quad (13)$$

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

Using results from Kendall (1954, equation (20)) one can derive the following first order approximation to θ_j :

$$\theta_j \approx - \left[\frac{(1+\rho)}{(1-\rho)} (1-\rho^j) + 2j\rho^j \right] \left[\frac{1}{T-(n-1)} \right]. \quad (14)$$

For $n = 2$, the bias is $-\theta_1/(1-\rho)$, which is one-half of the small sample bias in equation (2). For longer horizons, additional bias terms must be added. In Column 2 of Panel C in Table 3 below, 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 (1991)) 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. We will examine the small sample behavior of the t -statistics for the slope coefficients for all regression specifications in Table 4 below.

The third specification test, equation (5), is also subject to small sample bias.

Proposition 4:

Under equations (1) and (7), the expected value of the slope coefficient of equation (5) is

$$E[\hat{Y}_1] = 1 + \frac{-\theta_{n-1}}{(1 - \rho^{n-1})}, \quad (15)$$

where θ_{n-1} denotes the small sample bias in estimating ρ^{n-1} .

These biases are reported below in Column 2 of Panel D in Table 3 and are comparable in magnitude to those for specification (3).

We now turn to biases in the two VAR-based statistics under the maintained assumption that the short rate is generated by the AR(1) in equation (7). We will derive analytical approximations to the bias 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 VAR-order used in Campbell and Shiller (1991). To simplify notation, let $\eta(\rho, n) \equiv (1/n)(1 - \rho^n)/(1 - \rho) - 1$. Under the data generating process of equation (7) the term spread is $s(t, n) = \eta(\rho, n)r(t)$. The first-order VAR can then be written:

$$\begin{pmatrix} \Delta r(t+1) \\ \eta(\rho, n)r(t+1) \end{pmatrix} = A_0 + A \begin{pmatrix} \Delta r(t) \\ \eta(\rho, n)r(t) \end{pmatrix} + \varepsilon(t+1). \quad (16)$$

Let \hat{A} denote the OLS estimator of A. It follows from equations (7) and (16) that

$$\text{plim}(\hat{A}) = \begin{bmatrix} 0 & \frac{\rho-1}{\eta(\rho, n)} \\ 0 & \rho \end{bmatrix}. \quad (17)$$

We can write the theoretical spread as follows:

$$s'(t, n) = x_n \Delta r(t) + y_n s(t, n), \quad (18)$$

where x_n and y_n are the implied coefficients from equation (6). The coefficients x_n and y_n are functions of the estimated A parameters. When x_n and y_n are evaluated using $\text{plim}(\hat{A})$, one finds that $x_n = 0$, 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 one might think that both of the VAR-based statistics would inherit a bias from the biases in x_n and y_n , it turns out that, to a first approximation, only the standard deviation ratio is biased:

Proposition 5:

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:

$$E \left\{ \left[\frac{\text{var}(s'(t,n))}{\text{var}(s(t,n))} \right]^{1/2} \right\} \approx 1 + \frac{(1-\rho)}{\eta(\rho,n)} (x_n \text{ bias}) + (y_n \text{ bias}). \quad (19)$$

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

$$(x_n \text{ bias}, y_n \text{ bias}) = \frac{e l' n^{-1}}{n} \sum_{j=1}^{n-1} (n-j) \left\{ A^j + \sum_{i=1}^2 \sum_{k=1}^2 \left[\sum_{l=0}^{j-1} A^l J_{i,k} A^{j-l-1} \right] E[B_{i,k}] \right\} - (0, 1). \quad (20)$$

where $J_{i,k}$ is the indicator matrix with 1 in the i - k position and zeros everywhere else, and $E[B_{i,k}]$ denotes the bias in the $(i,k)^{\text{th}}$ element of \hat{A} .

To implement Proposition 5, 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 where 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-measured data.) It is likely that this approximation understates the magnitude of the bias.² In the next section, we will compare these first-order approximations to the average values of our Monte Carlo estimates.

² In the univariate case of equation (7), 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.

The VAR coefficients are non-linear functions of the following random variables:

$$Q \equiv \frac{\sum_{t=1}^T \varepsilon(t+1)r(t)}{\sum_{t=1}^T r(t)^2}, \quad (21)$$

$$V \equiv \frac{\sum_{t=1}^T \varepsilon(t+1)r(t-1)}{\sum_{t=1}^T r(t-1)^2}. \quad (22)$$

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

$$\hat{A} = \text{plim}(\hat{A}) + B \equiv \text{plim}(\hat{A}) + \begin{bmatrix} \frac{Q(Q+\rho) - V}{1 - (Q+\rho)^2} & \frac{Q+V}{\eta(\rho, n)(1+Q+\rho)} \\ \eta(\rho, n) \frac{Q(Q+\rho) - V}{1 - (Q+\rho)^2} & \frac{Q+V}{1+Q+\rho} \end{bmatrix}. \quad (23)$$

The bias matrix $E[B]$ is a non-linear function of the moments of Q and V : $\bar{Q} \equiv E[Q]$, $\bar{V} \equiv E[V]$, $\text{var}[Q]$, and $\text{cov}[Q, V]$:

Proposition 6:

To a first-order approximation, the elements of the bias matrix $E[B]$ in equation (23) are given by:

$$E(B_{11}) = \frac{\bar{Q}^2 + \rho\bar{Q} - \bar{V}}{1 - (\bar{Q} + \rho)^2} + \frac{1}{1 - (\bar{Q} + \rho)^2} \left[1 + \frac{2(2\bar{Q} + \rho)(\bar{Q} + \rho)}{1 - (\bar{Q} + \rho)^2} + \frac{(\bar{Q}^2 + \rho\bar{Q} - \bar{V})(1 + 3(\bar{Q} + \rho)^2)}{(1 - (\bar{Q} + \rho)^2)^2} \right] \text{var}[Q] \quad (24)$$

$$- \frac{2(\bar{Q} + \rho)}{(1 - (\bar{Q} + \rho)^2)^2} \text{cov}[Q, V]$$

$$E(B_{22}) = \frac{\bar{Q} + \bar{V}}{1 + \bar{Q} + \rho} - \frac{1}{(1 + \bar{Q} + \rho)^2} \left[1 - \frac{(\bar{Q} + \bar{V})}{(1 + \bar{Q} + \rho)} \right] \text{var}(Q) - \frac{1}{(1 + \bar{Q} + \rho)^2} \text{cov}(Q, V), \quad (25)$$

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

We will report analytical values for the biases in the VAR statistics in Column 2 of Panels E and

F in Table 3, and we compute the moments of Q and V using 20,000 Monte Carlo draws.³ Below, we find that the bias in the ratio of the standard deviations is substantial. The mean value of this statistic is over 30% higher than the asymptotic value of unity.

3. Monte Carlo Results

3.1 Characteristics of the Small Sample Distributions

Kendall (1954) noted that derivations of first-order bias may be of doubtful validity for ρ near unity. This intuition was later confirmed by numerous papers on unit roots.⁴ Kendall also noted 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 biases in the VAR statistics are based on first-order 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. We report the mean, the standard deviation and the left-tail behavior of the small sample distributions and also investigate the distributions of squared t-statistics for the slope coefficients. Tables 3 (slope coefficients) and 4 (squared t-statistics) report results for the 421 observation case coinciding with the Campbell-Shiller sample.

In general, 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. Figures 1 and 2 compare the empirical distributions to the respective asymptotic distributions of the slope coefficients corresponding to equation (2) without approximation error and equation (3) for the 120-month horizon. The asymptotic distributions are normally distributed with means of unity and standard errors from Newey and West (1987). The standard errors are the sample means of

³ We use Monte Carlo approximations to the moments of Q and V because analytic approximations to $\text{var}[Q]$ and $\text{cov}[Q,V]$ are extremely difficult to derive. Kendall (1954) shows that $E[Q] \approx -2\rho/T$, and $E[V] \approx -2\rho^2/T$. Our Monte Carlo estimates of these means were extremely close to the Kendall approximations.

⁴ Stock (1994) provides a recent survey of the unit root literature.

the Newey-West estimators from 5,000 Monte Carlo experiments. The bias and skewness of the distributions is quite apparent.

To provide some perspective on how large a sample one needs to overcome the small sample biases, Figures 3 and 4 contain the empirical distributions of the same coefficients as Figures 1 and 2 but for three sample sizes: 421, 2,000 and 20,000 observations. The biases are still apparent even with 2,000 monthly observations (166.67 years). Even with 20,000 observations (1,667 years), the dispersion in the distribution is substantial.

The squared t-statistics in Table 4 are also strongly biased to the right. For example, the 5% critical value of a $\chi^2(1)$ is 3.841, while the 95th percentile of the empirical distribution of the corresponding test statistic for equation (2) at a horizon of 120 is 5.652. When this specification test is conducted with the approximation in equation (10), the corresponding 95th percentile is 7.460. The biases in the test statistics for equation (3) are even worse because of the problem of estimating the standard error with overlapping data. At the horizon of 120, the 95th percentile is actually 149.806. While this is an extreme value, it might be expected since there are fewer than three non-overlapping observations.

The results of the Monte Carlo experiments in Table 3 support the accuracy of the theoretical bias calculations for the specifications corresponding to equations (2), (3) and (5). For equation (2), the theoretical estimates are no worse than 93% of the mean values of the Monte Carlo experiments. For equation (2) with the approximation error, the theoretical estimates are all 95% of the means of the Monte Carlo experiments. For equation (3), the theoretical estimates are between 95% and 98% of the means of the Monte Carlo experiments. For equation (5), the theoretical estimates are between 95% and 106% 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 $n=120$ ($n=60$), for the Campbell-Shiller sample the coefficient must only be less than 0.968 (1.075) to have a 5% rejection.

The two VAR statistics (Panels E and F of Table 3) have different small sample properties. The correlation statistic for n greater than or equal to 12 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 less well. For $n = 2$, the value of 1.310 is only 68% of the mean value of the Monte Carlo experiments. The substantial underestimate was anticipated since the theoretical estimate assumed that the means of the variables in the VAR are known. For larger n , the approximation is more accurate. For $n = 60$ ($n = 120$) the analytical estimate is 93% (105%) of the mean value of the Monte Carlo experiments. As might be expected since the data generating process is an AR(1), the Monte Carlo biases in the VAR statistics are slightly worse for a fourth-order VAR (the order used in Campbell and Shiller (1991)) than for a first-order specification.

3.2 Inference Using the Small Sample Distributions

Campbell and Shiller (1991) were careful not to base their conclusions completely on asymptotic distribution theory. They conducted Monte Carlo experiments using the estimated VAR as the basis of a data generating process. To generate artificial data that satisfy the expectations theory, they first generated artificial series for $\Delta r(t)$ and $s(t,n)$ using a random number generator and the point estimates of the VAR coefficients and the covariance matrix of the innovations. Then, they generated the theoretical spread, $s'(t,n)$ from equation (6), using the estimated VAR parameters and the series for $\Delta r(t)$ and $s(t,n)$ (which does not satisfy the theory). The series for the theoretical spread (which satisfies the theory) and the short rate series were then used as the raw data to reexamine the small sample properties of the various test statistics. A major difference between our data generating approach and Campbell and Shiller's approach is that we use bias-adjusted coefficients to generate our short rate series.

Now, consider how our inference about the expectations hypothesis differs from the inferences drawn with the asymptotic theory and Campbell and Shiller's Monte Carlo experiments. To be concrete, we focus the discussion on the 120-month horizon. The Campbell-Shiller results based on asymptotic

distribution theory in Table 1 indicate a strong rejection of the expectations hypothesis with the specification in equation (2) ($\hat{\alpha}_1 = -5.024$, s.e. = 2.316, p-value = .009). Campbell and Shiller find that 2.9% of their estimates are more negative than -5.024, and we find that none of our 5,000 estimated slope coefficients are more negative than the sample value. In fact, the minimum value of our statistics is 0.095. This is strong evidence against the null hypothesis. The specification corresponding to equation (3) did not reject the null hypothesis with asymptotic inference ($\hat{\delta}_1 = 1.157$, s.e. = 0.094, p-value = .196). While the asymptotic theory predicts that 1.157 is in the right tail of the distribution, Campbell and Shiller find that 71.3% of the observations are actually more positive than the point estimate, which reflects the strong positive bias in this specification. We find that 84.2% of the point estimates are more positive than 1.157. While this is not as strong evidence against the null hypothesis as specification (1), the differences could easily be due to differential power of the tests. The strong positive bias in the coefficient estimator seems to be the apparent source of the paradox involving the two specifications. The ratio of the standard deviations of the theoretical and actual spreads was slightly less than two asymptotic standard errors less than one (0.476, s.e. = 0.284, p-value = .065). Campbell and Shiller found that 4.3% of their simulations resulted in a smaller value, and we find that 1.1% are smaller. The correlation of the theoretical and actual spreads was less than one asymptotic standard error below one (0.979, s.e. = 0.045, p-value = .641). Campbell and Shiller found that 15.9% of their simulations resulted in a larger difference between the point estimate and one, while we find that 0.979 is less than the minimum value of 0.997. Hence, the small sample distributions of the VAR statistics also imply that there is little support of the null hypothesis, in contrast to the puzzling asymptotic inference.

In summary, when our Monte Carlo distributions are used to evaluate the specification tests of the expectations hypothesis, the hypothesis is rejected more decisively than if the asymptotic distributions or Campbell and Shiller's Monte Carlo distributions were used. The results are consequently uniformly less supportive of the null hypothesis and appear less paradoxical.

4. Robustness to the Data Generating Process

Although short-term interest rates are definitely very persistent, the data generating process used above is not a state-of-the-art model for the short rate. Perhaps the most important failure of the data generating process is that it is conditionally homoskedastic. In this section, we investigate whether our results are robust to the use of a more realistic data generating process for the short rate. Our model builds on the regime-switching models of Hamilton (1988) and Gray (1995a). We model the short interest rate as a two-regime Markov switching model with state-dependent transition probabilities.

Let $S(t)$ be an indicator variable identifying the regime at date t : $S(t) \in \{1, 2\}$. The short interest rate follows the law of motion

$$r(t+1) = \begin{cases} \mu_1 + \beta_1 r(t) + h_1[r(t)]\varepsilon(t+1), & \text{if } S(t+1) = 1, \\ \mu_2 + \beta_2 r(t) + h_2[r(t)]\varepsilon(t+1), & \text{if } S(t+1) = 2, \end{cases} \quad (26)$$

where $\{\varepsilon_t\}$ is a sequence of independent standard normal random variables, and the time-varying conditional standard deviation $h_i[r(t)]$ is given by

$$h_i[r(t)] = \sigma_i r(t)^{\gamma}, \quad \text{for } i = 1, 2. \quad (27)$$

Notice that the realization $r(t+1)$ is affected by two random shocks not known at date t : $\varepsilon(t+1)$ and $S(t+1)$.

The law of motion for $S(t)$ is as follows:

$$\text{Prob}[S(t+1) = 1 \mid S(t) = 1, r(t)] = \frac{\exp[a_1 + b_1 r(t)]}{1 + \exp[a_1 + b_1 r(t)]} \quad (28)$$

$$\text{Prob}[S(t+1) = 2 \mid S(t) = 2, r(t)] = \frac{\exp[a_2 + b_2 r(t)]}{1 + \exp[a_2 + b_2 r(t)]} \quad (29)$$

where $\{a_i, b_i, i = 1, 2\}$ are parameters of the model. Under equations (26)-(29), the conditional distribution of $r(t+1)$ given $r(t)$ and $S(t)$ is a mixture of normals with state-dependent mixing probabilities.

Under the expectations theory of the term structure, long rates are determined by two state variables: the

current short rate and the current regime.

Table 5 presents our parameter estimates, obtained using Gray's (1995b) recursive maximum-likelihood procedure. As in Table 1, we use the McCulloch (1990) data from 1952:1 through 1987:2. These parameter estimates capture a number of appealing features. First, there is strong evidence of conditional heteroskedasticity as the γ_i parameters are quite significantly different from zero. The difference in the conditional volatility of the two regimes is further influenced by the big difference between σ_2 and σ_1 . Regime 2 is clearly the high-volatility regime. There is also more evidence of mean reversion in regime 2, as $\beta_2 < \beta_1$. Furthermore, if the current regime is regime 1, the probability of switching into regime 2, the high-volatility regime, is increasing in $r(t)$. This follows from equation (28) because $b_1 < 0$. Similarly, if the economy is currently in the high-volatility regime, the probability of staying in that regime is increasing in $r(t)$. This follows from equation (29) because $b_2 > 0$. Therefore, regardless of the current regime, the probability that the next period will be the high-volatility regime is increasing in $r(t)$. The high-volatility regime is also the high-mean regime. This fact is determined by simulating 10,000 observations of the model and examining the means of the series conditional on being in the regimes. The mean for regime 1 is 5.211% and the mean for regime 2 is 7.764%.

There are good reasons to believe that this model is particularly vulnerable to small-sample problems. First, the difference between the means in the two regimes, the extremely high persistence within regimes, and the non-trivial conditional heteroskedasticity imply a highly leptokurtic unconditional distribution for short rates.⁵ It is well known that quite large samples are required for conventional asymptotic estimation to work well with leptokurtic distributions. Second, both regimes display near-unit-

⁵ Bollerslev, Chou and Kroner (1992) provide a review of the literature on GARCH models and discuss the problems inherent in doing time series estimation with leptokurtic data.

root behavior for the short rate, implying that the process is nearly non-stationary.⁶ Both considerations suggest that an extremely long time series is necessary for the empirical distribution of the sample to replicate the population distribution.

In order to calculate the distributions of regression coefficients associated with the term structure of interest rates when the short rate follows the regime switching model, we must compute the long-maturity yields implied by the expectations theory. Since the model is highly non-linear, this task is computationally intractable. Nevertheless, we can compute expected future short rates using a discrete-state approximation for the two state variables of the regime switching model: the short interest rate and the regime. We use an equally-spaced grid of 168 points on the space of possible realizations of the short rate for each of the two regimes for a total of 336 discretized states. For each of the states, the transition probabilities are computed by evaluating the conditional density of next period's short interest rate (a state-dependent mixture-of-normals) at each discrete state, and then normalizing these transition probabilities to sum to unity. When we simulate our Markov chain approximation and estimate our regime-switching model from the simulated data, the estimated parameters are all within one-half standard error of the original point estimates reported in Table 5.⁷

We use the linearized Markov process to generate 421 observations on short rates and long rates

⁶ Elliott and Stock (1994) discuss inference when there is near unit root behavior in a regressor. The high persistence of the short rate and the Markov regime structure for the short rate imply that the spread in our model is highly serially correlated as it is in the actual data.

⁷ It is also instructive to compare statistics for a simulation of 10,000 observations of the non-linear short-rate model to the corresponding statistics of our discrete-state approximation. The means are 5.860 and 5.850, respectively; the standard deviations are 2.845 and 3.050; the first-order autocorrelations are 0.971 and 0.973; the minimum values are 0.476 and 0.400; the maximums are 25.23 and 25.45; and the probabilities of regime one are 0.254 and 0.265.

that satisfy the expectations hypothesis. Then, we perform the statistical tests in Table 1 for 5,000 independent replications. The distributions of the specification test statistics under this alternative data-generating process are summarized in Table 6. These distributions are very similar to the distributions displayed for the homoskedastic AR(1) data generating process in Table 3. The mean values of the Monte Carlo experiments are only slightly lower for the alternative data generating process, and the standard deviations are slightly larger in some cases but slightly smaller in others. The differences in mean bias presumably arise because the parameter estimates in Table 5 imply a slightly less persistent short-rate process than the AR(1) process in Table 3. In particular, the first-order autocorrelation of the short interest rate in our regime-switching model is 0.9713, whereas the first-order autocorrelation in Table 3 is 0.9865. We conclude from this exercise that the biases we document are not artifacts of the simple data generating process given in equation (7). Rather, they are likely to be found whenever the short interest rate has extremely high serial persistence.

5. Conclusions

We explore the small sample properties of five commonly-used tests of the expectations hypothesis of the term structure of interest rates. We document that, even with what seem like relatively large sample sizes, the asymptotic distributions of most of these statistics are not to be trusted. Perhaps the most surprising result of the paper is the finding of extreme positive bias in the slope coefficients of traditional single-equation regression tests. Very large biases, extreme dispersion, and substantial skewness are present even in samples containing 35 years of monthly data. The problems arise because these statistics essentially estimate serial correlation coefficients, and there are well-known biases in OLS estimates of autocorrelation coefficients for very persistent data. An exception to this pattern is the Campbell-Shiller correlation statistic, which, somewhat surprisingly, displays virtually no 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, we find that the evidence against the expectations hypothesis is strengthened and appears to be less paradoxical than if either the asymptotic distributions or Campbell and Shiller's (1991) small sample distributions are employed. 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.

Table 1
The Campbell-Shiller (1991) Results

n (months)	Equation (2)	Equation (3)	Equation (5)	Correlation	Standard Dev. Ratio
2	0.002 (0.238)	0.501 (0.119)	0.501 (0.119)	0.736 (0.148)	0.681 (0.136)
12	-1.381 (0.683)	0.161 (0.228)	0.252 (0.216)	0.391 (0.468)	0.382 (0.119)
24	-1.815 (1.151)	0.302 (0.212)	0.518 (0.458)	0.543 (0.764)	0.304 (0.138)
36	-2.239 (1.444)	0.614 (0.230)	0.660 (0.473)	0.770 (0.531)	0.311 (0.227)
48	-2.665 (1.634)	0.873 (0.291)	0.967 (0.426)	0.867 (0.328)	0.338 (0.274)
60	-3.099 (1.749)	1.232 (0.182)	1.530 (0.333)	0.912 (0.218)	0.360 (0.290)
120	-5.024 (2.316)	1.157 (0.094)	0.845 (0.348)	0.979 (0.045)	0.476 (0.284)

Note: The estimates and asymptotic standard errors for columns 2, 3, 5, and 6 are reproduced from Campbell and Shiller (1991): Table 1(a) for equation (2), Table 2 for equation (3), Table 3(a) for the Correlation, and Table 4(a) for the Standard Deviation Ratio). The monthly data are from McCulloch (1990), and the sample is from 1952:1 to 1987:2. Column 4 contains estimates and standard errors for equation (5) which were computed by the authors from the same data. Equation (2) is a regression of the change in the yield on an n-period bond on $[1/(n-1)]$ times the term spread between the long yield and the short rate. At all horizons other than $n = 2$, equation (2) is estimated using the approximation described in the text. Equation (3) is a regression of the weighted average of changes in future short rates on the term spread. Equation (5) is a regression of the future short rate minus the current short rate on the forward premium. In that regression, the short rate has a one-month maturity for $n=2$ and $n=12$, a 3-month maturity for $n=24$, a 6-month maturity for $n=36$, and a one-year maturity for the remaining regressions. The last two columns report statistics related to the implied term spread calculated from a VAR. They are the correlation between the implied spread and the actual spread and the ratio of the standard deviation of the implied spread to the standard deviation of the actual spread.

Table 2**A First-Order Autoregressive Model for the Short Interest Rate**

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

$$u(t) \sim N(0,1)$$

$\hat{\rho}$	0.9771
bias adj. ρ	0.9865
$\hat{\mu}$	0.1281
bias adj. μ	0.0755
$\hat{\sigma}$	0.6481
bias adj. σ	0.4988

Note: The sample is monthly data from 1952:1 to 1987:2. 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 $\sigma = \hat{\sigma}[(1 - \rho^2)/(1 - \hat{\rho}^2)]^{0.5}$.

Table 3
Monte Carlo Distribution of the Slope Coefficients

Panel A: Equation (2)						
n	Analytical Estimate	Mean	σ	1%	5%	10%
2	2.393	2.573	1.864	-0.313	0.222	0.561
12	2.362	2.538	1.822	-0.284	0.239	0.571
60	2.233	2.392	1.649	-0.162	0.311	0.612
120	2.109	2.252	1.483	-0.045	0.381	0.651
Panel B: Equation (2) (With Approximation Error)						
2	3.370	3.549	1.851	0.682	1.214	1.551
12	3.296	3.471	1.810	0.667	1.187	1.517
60	2.985	3.143	1.640	0.604	1.075	1.374
120	2.687	2.830	1.476	0.544	0.968	1.237
Panel C: Equation (3)						
2	1.697	1.788	0.933	0.330	0.613	0.778
12	1.681	1.762	0.865	0.357	0.639	0.797
60	1.617	1.668	0.606	0.362	0.724	0.904
120	1.555	1.585	0.419	0.433	0.777	1.001
Panel D: Equation (5)						
2	1.697	1.788	0.933	0.330	0.613	0.778
12	1.674	1.740	0.835	0.338	0.641	0.804
60	1.577	1.561	0.515	0.333	0.712	0.898
120	1.489	1.400	0.422	0.253	0.684	0.867

Panel E: VAR Statistics (Order = 1) Correlation Coefficient						
n	Estimate	Mean	σ	1%	5%	10%
2	1.0	0.915	0.102	0.562	0.701	0.767
12	1.0	0.997	0.007	0.974	0.987	0.991
60	1.0	0.9999	0.0003	0.999	0.9996	0.9997
120	1.0	0.99998	0.00007	0.9998	0.9999	0.99994

Panel F: VAR Statistics (Order = 1) Standard Deviation Ratio						
2	1.310	1.940	0.959	0.430	0.703	0.902
12	1.315	1.716	0.817	0.356	0.635	0.794
60	1.316	1.419	0.461	0.402	0.678	0.819
120	1.316	1.248	0.289	0.459	0.729	0.854

Panel G: VAR Statistics (Order = 4) Correlation Coefficient						
2		0.644	0.136	0.307	0.413	0.464
12		0.968	0.036	0.816	0.902	0.929
60		0.999	0.002	0.990	0.996	0.997
120		0.9997	0.0006	0.998	0.999	0.999

Panel H: VAR Statistics (Order = 4) Standard Deviation Ratio						
2		2.703	1.099	0.802	1.222	1.458
12		1.787	0.846	0.375	0.670	0.841
60		1.442	0.479	0.359	0.674	0.827
120		1.259	0.297	0.409	0.722	0.860

Note: The Monte Carlo evidence is based on 5,000 replications. The data generating process is equation (7) with the bias adjusted parameters from Table 2, and the sample size is 421. The horizon is n months. The column labelled Analytical Estimate contains the expected value of the distribution predicted by the analytical derivations in Section 2. The columns labelled Mean, σ , 1%, 5%, and 10% are the sample mean, the standard deviation and the respective quantiles of the empirical distribution.

Table 4

Monte Carlo Distributions of the Squared t-statistics for the Slope Coefficients

Panel A: Equation (2)					
n	Mean	σ	90%	95%	99%
all	1.597	2.038	4.219	5.652	9.060
Panel B: Equation (2) (With Approximation Error)					
2	2.942	2.690	6.517	8.177	11.891
12	2.899	2.677	6.460	8.118	11.833
60	2.707	2.612	6.162	7.814	11.509
120	2.496	2.534	5.858	7.460	11.028
Panel C: Equation (3)					
2	1.608	2.055	4.258	5.650	9.157
12	2.712	3.863	7.277	10.281	18.362
60	10.992	25.259	27.104	45.072	114.806
120	38.921	97.146	91.339	149.806	396.448
Panel D: Equation (5)					
2	1.608	2.055	4.258	5.650	9.157
12	2.959	4.298	7.829	11.484	20.624
60	7.698	13.213	19.937	30.651	61.665
120	12.102	23.881	30.543	50.401	109.506

Note: The Monte Carlo evidence is based on 5,000 replications. The data generating process is equation (7) with the bias adjusted parameters from Table 2, and the sample size is 421. The horizon is n months. The distributional characteristics should be compared to a $\chi^2(1)$ distribution. The characteristics of the $\chi^2(1)$ distribution are as follows: the mean is 1.0, the standard deviation is 1.414 and the critical values are 2.705 (90%), 3.841 (95%), and 6.635 (99%).

Table 5**Estimates of Parameters of the Regime-Switching Model**

μ_1	0.0526 (0.0342)
μ_2	0.2003 (0.1719)
β_1	0.9975 (0.0087)
β_2	0.9605 (0.0292)
σ_1	0.1784 (0.0237)
σ_2	0.4807 (0.0922)
γ_1	0.3803 (0.0906)
γ_2	0.4611 (0.0777)
a_1	5.5354 (2.4795)
b_1	-0.3557 (0.3092)
a_2	1.3549 (0.9419)
b_2	0.1272 (0.1011)

Note: The regime-switching model is equations (26)-(29). Data are monthly observations on the one-month interest rates from 1952:01 through 1987:02. Robust standard errors (see White (1982)) are in parentheses.

Table 6

**Monte Carlo Distributions of the Slope Coefficients
Using a Regime-Switching Model as the Data Generating Process**

Panel A: Equation (2)					
n	Mean	σ	1%	5%	10%
2	2.567	2.294	-0.164	0.207	0.446
12	2.685	2.339	-0.112	0.242	0.501
60	2.043	1.441	0.146	0.420	0.625
120	1.761	1.042	0.266	0.506	0.677
Panel B: Equation (2) (With Approximation Error)					
2	3.452	2.243	0.774	1.145	1.387
12	3.488	2.315	0.715	1.076	1.332
60	2.638	1.512	0.633	0.915	1.152
120	2.147	1.087	0.577	0.827	1.015
Panel C: Equation (3)					
2	1.784	1.149	0.397	0.603	0.720
12	1.704	0.865	0.429	0.660	0.806
60	1.411	0.449	0.525	0.774	0.895
120	1.308	0.232	0.536	0.851	0.974
Panel D: Equation (5)					
2	1.783	1.147	0.418	0.603	0.723
12	1.650	0.793	0.459	0.683	0.812
60	1.276	0.339	0.458	0.735	0.868
120	1.183	0.315	0.383	0.676	0.799

Panel E: VAR Statistics (Order = 1) Correlation Coefficient					
n	Mean	σ	1%	5%	10%
2	0.912	0.110	0.522	0.688	0.765
12	0.996	0.019	0.973	0.987	0.991
60	1.000	0.014	0.999	1.000	1.000
120	0.999	0.014	0.992	0.996	0.997

Panel F: VAR Statistics (Order = 1) Standard Deviation Ratio					
2	1.943	1.192	0.480	0.704	0.837
12	1.629	0.779	0.456	0.662	0.790
60	1.241	0.310	0.589	0.765	0.832
120	1.112	0.161	0.697	0.839	0.908

Note: The Monte Carlo evidence is based on 5,000 replications. The data generating process is the regime-switching model of equations (26)-(29) using the parameter estimates of Table 5. The horizon is n months. The columns labelled Mean, σ , 1%, 5%, and 10% are the sample mean, the standard deviation and the respective quantiles of the empirical distribution.

Appendix

Proof of Proposition 1:

Under the AR(1) data generating process of equation (7), the regressor in equation (2) can be written as

$$\frac{1}{(n-1)}[r(t,n) - r(t)] = \frac{1}{(n-1)}\left[\frac{(1-\rho^n)}{n(1-\rho)} - 1\right]r(t), \quad (\text{A1})$$

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

$$r(t+1, n-1) - r(t, n) = \frac{1}{(n-1)}\left[\frac{(1-\rho^n)}{n(1-\rho)} - 1\right]r(t) + \frac{(1-\rho^{n-1})}{(n-1)(1-\rho)}\varepsilon(t+1). \quad (\text{A2})$$

By the algebra of OLS, equations (A1) and (A2) imply

$$\hat{\alpha}_1 = 1 + \frac{-n(1-\rho^{n-1})}{n(1-\rho) - (1-\rho^n)} \left[\frac{\text{cov}_T(\varepsilon(t+1), r(t))}{\text{var}_T(r(t))} \right] \quad (\text{A3})$$

where cov_T and var_T denote sample second moments. The conclusion of Proposition 1 follows from (A3), along with the observation that, under the assumed data generating process (7),

$$E(\hat{\rho}) - \rho = \left[\frac{\text{cov}_T(\varepsilon(t+1), r(t))}{\text{var}_T(r(t))} \right]. \quad (\text{A4})$$

QED.

Proof of Proposition 2:

If constant maturity bonds are used to construct the dependent variable in equation (2), the dependent variable can be written as

$$r(t+1, n) - r(t, n) = \frac{1}{n}(\rho^n - 1)r(t) + \frac{(1-\rho^n)}{n(1-\rho)}\varepsilon(t+1). \quad (\text{A5})$$

The regressor is again given in equation (A1). The algebra of OLS then implies equations (11)-(12) in the text. QED.

Proof of Proposition 3:

From the AR(1) data generating process, the regressor in equation (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 equation (3) is the weighted average of future short rates minus the current short rate:

$$\sum_{i=1}^{n-1} \left[1 - \frac{i}{n} \right] [r(t+i) - r(t+i-1)] = \frac{1}{n} [r(t) + r(t+1) \dots + r(t+n-1)] - r(t). \quad (\text{A6})$$

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

$$\frac{1}{n\eta(\rho, n)} \sum_{j=0}^{n-1} \sum_{t=1}^{T-n+1} \frac{[r(t+j)r(t)]}{r(t)^2} - \frac{1}{\eta(\rho, n)} = 1 + \frac{1}{n\eta(\rho, n)} \sum_{j=1}^{n-1} \theta_j \quad (\text{A7})$$

because each of the bivariate OLS slope coefficient terms in equation (A6) can be written as the true j -th autocorrelation plus a bias term:

$$\sum_{t=1}^{T-n+1} \frac{r(t+j)r(t)}{r(t)^2} = \rho^j + \theta_j. \quad (\text{A8})$$

Equation (13) follows from combining terms in ρ^j to get $\eta(\rho, n)$ and rewriting the coefficient multiplying the sum of the bias terms. QED.

Proof of Proposition 4:

From the definition of the forward rate in equation (4) and using the data generating process in equation (7), we have the following:

$$f(t, n-1) - r(t) = [\rho^{n-1} - 1]r(t). \quad (\text{A9})$$

The estimated OLS slope coefficient in equation (5) can then be written as

$$\frac{1}{(\rho^{n-1} - 1)} \sum_{t=1}^{T-n+1} \frac{[r(t+n-1)r(t)]}{r(t)^2} - \frac{1}{(\rho^{n-1} - 1)} = 1 + \frac{-\theta_{n-1}}{(1 - \rho^{n-1})}. \quad (\text{A10})$$

The result also uses equation (A8) above. QED.

Proof of Proposition 5:

We write the theoretical spread as

$$s'(t, n) = x_n \Delta r(t) + y_n s(t, n) \quad (\text{A11})$$

with the x_n and y_n coefficients evaluated according to equation (6) 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:

$$\text{var}_T(\Delta r(t)) \approx 2(1 - \rho - Q)\text{var}_T(r(t)) \quad (\text{A12})$$

$$\text{var}_T(s(t, n)) \approx \eta(\rho, n)^2 \text{var}_T(r(t)) \quad (\text{A13})$$

$$\text{cov}_T(s(t, n), \Delta r(t)) \approx \eta(\rho, n)(1 - \rho - Q)\text{var}_T(r(t)), \quad (\text{A14})$$

where the approximation involves $\text{var}_T(r_{t-1}) \approx \text{var}_T(r_t)$. Equations (A12) and (A14) employ the definition of Q in equation (21) in the text. Then, the sample variance of the theoretical spread is

$$\text{var}_T(s'(t, n)) = \left\{ x_n^2 2(1 - \rho - Q) + y_n^2 \eta(\rho, n)^2 + 2x_n y_n \eta(\rho, n)(1 - \rho - Q) \right\} \text{var}_T(r(t)). \quad (\text{A15})$$

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 equations (A14) and (A15). 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 0, 1, and 0, respectively. The result is given in Proposition 6. 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). QED.

Proof of Proposition 6:

First, partition the VAR coefficient matrix A as

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad (\text{A16})$$

The estimated coefficients are given by:

$$\begin{bmatrix} \hat{a}_{11} \\ \hat{a}_{12} \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^T \Delta r(t)^2 & \eta(\rho, n) \sum_{t=1}^T \Delta r(t)r(t) \\ \eta(\rho, n) \sum_{t=1}^T \Delta r(t)r(t) & \eta(\rho, n)^2 \sum_{t=1}^T r(t)^2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum_{t=1}^T \Delta r(t)y^1(t+1) \\ \sum_{t=1}^T \eta(\rho, n)r(t)y^1(t+1) \end{bmatrix} \quad (\text{A17})$$

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 assumption:

$$\sum_{t=1}^T r(t)^2 = \sum_{t=1}^T r(t-1)^2. \quad (\text{A18})$$

The following results are also useful:

$$\frac{\sum_{t=1}^T \Delta r(t)^2}{\sum_{t=1}^T r(t)^2} = 2(1 - Q - \rho), \quad (\text{A19})$$

and

$$\frac{\sum_{t=1}^T \Delta r(t)r(t)}{\sum_{t=1}^T r(t)^2} = 1 - (Q + \rho). \quad (\text{A20})$$

Equations (A11)-(A13) imply:

$$D^* = \frac{D}{\sum r(t)^2} = \eta(\rho, n)^2 [1 - (Q + \rho)^2]. \quad (\text{A21})$$

With V defined as in equation (22) in the text, using the AR(1) data generating process to evaluate y_t^i , and using equations (A16)-(A21), equation (23) in the text follows from dividing the numerator and denominator of the expression in (A17) by $[\sum r(t)^2]^2$. The biases in the VAR parameters in Proposition 5 are found by taking the expected values of second-order Taylor's series expansions of equation (23) around the unconditional mean values of Q and V . QED.

References

- Bollerslev, T., R.Y. Chou, and K.F. Kroner, 1992, "ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence," Journal of Econometrics 52, 5-60.
- Campbell, J.Y., and R.J. Shiller, 1991, "Yield Spreads and Interest Rate Movements: A Bird's Eye View," Review of Economic Studies 58, 495-514.
- Elliott, G., and J.H. Stock, 1994, "Inference in Time Series Regression When the Order of Integration of a Regressor is Unknown," Econometric Theory 10, 672-700.
- Fama, E.F., 1984, "The Information in the Term Structure of Interest Rates," Journal of Financial Economics 13, 509-528.
- Fama, E.F., and R. Bliss, 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, New York: John Wiley & Sons.
- Gray, S.F., 1995a, "Modelling the Conditional Distribution of Interest Rates as a Regime-Switching Process," Duke University, manuscript.
- Gray, S.F., 1995b, An Analysis of Conditional Regime-Switching Models, Duke University, manuscript.
- Hamilton, J.D., 1988, "Rational Expectations Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates," Journal of Economic Dynamics and Control 12, 385-423.
- 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., and F. Mishkin, 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.
- Lutz, F.A., 1940, "The Structure of Interest Rates," Quarterly Journal of Economics, 55, 36-63.
- Mankiw, N.G., and M. Shapiro, 1986, "Do We Reject Too Often: Small Sample Properties of Tests of Rational Expectations Models," Economic Letters, 20, 139-145.
- Marriott, F., and J. Pope, 1954, "Bias in the Estimation of Autocorrelations," Biometrika 41, 390-402.
- McCulloch, J.H., 1990, "U.S. Term Structure Data, 1946-87," Handbook of Monetary Economics Volume I, pp. 672-715.
- Newey, W., and K. West, 1987, "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," Econometrica, 55, 703-708.
- Richardson M., and J.H. Stock, 1989, "Drawing Inferences from Statistics Based on Multi-Year Asset Returns," Journal of Financial Economics 25, 323-348.
- Stambaugh, R., 1986, "Bias in Regression with Lagged Stochastic Regressors," C.R.S.P. Working Paper No. 156, University of Chicago.
- 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," Chapter 46 in Handbook of Econometrics, Volume IV, Edited by R.F. Engle and D.L. McFadden, 2739-2841.
- White, H., 1982, "Maximum Likelihood Estimation of Misspecified Models," Econometrica 50, 1-25.

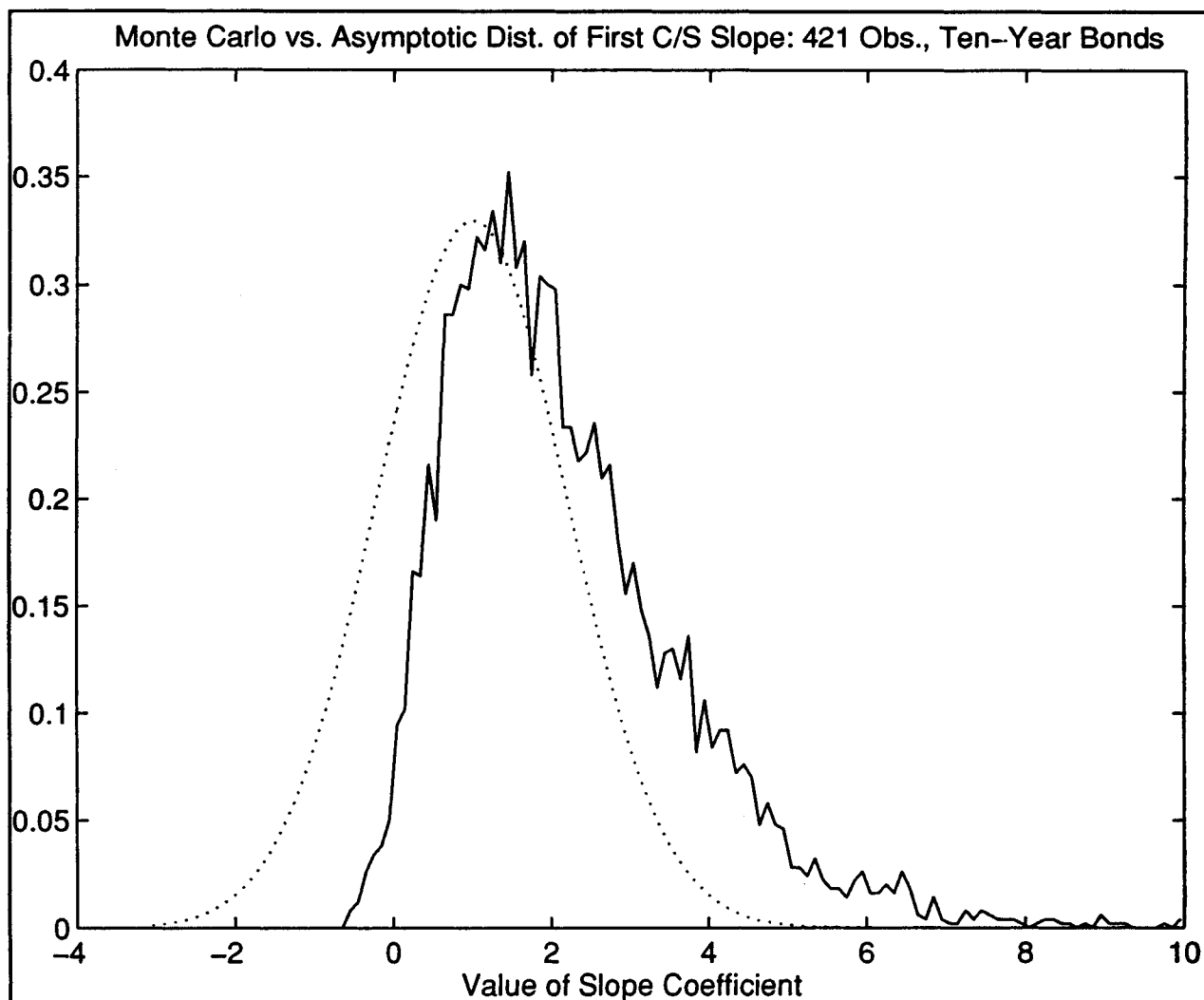


Figure 1 Monte Carlo Distribution vs. Asymptotic Distribution of First Campbell/Shiller Specification Statistic

The solid line displays the Monte Carlo density function for the OLS estimate of α_1 , the slope coefficient in equation (2), when the maturity of the long bond equals 120 months, the sample size equals 421 months, and the short interest rate is generated by the AR(1) process given in equation (7) with the bias-adjusted parameters from Table 2. The Monte Carlo evidence is based on 5,000 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 to equal 0.10, the average Newey-West standard error over the 5,000 Monte Carlo replications. In computing σ_α , a single Newey-West lag was used.

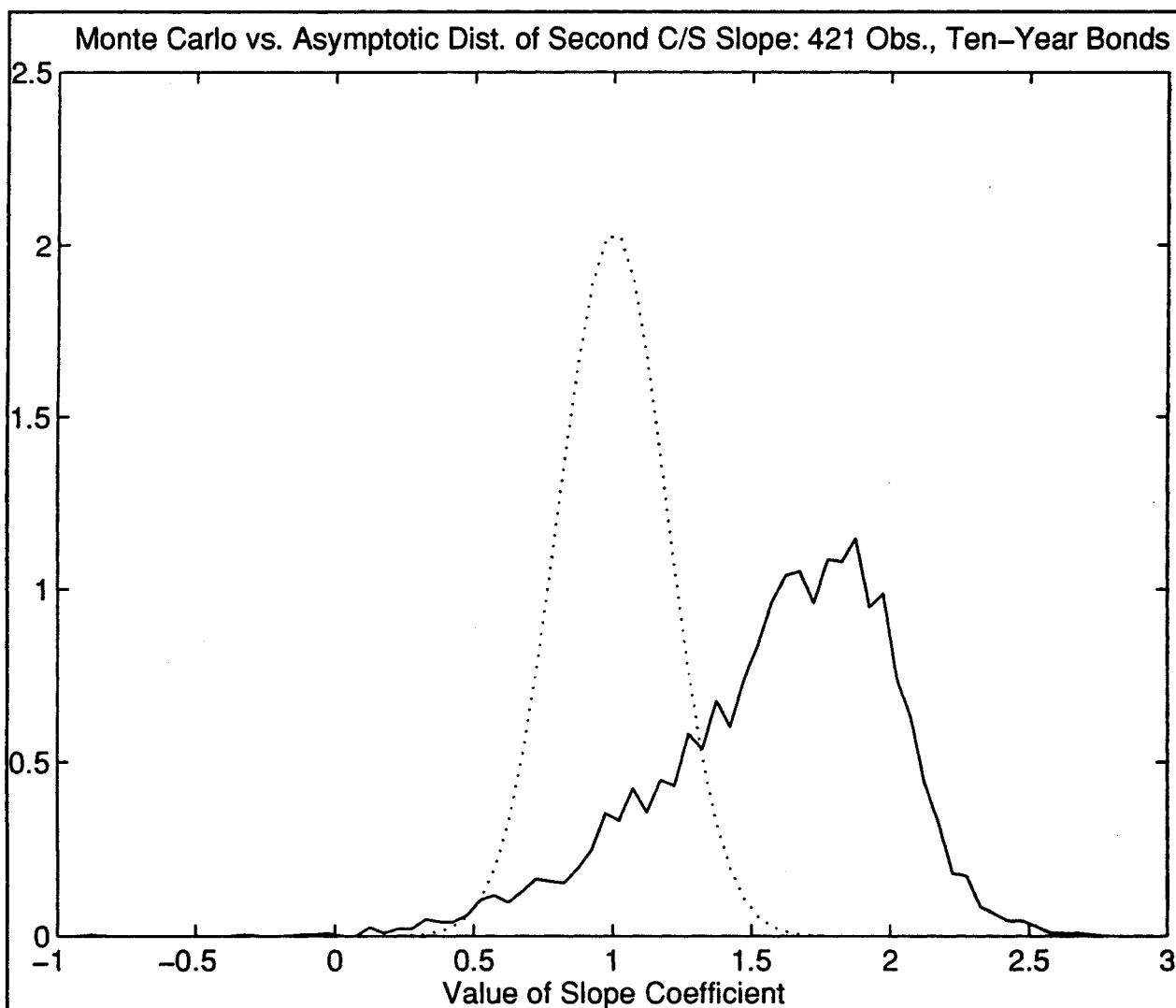


Figure 2 Monte Carlo Distribution vs. Asymptotic Distribution of Second Campbell/Shiller Specification Statistic

The solid line displays the Monte Carlo density function for the OLS estimate of δ_1 , the slope coefficient in equation (3), when the maturity of the long bond equals 120 months, the sample size equals 421 months, and the short interest rate is generated by the AR(1) process given in equation (7) with the bias-adjusted parameters from Table 2. The Monte Carlo evidence is based on 5,000 replications. The dotted line displays the density implied by the asymptotic approximation $\sqrt{T}[\delta_1 - 1] \sim \text{normal}(0, T\sigma_\delta^2)$. The asymptotic standard error σ_δ was set to equal 0.196, the average Newey-West standard error over the 5,000 Monte Carlo replications. In computing σ_δ , 120 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).

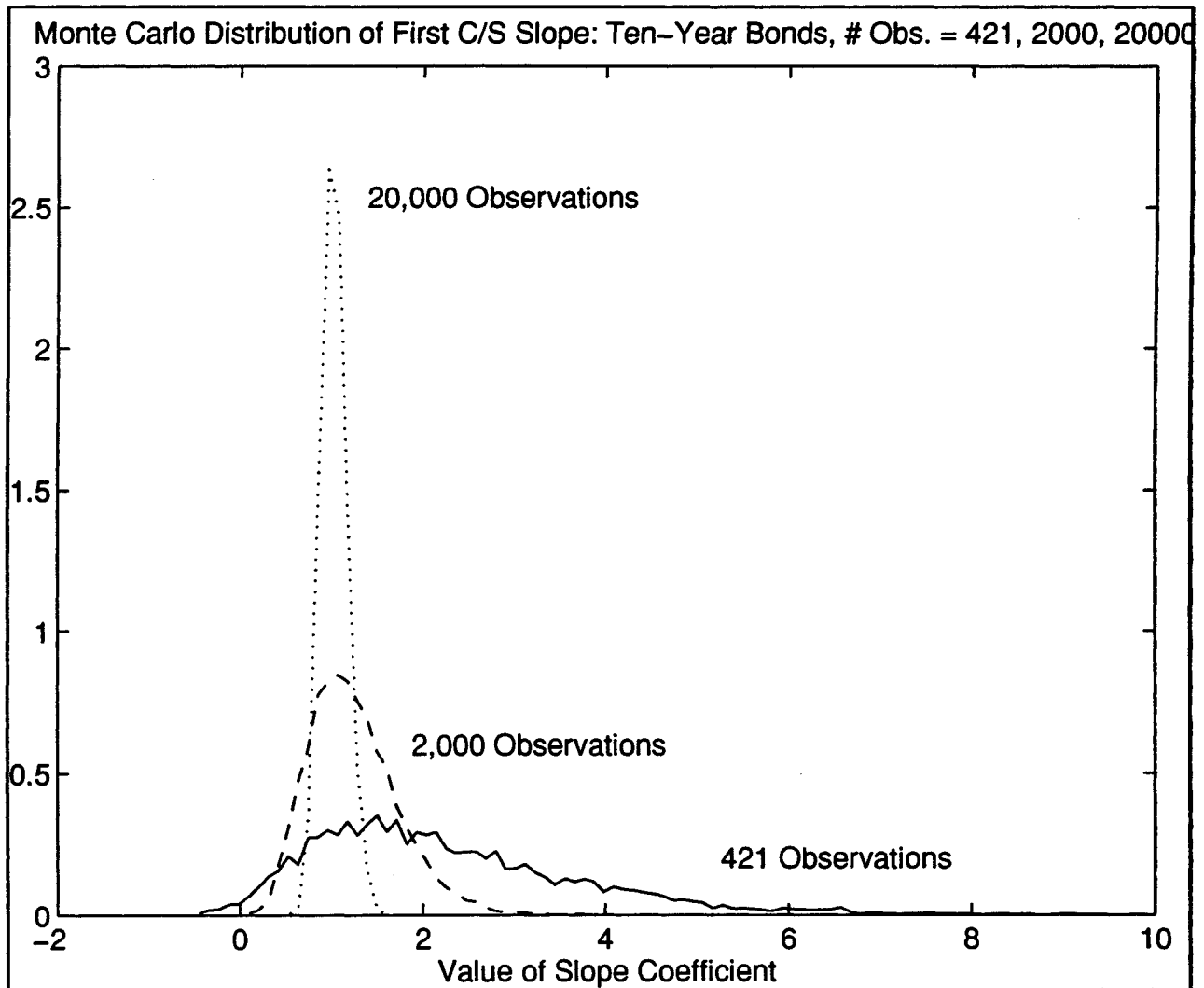


Figure 3 Monte Carlo Distributions of First Campbell/Shiller Specification Statistic as Sample Size Increases

This figure displays three Monte Carlo density functions for the OLS estimate of α_1 , the slope coefficient in equation (2), when the maturity of the long bond equals 120 months and the short interest rate is generated by the AR(1) process given in equation (7) with the bias-adjusted parameters from Table 2. The solid line displays the density function with 421 monthly observations. (This plot is identical to the solid line in Figure 1.) The dashed line is the corresponding density with 2,000 observations, and the dotted line is the density with 20,000 observations. This Monte Carlo evidence is based on 5,000 replications.

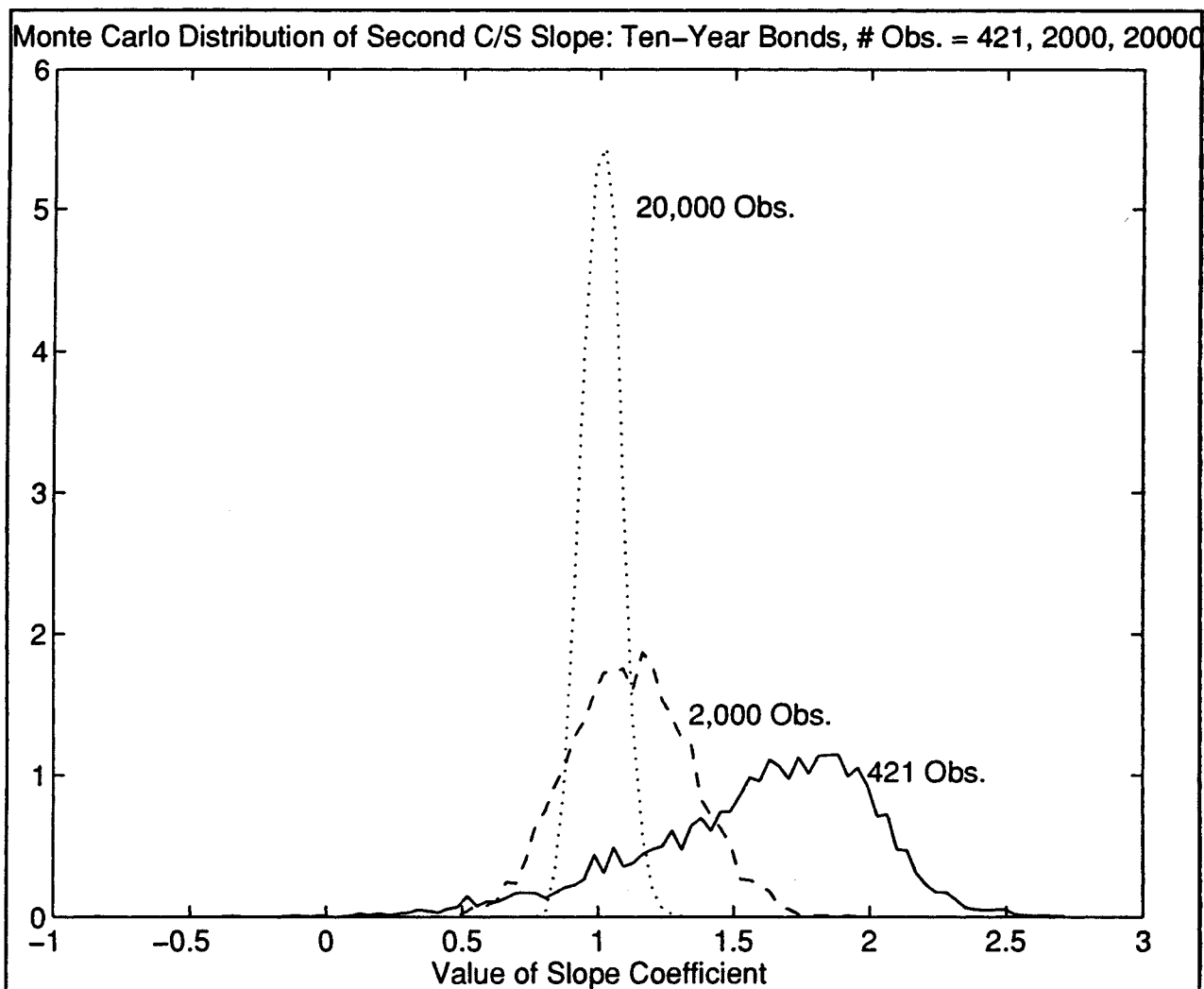


Figure 4 Monte Carlo Distributions of Second Campbell/Shiller Specification Statistic as Sample Size Increases

This figure displays three Monte Carlo density functions for the OLS estimate of δ_1 , the slope coefficient in equation (3), when the maturity of the long bond equals 120 months and the short interest rate is generated by the AR(1) process given in equation (7) with the bias-adjusted parameters from Table 2. The solid line displays the density function with 421 monthly observations. (This plot is identical to the solid line in Figure 2.) The dashed line is the corresponding density with 2,000 observations, and the dotted line is the density with 20,000 observations. This Monte Carlo evidence is based on 5,000 replications.