International Finance Discussion Papers

Number 258

July 1985

LONG MEMORY MODELS OF INTEREST RATES, THE TERM STRUCTURE, AND VARIANCE BOUNDS TESTS

by

Gary S. Shea

Abstract

Variance bounds tests of the rational expectations hypothesis of the interest rate term structure are sensitive to the stochastic characterization of short-term interest rates used. When a long memory or fractional difference nonstationary time series model is used in preference to a mean stationary model, the rational expectations hypothesis is not rejected. Long memory models of interest rates are estimated and tested against alternatives. Their forecasting properties are also examined. Hypothesis tests are based upon bootstrapping (Monte Carlo) methodologies.

Long Memory Models of Interest Rates,

the Term Structure, and Variance Bounds Tests

by

Gary S. Shea*

I. Introduction

Equilibrium models of the interest rate term structure under linear rational expectations are rejected by a number of researchers upon the basis of empirical evidence that long-term interest rates and holding period yields are too volatile relative to short-term interest rates. Prominent among those who believe so are Shiller (1979) and Singleton (1980). Their views echo the original arguments of Culbertson (1957) that interest rates are not smooth enough to coincide with the averaging of expected future short-term rates which are equal to long-term interest rates under the linear rational expectations hypothesis. Without resort to a term premium or without direct appeal to risk aversion (Michener 1982), it has been difficult to reconcile observed violations of variance bounds with linear rational expectations models of asset pricing. Empirical support for these views is also found in papers on excess volatility in the stock markets by Leroy and Porter (1981) and Shiller and Grossman (1981).

The empirical methodologies used in these studies have been questioned, however, and it is becoming clear that a firm empirical rejection of the rational expectations hypothesis is not yet at hand. The primary reason for this is the questionable assumption commonly found in the excess volatiltity literature that asset returns are mean stationary linear stochastic processes having finite variance. With that assumption, rational expectations models of the term structure will imply particular variance relationships between expected future returns, certain short-term interest rates, and other returns, such as yields to maturity or holding period yields. Although most writers have explicitly admitted that the assumed stationarity of short-term interest rates or dividend flows is crucial to the working of their tests (a catalog of such admissions is cited in Flavin's (1983) concluding remarks), there is only a little dissatisfaction with the assumption yet evident in the literature. This is unfortunate because, even if the assumption is valid, there are practical problems in implementing reliable hypothesis tests. Flavin, for example, shows that it is difficult to reject the expectations hypothesis using the small samples used by Shiller and Singleton in their studies.

In contradiction to the stationarity assumption, there are many studies which have found interest rate time series to be best characterized as nonstationary series. Nelson and Plosser (1982) found interest rates to be perhaps nonstationary as did Brick and Thompson (1978) and Cargill (1975). Moreover, Granger (1966) in his survey of the spectral shapes of economic time series found that many typically have very large power at low frequencies implying that good models for what is usually called trend are important components in any characterization of these time series. In concentrating upon interest rates and their typical spectral shape, Granger and Rees (1968) concluded that low frequency components were particularly powerful in the spectra of the time series they examined. In response to the Leroy and Porter and the

Shiller and Grossman studies, Marsh and Merton (1984) have also questioned the assumption of whether dividend flows are adequately described as stationary stochastic processes.

If interest rate time series are nonstationary, would there be interesting implications for how excess volatility tests are applied to equilibrium notions of the term structure? This paper answers that in the affirmative and introduces a particular class of discrete linear nonstationary stochastic processes which are better models of short-term interest rates than either the random walk or the discrete stationary linear models. These models are the general integrated models investigated by Granger and Joyeux (1980), Geweke and Porter-Hudak (1983) and Mandelbrot (1972) and are otherwise known as fractional difference or long memory time series models. In the remainder of this paper there are developed a number of general linear integrated models for short-term interest rates, which are estimated and then compared to alternatives. The variation of short-term rates relative to that of holding period yields for longer term bonds under the linear rational expectations hypothesis is examined. It is shown that, joint with the estimated general integrated model for short-term rates, the linear rational expectations hypothesis cannot be rejected using a variance bounds test. The controversial bootstrap methodology is used to construct confidence intervals for hypothesis testing.

II. A General Integrated Model of Short-Term Interest Rates

I assume that short-term interest rates can be represented by a discrete linear stochastic model of the form

$$r_t = \psi(B)\varepsilon_t \, , \tag{1}$$

where r_t is a short-term rate, ε_t is an i.i.d. normal random variable and $\psi(B)$ is a polynominal (possibly of infinite order) in the backshift operator B which operates on the subscript t. Equation (1) is the moving average form of the model and the coefficients of $\psi(B)$ will be called the moving average coefficients. Equation (1) can serve as a model for a nonstationary interest rate time series and I shall assume that a differencing filter of the form $(1-B)^d$, when applied to r_t, will render the interest rate series stationary. The stationary component of the time series can then be modelled as an ARMA process.

Economists are used to assuming some prior knowledge of d, the order of the differencing filter $(1-B)^d$. At most, the assumed value of d has been tested against an alternative which is usually 0 or 1 when the possibility of a unit root in the difference equation representation of a time series arises. In any case, we are mostly used to thinking of ARMA parameter estimates as conditional upon known integer values of d.

The general integrated model, however, allows for noninteger values of d and thus gives the modeller some more flexibility in the representation of the low frequency components of a time series. In testing for excess volatility this is obviously important since improper allowance for low frequency components in a time series will affect the reliability with which long-term forecasts can be made and will certainly alter the expression of theoretical variance bounds on yields from long-term bonds.

Proceeding to the expression of the general integrated model, we let

$$\psi(B) = \phi^{-1}(B)\theta(B)(1 - B)^{-d}, \tag{2}$$

where $\phi(B)$ and $\theta(B)$ are finite-order polynomials in the backshift operator B and constitute the short-term ARMA model of the process, so that

$$\phi(B)(1 - B)^{d} r_t = \theta(B)\varepsilon_t . \tag{3}$$

This is the complete linear general integrated model.

We now adopt the notation used by Geweke and Porter-Hudak (1983) to express the model in the frequency domain in which some of the model's features are more comprehensible and where the estimation of d takes place. Letting

$$u_t = \phi(B)\theta^{-1}(B)r_t,$$

the spectral density of r_t is

$$f(\lambda{:}d)f_u(\lambda) = (\sigma^2/2\pi)|1 - e^{-i\lambda}|^{-2d} f_u(\lambda), \tag{4}$$

where $f_u(\lambda)$ is the spectral density of u_t. Alternatively, $f(\lambda:d)$ may be expressed (Granger and Joyeux 1980) as

$$(\frac{\sigma}{2\pi})^2 \ [2\ (1 - \cos\lambda)]^{-2d} \quad , \tag{5}$$

which clearly gets very large as $\lambda \to 0$. On the other hand, the spectral density of the differenced series is

$$|\ 1 - e^{-i\lambda}|^2 \ f(\lambda:d) = (\frac{\sigma^2}{2\pi}) \ [2(1 - \cos\lambda)]^{2(1 - d)} \tag{6}$$

which goes to zero as $\lambda \to 0$ and $d \neq 1$. Thus, for a time series having spectral density (4) with fractional d, the econometrician runs the risk of losing low frequency components of the series if he approaches the problem of achieving stationarity in the conventional manner by taking an integer-order difference of the series. Only for $|d| < .5$ will the time series be stationary. That is, for $|d| > .5$, $\sum_{k=0}^{\infty} \psi^2(B)$ is unbounded (Granger and Joyeux 1980) and for those values of d we must be able to prefilter the data prior to model estimation.

Applying a differencing filter of fractional order is not as straightforward as with integer-order prefilters. To see this consider the expansion of $(1-B)^d$ when d is integer. The autoregressive representation of a time series, y_t, such that

$$(1 - B)^d \ y_t = \varepsilon_t \quad \text{is} \tag{7}$$

$$y_t = d\, y_{t-1} - \frac{1}{2} d\,(d - 1)\, y_{t-2} + \ldots + (-1)^{d+1}\, y_{t-d} + \varepsilon_t. \quad (8)$$

For fractional d, however, the autoregressive representation is of infinite order and is expressed

$$y_t = d \sum_{k=1}^{\infty} \frac{\Gamma(k - d)\, B^k}{\Gamma(1 - d)\, \Gamma(k + 1)}\, y_t + \varepsilon_t. \quad (9)$$

Γ is the gamma or generalized factorial function. So, direct application of a fractional differencing filter will in principle entail the use of a long approximating autoregressive filter which may be prohibitively consumptive of scarce data. The approximation itself may also be less than satisfactory since the autoregressive coefficients get small only at long lags, depending on the value of d (Granger and Joyeux 1980; Geweke and Porter-Hudak 1983); this is a manifestation of the potential long memory characteristics of these models.

We shall now take a preliminary look at the data to see if traces of a general linear integrated model can be found there. The data described in the Appendix are a long time series of continuous term structure observations stretching from 1951 through 1964. Compared to what was to follow, this was a quiescent period for interest rates, so if a nonstationary model is not rejected against a stationary model, that will be telling evidence for the general nonstationary character of interest rates. A glance at Figure 1 suggests that the yields on Treasury Bills with one week to maturity have the 'typical' spectral

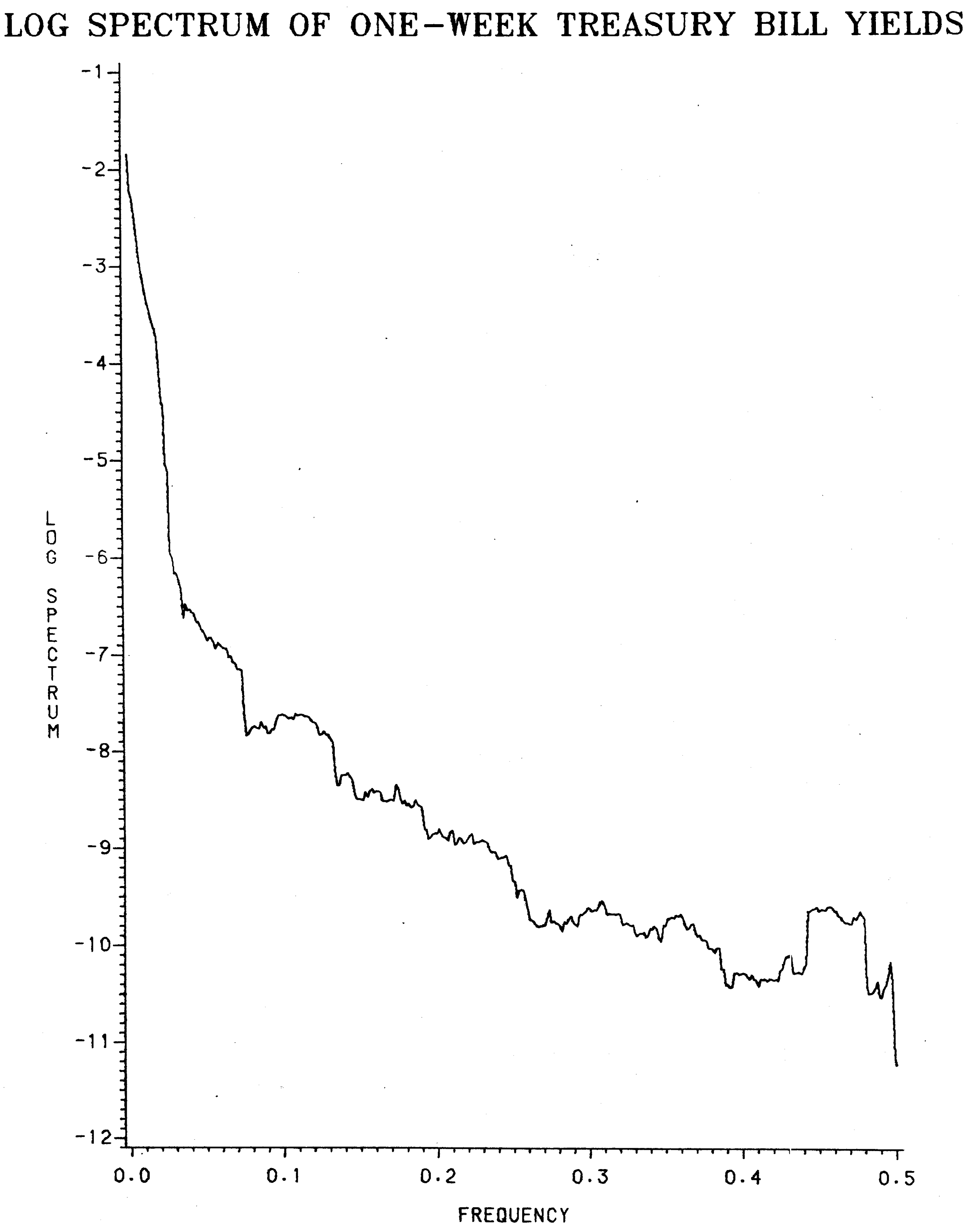
LOG SPECTRUM OF ONE-WEEK TREASURY BILL YIELDS
LOG SPECTRUM
FREQUENCY
-1
-2
-3
-4
-5
-6
-7
-8
-9
-10
-11
-12
0.0
0.1
0.2
0.3
0.4
0.5

Figure 1

shape described by Granger. For interest rate levels the spectrum is very large and perhaps explosive which greatly obscures other features of the spectrum at low frequencies due to leakage. On the other hand, the spectrum of the differenced series in Figure 2 clearly goes to zero at the lowest frequencies instead of flattening which is generally symptomatic of overdifferencing. Reflecting on his studies after more than a decade, Granger (1981) was willing to treat this as an additional feature of an economic time series having the 'typical' spectral shape. Finally, in Figure 3 is the estimated autocorrelation function out to 104 lags. The autocorrelations die out very slowly as we would expect for a time series with 'long memory.'

This particular time series of interest rates, not unlike many others, appears to be nonstationary. Its nonstationary character is somewhat shy of that possessed by a random walk, however. In the next section we estimate a fractional difference filter and accompanying ARMA model to deal with the data's peculiar character.

III. Selecting a General Linear Integrated Model

The previous discussion raises several expectations for an estimated fractional difference model for short-term interest rates. We would first naturally expect to get a reliable estimate of the order of the differencing filter and to be able to ascertain whether it is statistically different from 0, 1 or some other value. Second, a general integrated model is supposed to preserve the low frequency components of a time series and to use them to greater effect for prediction than do

SPECTRUM OF DIFFERENCED ONE-WEEK YIELDS

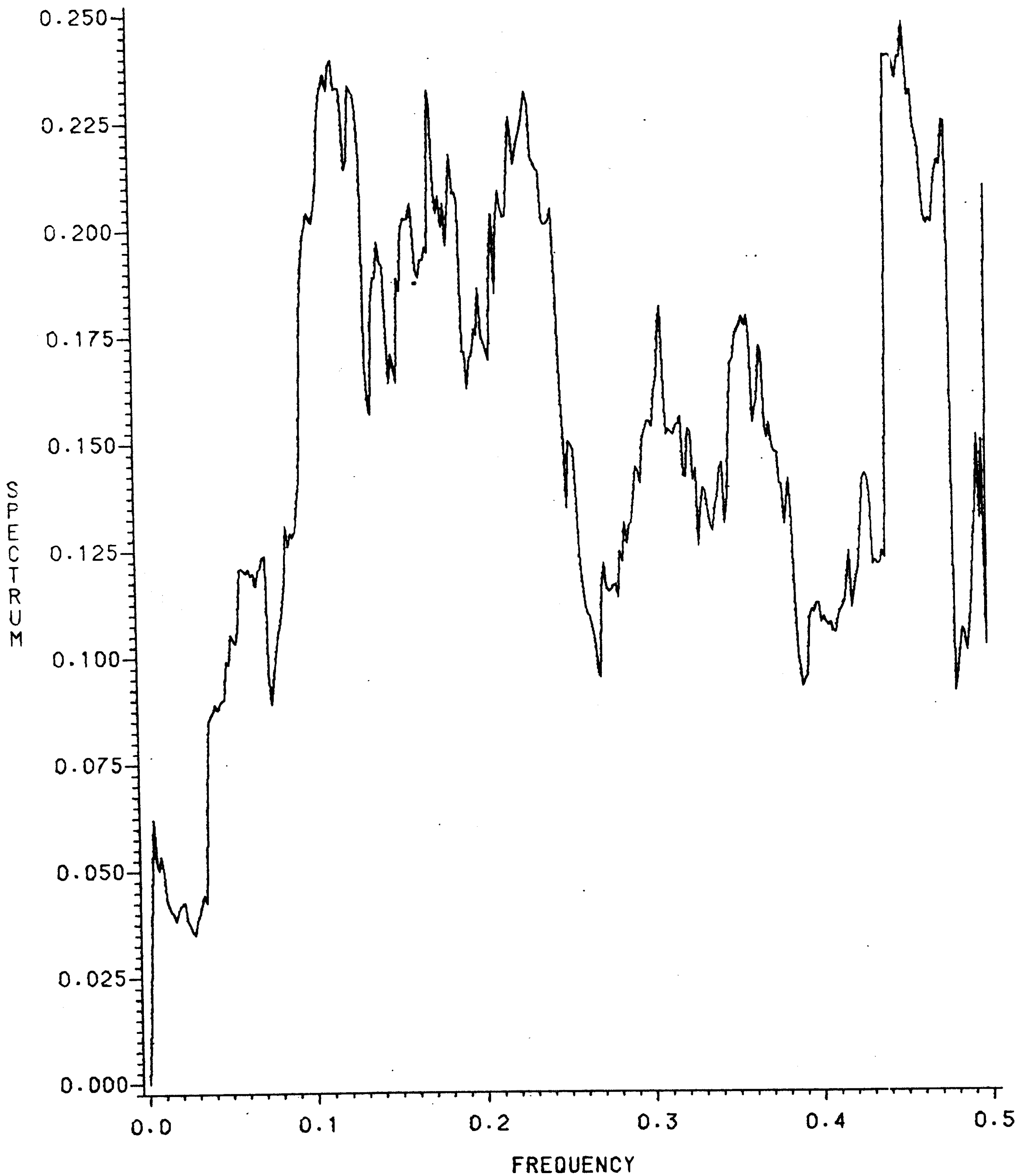

Figure 2

AUTOCORRELATIONS OF ONE-WEEK YIELDS

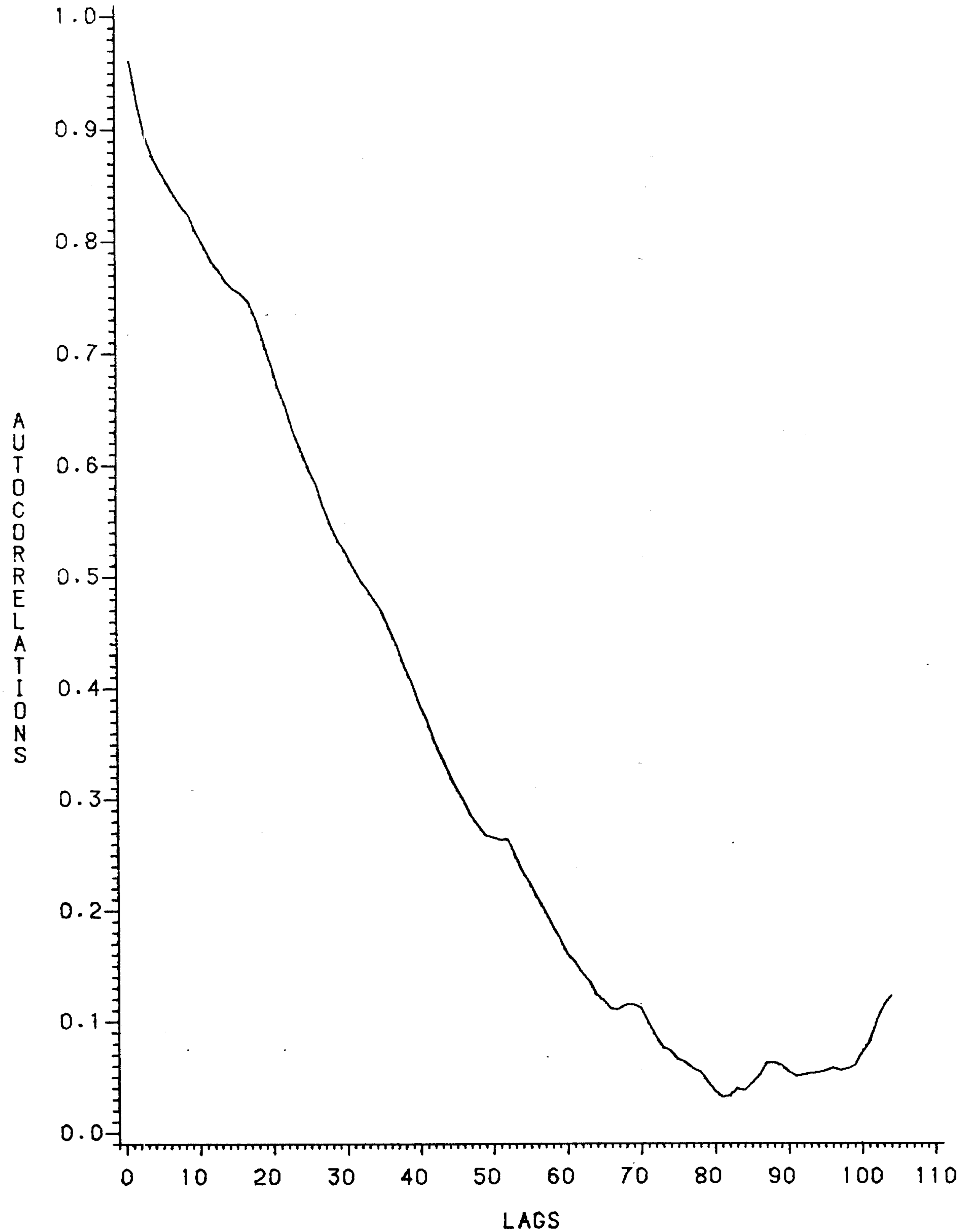

Figure 3

other ARIMA models which impose an incorrect order of differencing on the time series. We would therefore expect detectably better forecasting performance from a general integrated model and especially better long-term forecasting capabilities if the time series truly has 'long memory'. In this section we estimate several such long memory models and select some to address the ultimate objective of this paper, a variance bounds test for the linear rational expectations theory of the term structure. Forecasting capabilities are an important criterion in the model selection process. By this means it is hoped the reader will be convinced that only by using suboptimal long-term linear forecasting models for interest rates will the rational expectations hypothesis be rejected.

Identification, estimation, forecasting and hypothesis testing with general linear integrated models present some special difficulties. For example, in trying to identify a preliminary time series model, inspection of the autocorrelation function of a general integrated series is not as helpful in obtaining preliminary parameter estimates as is the case when trying to identify an ARMA model when d is assumed to be integer and known. This is because the autocorrelations are complex combinations of d and the ARMA parameters and cannot be readily factored into simple expressions suggestive of the model's structure (Porter-Hudak 1982). This complexity also presents problems for estimation. Granger and Joyeux (1980) and Porter-Hudak (1982) review estimation theory for these models, most of which is designed to circumvent the problem of not being able to write down a manageable likelihood function for a fractional differencing filter. A line of attack hinted at by Granger and Joyeux (1980) was extensively developed by Geweke and Porter-Hudak (1983) and we use it below to estimate our long memory models.

By taking logarithms of both sides of equation (4) and by adding log $f_u(0)$ to both sides Geweke and Porter-Hudak obtained

$$\log f(\lambda) = \log\{\sigma^2 f_u(0)/2\pi\} - d \log\{4\sin^2(\lambda/2)\} + \log\{f_u(\lambda)/f_u(0)\}. \tag{10}$$

The log periodgram (log $I(\lambda)$) evaluated at the sample's harmonic frequencies can be added to both sides of (10) to obtain

$$\log\{I(\lambda)\} = \log\{\sigma^2 f_u(0)/2\pi\} - d \log\{4\sin^2(\lambda/2)\} + \log\{f_u(\lambda)/f_u(0)\} + \log\{I(\lambda)/f(\lambda)\}. \tag{11}$$

The theory for estimating d was developed by noting that $\log\{I(\lambda)/f(\lambda)\}$ is independently distributed across the harmonic frequencies and that, for the first n harmonic frequencies such that n satisfies a function, g, of sample size T so that $\lim_{T\to\infty} g(T)/T = 0$ and $\lim_{T\to\infty} (\log T)^2/g(T) = 0$, $\log\{f_u(\lambda)/f_u(0)\}$ is small compared to $\log\{\sigma^2 f_u(0)/2\pi\}$ in (11). In these circumstances use of (11) as a regression equation will give consistent estimates of d (for $d < 0$) without account being taken of the other ARMA parameters. Moreover, the theoretical error's variance for this regression is known and is $\pi^2/6$. Simulation studies by Geweke and Porter-Hudak suggest that these theoretical results extend to cases in which $0 < d < 1/2$.

This method of estimation for d has been applied to the series of short-term interest rates from the Appendix. We have seen that the series clearly is a candidate for nonstationarity and, from our differencing experiment, that d probably is between .5 and 1. Prior to estimation the series was first put through the filter $(1-B)^{.5}$ so that the filtered series has bounded variance. The differencing method was that suggested by Geweke and Porter-Hudak (1983) and consists of applying the discrete Fourier transformation to the interest rates and then applying the difference filter $(1 - e^{-i\lambda})^{.5}$, which is the analog of $(1 - B)^{.5}$ in the frequency domain. The inverse Fourier transformation will then give back the fractionally differenced time series of interest rates.

Following a suggestion by Geweke and Porter-Hudak, g(T) was chosen to be T^{α} where α varied between .5 and .7. The suggested functional form for g certainly satisfies the necessary conditions for consistent estimates of d. Values of α between .5 and .7 were found by the authors to give the most stable estimates and also the more reliable estimates of d based upon Monte Carlo simulations. Certainly the specification of g(T) and its parameters are judgmental exercises at the current stage of development for this estimation theory, but our experiences with g(T) experimentation seemed to have paralleled those of Geweke and Porter-Hudak and we have accepted their guidelines as reliable for the investigations undertaken here.

With a sample size of T = 714 (11) was used as a regression equation for α between .64 and .75. In these instances estimated d varied between .70 and .87. Of course many other trial regressions were performed, but as will be shown, the combination of results with respect to parameter significance and the forecast worthiness of the long memory model suggest the best values for fractional estimated d's belong to models with d precisely in the range of .70 to .87.

In experiments with the data prefiltered by $(1-B)^d$, $.70<d<.87$, several preliminary ARMA models were identified by inspection of the filtered data's estimated autocorrelation and partial autocorrelation functions. What emerged from these experiments was a multiplicative, seasonal autoregressive model of the form

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4)(1 - \phi_{26} B^{26})(1 - \phi_{52} B^{52})(1-B)^d r_t = \varepsilon_t. \quad (12)$$

Since d is unknown and has to be estimated, standard errors of regression and significance of autoregressive parameters in (12) are difficult to judge. The approach we take in this paper is to simulate the sampling distributions for d and other parameters using the nonparametric bootstrapping methods described by Freedman and Peters (1984) and to study the distributions of forecast errors using the same methods described by Peters and Freedman (1984). In the next section it will also be shown that there is no other practical way to get at confidence intervals for some crucial test statistics without resort to these Monte Carlo methods. The bootstrapping methods are appropriate for simulating confidence intervals in ordinary least squares situations such as in the estimation of d and for models with simple dynamics such as the autoregressions in equation (12).

In the course of the bootstrap experiments, several notable features of the model came to light. When d is fixed, the estimated standard errors for ϕ_2, ϕ_4 and ϕ_{26} at first made it appear that these were statistically significant parts of the model, but the bootstrapped standard errors seemed to demonstrate that their presence was illusory. Moreover, for α less than .67 or thereabouts, ϕ_1 did not seem to be significant whereas for α greater than .67, ϕ_3 did not appear to be significant. Tables 1 and 2 therefore display the results of estimation of the models

$$(1 - \phi_1 B)(1 - \phi_{52}B^{52})(1- B)^d r_t = \varepsilon_t \tag{13}$$

for values of α between .67 and .75 and

$$(1 - \phi_3 B^3)(1 - \phi_{52}B^{52})(1- B)^d r_t = \varepsilon_t \tag{14}$$

for values of α between .64 and .66. Using (11) as a regression equation and letting $\nu = \log \{ I(\lambda)/f(\lambda) \}$, Table 1 presents estimated d and estimated σ^2_ν and their associated statistics for each of the two respective ARMA models for selected values of α. Table 2 similarly displays results for estimated ϕ_1, ϕ_3, and ϕ_{52}.

The bootstrap starts by saving the fitted residuals. From the original 714 post Treasury Accord observations of Treasury Bill yield curves 200 observations have been saved for the out-of-sample forecasting experiments to follow. The empirical distribution of fitted residuals is

TABLE 1

Geweke-Porter-Hudak Regression (T=514, with 1000 bootstrap replications)

		σ_ν^2			d				
			Bootstrap				Bootstrap		
Model	α	Estimate	Mean	SD	Estimate	Nominal SE	Mean	SD	RMS Nominal SE
I.	.67	1.84	1.70	.42	.76	.11	.81	.10	.10
II.	.69	1.76	1.69	.41	.73	.10	.78	.09	.09
III.	.71	1.72	1.70	.38	.72	.09	.78	.09	.09
IV.	.73	1.66	1.71	.36	.70	.08	.77	.08	.08
V.	.75	1.73	1.70	.33	.70	.08	.78	.08	.08
VI.	.64	1.84	1.71	.47	.87	.12	.85	.11	.11
VII.	.65	1.78	1.70	.47	.84	.11	.83	.11	.11
VIII.	.66	1.90	1.67	.43	.75	.11	.73	.10	.10

then used as a sampling distribution with equal mass assigned to each member. Each model (13) and (14) is then used to simulate data for 714 periods. The simulation model in each case was an autoregressive approximation to (13) or (14). The filter $(1-B)^{\tilde{d}}$ was approximated by a truncated version of equation (9). This filter was then convoluted with $(1-\tilde{\phi}_1 B)$, $(1-\tilde{\phi}_3 B^3)$, or $(1-\tilde{\phi}_{52} B^{52})$ as the situation required. In all examples to follow the autoregressive approximating polynomial was of degree 56. Generally, 800 fitted residuals were sampled from the empirical distribution to start up the simulation model and to obtain 714 simulated observations.

The simulated data, which are denoted r_t^*, were then used to reestimate the model. Parameter estimates resulting from the use of this data are denoted $\tilde{\sigma}_\nu^{*2}$, $\tilde{d}^*$, $\tilde{\phi}_1^*$, $\tilde{\phi}_3^*$, and $\tilde{\phi}_{52}^*$. The bootstrap procedure was repeated 1000 times in all cases presented here. In Tables 1 and 2 columns containing sample estimates are simply labelled "estimate." The mean and standard deviation of bootstrapped sampling distributions for $\tilde{d}^*$ and $\tilde{\phi}_1^*$, $\tilde{\phi}_3^*$, and $\tilde{\phi}_{52}^*$ are also provided. Finally, the estimated standard error for $\tilde{d}$ using the conventional formula under ordinary least squares and the estimated standard errors for the autoregressive parameters using the diagonal of the inverse information matrix from a maximum likelihood estimation procedure (Box and Jenkins 1976, p.227) are presented. These are called the nominal standard errors and their repeated calculation in the bootstrap estimation process allows us to calculate their root mean square. It is important to note that the bootstrap error terms, ε^*, are independently sampled from the common distribution of fitted residuals, $\tilde{\varepsilon}$. The sampled ε^*, however, are used only to simulate the bootstrapped data, r_t^*. Sample nominal standard errors are formed by using $\tilde{\varepsilon}$ and the bootstrap standard errors are formed by using $\tilde{\varepsilon}^*$, the fitted errors from each bootstrap replication.

Models I through V correspond to equation (13) and models VI through VIII correspond to equation (14). In Table 1 estimated σ_ν^2 in most all cases is very close to its theoretical value of $\pi^2/6 = 1.645$ and to the mean of its bootstrapped sampling distribution. Estimated d also seems to be well within its bootstrapped sampling distribution. In each of the simulated worlds represented by these models, nominal standard errors are close to their 'true' bootstrapped standard deviations and, moreover, are probably reliable guides for hypothesis testing judging from the similarity between estimated d's bootstrapped standard deviations and the root mean square nominal standard errors. This answers in part the question raised by Geweke and Porter-Hudak (1983) as to whether it is sufficient to use conventional formulas for calculating standard errors or should the known characteristics of the theoretical distribution of $\nu = \log \{ I(\lambda)/f(\lambda) \}$ be used instead.

Table 1 would seem to suggest that a good case can be made for d lying between .5 and 1. It was found in several experiments that estimated d was inversely related to α until α was between .70 and .75, at which point values of estimated d were stable and apparently different from either .5 or 1. There are instances in Table 1 in which estimated d or the mean of its bootstrapped sampling distribution are not too far from 1 judging by the standard errors, but these instances are associated with models of detectably inferior forecasting performance, as will be shown. In no case, on the other hand, does it appear that estimated or simulated d is close to .5. Experiments with different prefilters, values of α and ARMA model specifications yielded more or less the same results; sample and bootstrapped estimates of d were always well bounded away from .5. Therefore, this particular series of short-term interest rates, r_t, is of the nonstationary type even though it may not be possible to reject the hypothesis that d is less than 1.

Treating d as if it were fixed and known can at times give misleading estimates for standard errors for the ARMA parameters. This is particularly true for lower order autoregressive parameters which, as shown in Table 2, generally have estimated standard errors that overstate their significance. In the simulated bootstrap world the estimated first order autoregression coefficients are lower than those estimated from the sample and their standard deviations demonstrate there is a considerable chance they are lower still. This is not quite the case for the third order autoregressive coefficients although in this instance too it is not apparent that a correct ARMA representation of r_t or r_t^* have any low order autoregressive components. For the seasonal parameter estimates, however, the nominal standard errors are close to their 'true' values in the bootstrap experiments and strongly support the inclusion of the seasonal autoregressive parameter in an ARMA representation of r_t and r_t^*. The simulation experiments presented here can be altered, of course, to further reflect some of the assumptions inherent in the estimation procedures for the autoregressive models. We could abandon the nonparametric bootstrap in favor of a parametric bootstrap experiment instead. In particular, this would mean that bootstrap error terms, ε^*, will be sampled form a normal distribution having the estimated sample standard error; this is consistent with the maximum likelihood procedure used in estimation of the autoregressive coefficients. From the limited experiments carried out here, however, it is fairly clear that sampling error may have played a role in the appearance of nonexistent low-order autoregressive parameters in these models. On the other hand, a seasonal autoregressive parameter and an order of differencing between .5 and 1 are tenable components of a general linear integrated model for r_t.

Evidence of the forecasting abilities of long memory models for r_t is presented in Table 3. Forecast statistics are displayed for the models I through VIII. Other models are also presented as benchmarks with which we can compare the long memory models. The first benchmark model is a straightforward autoregression of order 56. Since forecasting experiments were performed by constructing autoregressive projecting filters with 56 lags for all models, it is important to show that a long autoregression, which might naively be estimated without account being taken of the data's nonstationary character, does not have forecasting capabilities comparable to autoregressive filters of similar size constructed from long memory models. This long autoregression, by necessity of its length and the large sample sizes employed in this study, was estimated by Whittle's (1983) algorithm for the recursive

TABLE 2

Autoregressive Parameter Estimates

ϕ_1

Model	α	Estimate	Nominal SE	Bootstrap Mean	Bootstrap SD	Bootstrap RMS Nominal SE
I.	.67	.15	.02	.10	.10	.05
II.	.69	.19	.03	.14	.09	.06
III.	.71	.19	.05	.14	.09	.05
IV.	.73	.22	.10	.14	.08	.06
V.	.75	.22	.03	.13	.08	.06
VI.		--				
VII.		--				
VIII.		--				

ϕ_3

Model	α	Estimate	Nominal SE	Bootstrap Mean	Bootstrap SD	Bootstrap RMS Nominal SE
I.		--				
II.		--				
III.		--				
IV.		--				
V.		--				
VI.	.64	-.12	.04	-.11	.06	.06
VII.	.65	-.11	.05	-.11	.06	.05
VIII.	.66	-.08	.05	-.07	.06	.06

ϕ_{52}

Model	α	Estimate	Nominal SE	Bootstrap Mean	Bootstrap SD	Bootstrap RMS Nominal SE
I.	.67	.25	.07	.24	.05	.05
II.	.69	.24	.05	.24	.05	.06
III.	.71	.24	.05	.24	.05	.05
IV.	.73	.24	.08	.24	.04	.06
V.	.75	.24	.06	.24	.05	.06
VI.	.64	.25	.05	.25	.04	.06
VII.	.65	.25	.05	.24	.05	.05
VIII.	.66	.23	.05	.23	.05	.05

calculation of autoregression coefficients. The other benchmark models, numbers IX through XI, have the structure

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4)(1 - \phi_{52} B^{52})(1 - B)^d r_t = \varepsilon_t \quad . \quad (15)$$

Models IX, X, and XI have values of d which respectively are estimated to be .34 (α was set at .65 just as in model VII) and are constrained to have the values of .5 and .1. The autoregressive parameters were estimated in the usual way conditional upon d. These final three models are meant to show how a long memory version of a model (IX) compares with versions of the same model displaying two extremes (models X and XI) in nonstationary behavior.

The two panels in Table 3 pertain to the two forecasting horizons of 13 and 26 weeks. For each model the estimated standard error of forecast is presented. By this we mean the standard errors of forecast based upon the model parameters and the standard error of estimate. For example, the sample estimate in column (2) of Table 3 is

$$\tilde{\sigma}_\varepsilon \ \mathrm{SQRT}\left[\sum_{i=0}^{n} \tilde{\psi}_i^2\right] \quad .$$

The standard deviation of the empirical distribution of $\tilde{\varepsilon}$ is $\tilde{\sigma}_\varepsilon$. n is the forecast horizon of either 13 or 26 weeks. Each moving average parameter, $\tilde{\psi}_i$, is constructed from $\tilde{d}$, $\tilde{\phi}_1$, $\tilde{\phi}_2$, $\tilde{\phi}_3$, $\tilde{\phi}_4$ and $\tilde{\phi}_{52}$ as warranted. The standard error of forecast in the bootstrap experiments is

$$\tilde{\sigma}_{\varepsilon^*} \, \mathrm{SQRT}\left[\sum_{i=0}^{n} \tilde{\psi}_i^{*2}\right]$$ where $\tilde{\sigma}_{\varepsilon^*}$ is the standard deviation of the empirical distribution of estimated bootstrap errors, $\tilde{\varepsilon}^*$, and where each $\tilde{\psi}_i^*$ is properly constructed from $\tilde{d}^*$, $\tilde{\phi}_1^*$, $\tilde{\phi}_2^*$, $\tilde{\phi}_3^*$, $\tilde{\phi}_4^*$, and $\tilde{\phi}_{52}^*$. As in the preceding tables, the mean and standard deviations for these empirical bootstrap sampling distributions are presented.

To contrast with the parameter-based standard error of forecast estimates, calculated root mean square forecast errors are also presented in both in-sample and out-of-sample settings. Since all forecast models use 56-order autoregressive projecting filters, in all instances presented in Table 3 there are T-56 = 514-56 = 458 in-sample forecasts and 200 out-of-sample forecasts. The fifth and eighth columns in Table 3, which are simply labeled 'estimate' refer to the sample root mean square forecast error for the estimated model. The bootstrap calculations demonstrate how well the models forecast data which are generated by the estimated model or, in other words, how well an estimated model forecasts its own pseudo-series, r_t^*.

The estimated standard error of forecast and the in-sample root mean square forecast error are both close to their bootstrapped counterparts in most cases. There possibly are departures from this pattern for models I through V at the longer forecast horizons where some root mean square forecast errors are just in the rightmost 5 percent tail of the bootstrap sampling distribution. With respect to out-of-sample performance there is the notable feature that all figures in

TABLE 3

Long Memory Model Forecast Accuracy

Forecast Horizon -- 13 Weeks

	Estimated SE of Forecast			Root Mean Square Forecast Error					
				In-Sample			Out-of-Sample		
		Bootstrap			Bootstrap			Bootstrap	
Model (1)	Estimate (2)	Mean (3)	SD (4)	Estimate (5)	Mean (6)	SD (7)	Estimate (8)	Mean (9)	SD (10)
I.	.60	.65	.10	.67	.65	.07	.31	.65	.10
II.	.57	.62	.09	.68	.62	.06	.31	.63	.09
III.	.57	.62	.09	.68	.62	.07	.31	.62	.09
IV.	.55	.62	.08	.68	.61	.06	.31	.61	.09
V.	.55	.62	.08	.68	.61	.06	.31	.60	.09
VI.	.64	.64	.14	.75	.77	.12	.36	.81	.20
VII.	.61	.62	.14	.74	.70	.10	.34	.66	.19
VIII.	.52	.52	.10	.72	.56	.06	.32	.56	.09
AR(56)				.74	.85	.09	.57	.86	.14
IX.	.61	.59	.08	.74	.70	.10	.34	.72	.16
X.	.45	.44	.03	.71	.62	.06	.36	.62	.09
XI.	.72	.72	.04	.74	.94	.21	.40	1.02	.32

Forecast Horizon -- 26 Weeks

	Estimated SE of Forecast			Root Mean Square Forecast Error					
				In-Sample			Out-of-Sample		
		Bootstrap			Bootstrap			Bootstrap	
Model (1)	Estimate (2)	Mean (3)	SD (4)	Estimate (5)	Mean (6)	SD (7)	Estimate (8)	Mean (9)	SD (10)
I.	.74	.83	.17	.97	.81	.12	.40	.81	.17
II.	.70	.77	.15	.97	.77	.11	.41	.76	.15
III.	.69	.78	.14	.97	.76	.11	.41	.75	.15
IV.	.66	.78	.13	.97	.73	.10	.42	.73	.15
V.	.66	.78	.13	.97	.72	.10	.41	.73	.14
VI.	.81	.82	.23	1.16	1.12	.25	.53	1.20	.40
VII.	.75	.78	.23	1.14	.99	.20	.51	1.05	.32
VIII.	.62	.62	.16	1.06	.71	.11	.44	.72	.16
AR(56)				1.27	1.54	.21	.94	1.60	.25
IX.	.76	.73	.13	1.13	.98	.20	.50	1.03	.31
X.	.54	.54	.04	.95	.70	.09	.48	.70	.13
XI.	.95	.95	.06	1.22	1.56	.46	.62	1.72	.67

column (8) are smaller than figures in column (5). These figures are all too low to have been sampled from the simulated out-of-sample forecast error distributions. The primary reason for this is that there is some marked difference between the variability of the actual data in the 200 post sample observations and that of the other 514 observations of short-term interest rates. We speculate that since the post sample period covers most of the time in which Operation Twist was in force, there was an effort to raise short-term interest rates relative to long-term rates along with some coincidental smoothing of interest rates resulting from policy actions. On the other hand, there are some models in Table 3 (notably models VI and VII) in which the estimated out-of-sample root mean square forecast errors are actually well within their simulated distributions. All in all, it does appear there may be some preference for the models with relatively lower values of $\tilde{d}$ on the basis of in-sample and out-of-sample forecasting accuracy for both actual and simulated data.

The bootstrap methods can also be used to address the question of which models are likely to yield biased forecasts. We found that the long memory models with lower $\tilde{d}$ were less likely to yield biased forecasts. On average, models I through V and model VIII, all of which generally have $\tilde{d}$ and $\tilde{d}^*$ at or below .76, produce forecasts of r_t^* very close to actual r_t^*. This was not as prevalent a feature in models VI, VII, IX, and X. For the less biased forecasting models we would also expect calculated root mean square forecasting errors to be closer to their standard errors of forecast based upon parameter estimates. This is indeed the case as can be confirmed by comparing column (3) with column (6) or column (9). Although models having $\tilde{d}$ and $\tilde{d}^*$ generally greater than .8 appear to have sample and bootstrapped root mean square forecasting errors that are very similar, they are models which really do not forecast very well based upon their own assumptions. Again, comparisons of column (3) with column (6) or (9) demonstrate this.

With respect to the benchmark models, it can immediately be noted that the long autoregression, denoted AR(56) in Table 3, is inferior to the long memory models, particularly in its out-of-sample performance. Not much better is model XI which presumes a true value of d equal to 1. Finally, the model with the least variable accuracy in forecasting we would expect to be the model with the smallest and fixed value of d. In Table 3 this is model X in which d is constrained to be .5. Model X has forecasting characteristics comparable to the other fitted long memory models even though there does not seem to be any other statistical evidence to favor a value for d as low as .5. Nor is there a case to be made for even lower values of d improving forecasting performance. In experiments with conventional stationary ARMA models ($d=0$) forecasting performance was on a par or worse than the performance of the AR(56) model. Other stationary long memory models ($0<d<.5$) displayed poorer performance than their nonstationary brothers.

In the balance of the estimation and forecasting results there are several reasons for preferring long memory nonstationary models as stochastic characterizations of interest rates. We have seen that hypothesis testing based upon bootstrapped confidence intervals around

$\tilde{d}^*$ strongly suggest that $\tilde{d}$ is significantly greater than .5 and probably less than 1. There is some question about how to properly model the high frequency components of the series, however. This is not too surprising in light of the experience of Geweke and Porter-Hudak (1983) in simulation experiments in which the chances of properly recognizing high frequency components of a long memory model were reasonable using only samples of fairly large size. The forecasting performances of models I through VIII compare quite favorably to those of the AR(56) model, other stationary ARMA models and an ARMA model with a unit root in its solution. There is some evidence favoring models I through V on the basis of their relative forecasting performance. In closing this section it will be important to remember that long-term forecasting ability seems to be strongly influenced by the level of $\tilde{d}$ and, judging from the bootstrap standard deviations in Table 3, the level of $\tilde{d}$ also strongly influences the simulated variability of forecasts. Forecast accuracy and variability play crucial roles in the tests conducted in the next section.

IV. Linear Rational Expectations and Variance Bound Tests

This section will look at one notion of variance bounds implicit in the term structure of interest rates, the variance bounds relation between holding period yields and short-term interest rates. The stochastic properties of holding period yields from long-term bonds and short-term interest rates are dramatically different and have been the basis of rejections of the linear rational expectations hypothesis of the term structure by Shiller and Singleton. Wide swings in holding period yields are to be expected on the basis of the capital gains or losses that can be incurred over a very short holding period, but many investigators have considered the spectacular capital gains and losses sometimes observed (especially when interest rates are at low levels) still too great to be in accord with the rational expectations hypothesis. For example, Figure 4 overlays the time series of one-week yields to maturity and one-week holding period yields for Treasury Bills with 12 weeks to maturity. Although the holding period yields seem to deviate from the short-term rates by many hundreds of basis points at certain times, is the deviation too great to agree with the expectations hypothesis?

To answer this we will construct the theoretical variance of holding period yields under the hypothesized expectation mechanism. Recall the infinite moving average representation of the short-term interest rate series,

$$r_t = \sum_{j=0}^{\infty} \psi_j \, \varepsilon_{t-j}.$$

This linear mechanism can be used to characterize a projection of r_t into future periods and using that projection as a model of market expectations we obtain

$$E_{t-m}(r_{t+k}) = \sum_{j=m}^{\infty} \psi_{j+k} \, \varepsilon_{t-j}. \tag{16}$$

E is the expectations operator and equation (16) simply states that the expectation held in period t-m of the interest rate level to prevail in

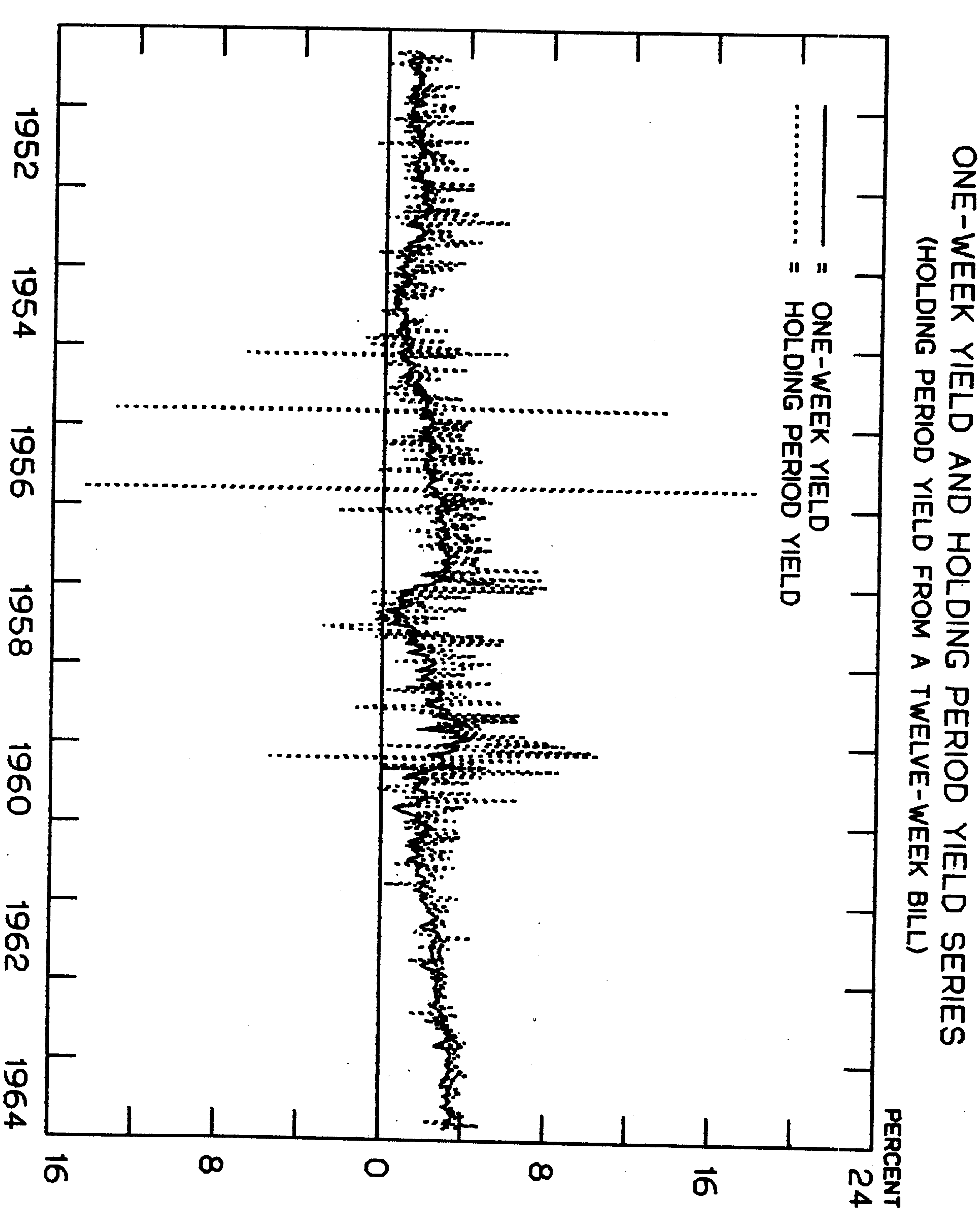
ONE-WEEK YIELD AND HOLDING PERIOD YIELD SERIES
(HOLDING PERIOD YIELD FROM A TWELVE-WEEK BILL)
= ONE-WEEK YIELD
= HOLDING PERIOD YIELD
1952
1954
1956
1958
1960
1962
1964
PERCENT
16
8
0
8
16
24

Figure 4

period t+k is just the appropriate linear combination of expected shocks (ε's) which are zero for periods subsequent to t-m and are known for all periods prior to and up through t-m.

The rational expectations equilibrium between a long-term rate and the short-term rate works through the expected equality of a long-term interest rate with an average of short-term rates expected to prevail over the life of the bond. That is, if $R_t^{(n)}$ is the yield to maturity of a bond with n periods left to redemption,

$$R_t^{(n)} = 1/n \sum_{k=0}^{n-1} E_t(r_{t+k}) = \frac{1}{n} \sum_{k=0}^{n-1} \sum_{j=0}^{\infty} \psi_{j+k} \varepsilon_{t-j}. \quad (17)$$

The one-period yield on a long-term bond can be calculated as the return on the purchase of a bond and its sale in the succeeding period. The yield realized in period t from holding and selling a bond in period t when it has n periods left to maturity is therefore

$$H_t^{(n)} = (n+1)\ R_{t-1}^{(n+1)} - n\ R_t^{(n)}\ . \quad (18)$$

The yield to maturity in period t-1 from a bond with n+1 periods to maturity is

$$R_{t-1}^{(n+1)} = \frac{1}{n+1} \sum_{k=0}^{n} E_{t-1}(r_{t-1+k}) = \frac{1}{n+1} \sum_{k=-1}^{n-1} \sum_{j=1}^{\infty} \psi_{j+k} \varepsilon_{t-j}\ . \quad (19)$$

By combining the right hand sides of equations (17) and (19) we can obtain an expression for a holding period yield in terms of the ψ

parameters of the short-term interest rate process. From equation (18) we have

$$
\begin{aligned}
H_t^{(n)} &= \sum_{k=-1}^{n-1} \sum_{j=1}^{\infty} \psi_{j+k}\, \varepsilon_{t-j} - \sum_{k=0}^{n-1} \sum_{j=0}^{\infty} \psi_{j+k} \varepsilon_{t-j} \\
&= \sum_{j=1}^{\infty} \psi_{j-1} \varepsilon_{t-j} - \sum_{k=0}^{n-1} \psi_k\, \varepsilon_t \qquad (20) \\
&= r_{t-1} + \varepsilon_t \sum_{k=0}^{n-1} \psi_k = r_{t-1} + \delta_n \varepsilon_t \ , \text{ where } \delta_n \text{ is}
\end{aligned}
$$

the sum of the moving average coefficients of the model from index 0 through index n-1. Equation (20) points out that i) an expected holding period yield in the next period is the current short rate and ii) the expected variability of holding period yields relative to short-term interest rates is a function of the bond's term to maturity and the short-term interest rate model's parameters. In other words,

$$
E(H_t^{(n)} - r_{t-1})^2 = \sigma_\varepsilon^2\, \delta_n^2 \ . \qquad (21)
$$

An analytic study of $\delta_n = \sum_{k=0}^{n-1} \psi_k$ as a function of d is potentially interesting, but is the proper object for another paper or for a later version of this one. It is possible to write down succinct expressions for δ_n and δ_n-squared and their derivatives with respect to d. In that way an investigator can search for qualitatively changing behavior in δ_n^2, say, as d approaches the stationarity boundary of .5. Here it will suffice to show that δ_n is, in general, a complex combination of d and other parameters of a long memory model. Taking equation (13) as an example, we can write down the first four ψ coefficients as

$$\psi_0 = 1,$$

$$\psi_1 = \frac{\Gamma(d+1)}{\Gamma(d)\Gamma(1)} - \phi_1 = d - \phi_1,$$

$$\psi_2 = \frac{\Gamma(d+2)}{\Gamma(d)\Gamma(2)} - \phi_1(d - \phi_1)$$

$$= \quad d^2 + d + \phi_1 d - \phi_1^2 \text{ , and}$$

$$\psi_3 = \frac{\Gamma(d+3)}{\Gamma(d)\Gamma(3)} + [d^2 + d + \phi_1 d - \phi_1^2]\ \phi_1$$

$$= \frac{(d+2)(d+1)d}{2} + [d^2 + d + \phi_1 d - \phi_1^2]\ \phi_1 \ .$$

If equation (21) is to be used as the basis of a test of the rational expectations hypothesis when interest rates follow a long memory process, sampling distributions for δ_n will have to be constructed and we will have to be concerned with the joint distributional properties of d and the ϕ coeffcients. Moreover, one has to be able to evaluate the moments of δ_n's sampling distribution. This distribution will, in general be a function of the factorial and product moments of the jointly distributed d and ϕ parameters. Even if these parameters were joint normally distributed and model estimation was based upon that fact, the evaluation of sampling distributions for δ_n is difficult. As useful as it was to use the Monte Carlo methods to simulate distributions in the previous section, it is essential in this section to use the same methods to simulate the distributions of sample δ_n's.

Using the models of interest rates presented in the previous section, the $\tilde{\psi}$ coefficients in the infinite moving average representation can be calculated to give estimated $\sigma_\varepsilon^2\delta_n^2$ for any value of n. From the Treasury Bill yield curve data an unbiased estimate of the variance of $H_t^{(n)} - r_{t-1}$ is obtained by taking the noncentral second moments of $H_t^{(n)} - r_{t-1}$ for any particular n. In Table 4 sample and bootstrap estimates of $E(H_t^{(n)} - r_{t-1})^2$ are presented. The in-sample estimates refer to estimates using the 514 observations which figured in model estimation. The so-called full-sample examples use the full complement of 713 available observations of r_{t-1} and $H_t^{(n)}$. For each model I through VIII, X, and XI a panel of estimated and bootstrap simulated $\sigma_\varepsilon^2\delta_n^2$ (the latter being based upon estimated $\tilde{\psi}^*$) are presented.

The results are easily enough stated. For all models except X the hypothesis that holding period yields are not too variable to agree with the rational expectations theory and a long memory characterization of interest rates is not rejected. Only in the case in which d is fixed and known to be .5 is a long memory model (X) close to describing a stationary interest rates series. In that case the mean and standard deviation of bootstrapped $\sigma_\varepsilon^2\delta_n^2$ are small enough to reject equation (21) as a tenable hypothesis. If we are inclined to assume that interest rates are mean stationary stochastic variables and especially if we assume that d equals zero rather than some other value, it is rather easy to bias tests toward the rejection of the rational expectations

TABLE 4

Excess Variability Test Statistics

	$E(H_t^{(n)} - r_{t-1})^2$						Model I $\sigma_\varepsilon^2\,\delta_n^2$		
	In-Sample			Full-Sample					
		Bootstrap			Bootstrap			Bootstrap	
n	Estimate	Mean	SD	Estimate	Mean	SD	Estimate	Mean	SD
1	.072	.072	.006	.060	.060	.005	.063	.063	.005
2	.221	.222	.024	.183	.184	.018	.162	.184	.045
3	.438	.440	.059	.356	.357	.044	.352	.394	.094
4	.721	.725	.105	.579	.579	.075	.604	.673	.161
5	1.09	1.09	.159	.861	.863	.116	.904	1.01	.251
6	1.52	1.52	.234	1.19	1.20	.167	1.24	1.40	.365
7	2.00	1.99	.346	1.55	1.55	.238	1.62	1.85	.505
8	2.55	2.56	.481	1.95	1.95	.333	2.03	2.34	.670
9	3.11	3.08	.608	2.36	2.37	.462	2.48	2.88	.861
10	3.72	3.65	.756	2.79	2.81	.575	2.95	3.46	1.08
11	4.53	4.59	.937	3.38	3.36	.677	3.45	4.08	1.31

	Model II $\sigma_\varepsilon^2\,\delta_n^2$		
		Bootstrap	
n	Estimate	Mean	SD
1	.063	.062	.005
2	.149	.169	.04
3	.323	.363	.08
4	.557	.621	.14
5	.835	.933	.22
6	1.15	1.29	.32
7	1.49	1.69	.44
8	1.87	2.13	.58
9	2.26	2.61	.74
10	2.69	3.13	.92
11	3.13	3.68	1.12

	Model III $\sigma_\varepsilon^2\,\delta_n^2$				Model IV $\sigma_\varepsilon^2\,\delta_n^2$				Model V $\sigma_\varepsilon^2\,\delta_n^2$		
		Bootstrap				Bootstrap				Bootstrap	
n	Estimate	Mean	SD	n	Estimate	Mean	SD	n	Estimate	Mean	SD
1	.063	.063	.005	1	.063	.063	.005	1	.063	.062	.005
2	.147	.170	.04	2	.137	.169	.03	2	.137	.170	.03
3	.320	.366	.08	3	.296	.365	.07	3	.297	.368	.07
4	.552	.629	.14	4	.513	.627	.13	4	.515	.630	.12
5	.826	.944	.22	5	.771	.942	.20	5	.773	.946	.19
6	1.14	1.31	.31	6	1.06	1.30	.29	6	1.06	1.31	.28
7	1.48	1.71	.43	7	1.38	1.71	.39	7	1.38	1.71	.38
8	1.85	2.16	.56	8	1.72	2.15	.52	8	1.72	2.16	.50
9	2.24	2.65	.71	9	2.08	2.45	.66	9	2.09	2.64	.64
10	2.66	3.17	.89	10	2.46	3.15	.82	10	2.47	3.16	.80
11	3.10	3.73	1.08	11	2.86	3.71	1.00	11	2.87	3.72	.97

TABLE 4 (cont'd.)

Model VI

	$\sigma^2_\varepsilon \; \delta^2_n$		
		Bootstrap	
n	Estimate	Mean	SD
1	.063	.064	.005
2	.219	.219	.03
3	.450	.450	.09
4	.800	.803	.22
5	1.11	1.13	.34
6	1.47	1.50	.50
7	1.86	1.90	.69
8	2.31	2.38	.92
9	2.79	2.89	1.18
10	3.32	3.46	1.49
11	3.88	4.07	1.84

Model VII

	$\sigma^2_\varepsilon \; \delta^2_n$		
		Bootstrap	
n	Estimate	Mean	SD
1	.063	.063	.005
2	.213	.213	.03
3	.427	.432	.09
4	.747	.762	.21
5	1.03	1.06	.33
6	1.35	1.41	.49
7	1.70	1.78	.67
8	2.09	2.22	.90
9	2.52	2.70	1.16
10	2.98	3.22	1.46
11	3.47	3.77	1.80

Model VIII

	$\sigma^2_\varepsilon \; \delta^2_n$		
		Bootstrap	
n	Estimate	Mean	SD
1	.065	.065	.005
2	.198	.195	.03
3	.374	.369	.08
4	.616	.606	.17
5	.835	.825	.25
6	1.08	1.07	.36
7	1.33	1.33	.48
8	1.62	1.62	.63
9	1.91	1.93	.80
10	2.23	2.26	.99
11	2.56	2.61	1.20

Model X

	$\sigma^2_\varepsilon \; \delta^2_n$		
		Bootstrap	
n	Estimate	Mean	SD
1	.064	.063	.005
2	.072	.072	.009
3	.128	.127	.02
4	.232	.230	.03
5	.340	.336	.03
6	.485	.479	.05
7	.666	.655	.06
8	.880	.864	.08
9	1.12	1.09	.10
10	1.37	1.34	.13
11	1.64	1.61	.17

Model XI

	$\sigma^2_\varepsilon \; \delta^2_n$		
		Bootstrap	
n	Estimate	Mean	SD
1	.063	.062	.005
2	.073	.270	.02
3	.611	.603	.05
4	1.04	1.02	.09
5	1.43	1.41	.13
6	1.80	1.77	.18
7	2.24	2.21	.24
8	2.79	2.75	.31
9	3.42	3.37	.39
10	4.11	4.06	.48
11	4.85	4.78	.58

hypothesis. For all other long memory models, however, sample estimates of $E(H_t^{(n)} - r_{t-1})^2$ are well within the bootstrapped sampling distributions of $\sigma_\varepsilon^2\delta_n^2$ and sample estimates of $\sigma_\varepsilon^2\delta_n^2$ are well within the bootstrapped sampling distributions of $E(H_t^{(n)} - r_{t-1})^2$.

V. Concluding Remarks

The results in Table 4 suggest that the rational expectations hypothesis is consistent with our observations of how a set of high quality term structure data behave. At least we do not find evidence such as that reported by Singleton (1980) that holding-period yield variances exceed their implied bounds by a factor of 10^1. Confidence intervals around sample $E(H_t^{(n)} - r_{t-1})^2$ and estimated $\sigma_\varepsilon^2\delta_n^2$ suggest that both are quite close enough to affirm the rational expectations hypothsis. There is much improvement to be made in the theory of estimation and forecasting with long memory models and there is some question (Schenker 1985) as to the reliability of bootstrapped confidence intervals for complexly constructed test statistics such as those presented in Table 4. We have tried to overcome some of the qualms about the bootstrap procedures by using as large sized samples of high quality data as possible. There are outstanding theoretical questions as to whether bootstrap techniques are useful for bias correction in long memory models and what effect that would have in aiding the long memory model selection process. For the data studied here and for the period of which it is representative, however, it is fairly clear that the mean-variance stationarity assumptions found in the prevalent excess volatility literature are unwarranted. How likely is it that the same assumptions are valid for asset returns in the 1970's and 1980's?

APPENDIX

Data used in this study were collected and described by Roll (1971, Chapter 5). This high-quality data set is a collection of bid-ask quotes carefully culled from broker quote sheets and corrected for calender effects to the greatest extent possible so that quotes are for Treasury Bills in the brokers' discount market for integer numbers of weeks to maturity. To construct the longest time series of yield curves possible only quotes for bills out to 13 weeks to maturity were used. On some dates quotes for bills with an integer number of weeks to maturity were not available. For this reason bid-ask quote averages were smoothed with a constrained cubic B-spline function, a method for generating an approximation to the term structure described by Shea (1984). This approximation to the discount curve is constrained so that it passes through the origin; that is, it is constrained so that the predicted price of a zero-maturity Treasury Bill is its face value. In this way it was possible to automatically generate 796 consecutive weekly discount curves from October 1949 through the last week of December 1964. The smoothing method is not without its faults. For one, it is particularly dangerous to use estimates of the term structure implied by the slope of the approximation at the end of the approximation interval, in this case, 13 weeks to maturity. There is little that can be done to correct for the poor yield curve slope approximations that come from these methods except by undertaking *ad hoc* constraint experimentation to manipulate the slopes until they seem reasonable. What has been done here is to use the term structure estimates out to only 12 weeks to maturity. This is done with the realization that term structure approximations at all maturities are possibly affected by the faulty local approximations at the end of the yield curve. Researchers who use such approximations should be aware of these problems.

The approximations, not surprisingly, closely replicate the data which appear in Roll's Figure 5-1 and the actual average yields by maturity contained in his Table 5-3. The convenience of the approximations is that only the few estimated coefficients of the approximating function need be saved and manipulated to quickly generate any transformation of the interest rate term structure that is desired. The data used in this study are actually from the final 714 weeks of the sample's post Treasury Accord period.

FOOTNOTES

* International Finance Division, Board of Governors of the Federal Reserve System and The Pennsylvania State University. This paper represents the views of the author and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or other members of its staff. This paper builds upon material in the author's dissertation (Shea 1982) and thereby benefitted from comments from Charles R. Nelson. A number of my colleagues at the Board have patiently discussed with me material covered in this paper. Under no circumstances would any blame attach to them for my own errors. These helpful discussants were Ed Green, Richard Freeman, and Jeffrey Marquardt. Conversations with Susan Porter-Hudak and John F. Geweke were particularly helpful in solving or in suggesting solutions to some of the technical problems encountered. David Laughton provided valuable programming assistance at several moments of crisis. Margaret Gray and Kathy Krasney ably typed the manuscript.

1 Tests performed by Singleton (1980) were repeated using the data from the Appendix. The results were much the same as those reported by Singleton, so the objection that the data are especially favorable towards the rational expectations hypothesis cannot be raised on the point that the hypothesis might be rejected using his test procedures.

BIBLIOGRAPHY

Box, G.E.P. and Jenkins, G.M. Time Series Analysis: Forecasting and Control. Revised Edition. San Francisco: Holden-Day, 1976.

Brick, R. and Thompson, H.E. "Time Series Analysis of Interest Rates: Some Additional Evidence," Journal of Finance 33 (March 1978): 93-103.

Cargill, T.F. "The Term Structure of Interest Rates: A Test of the Expectations Hypothesis." Journal of Finance 30 (June 1975): 761-771.

Culbertson, J.M. "The Term Structure of Interest Rates." Quarterly Journal of Economics 71 (November 1957): 485-517.

Flavin, M.A. "Excess Volatility in the Financial Markets: A Reassessment of the Empirical Evidence." Journal of Political Economy 91 (December 1983): 929-956.

Freedman, D.A. and Peters, S.C. "Bootstrapping a Regression Equation: Some Empirical Results." Journal of the American Statistical Association 79 (march 1984): 97-106.

Geweke, J. and Porter-Hudak, S. "The Estimation and Application of Long Memory Time Series Models." Journal of Time Series Analysis 4 (1983): 221-238.

Granger, C.W.J. "Some Properties of Times Series Data and their Use in Econometric Model Specification." Journal of Econometrics 16 (May 1981): 121-130.

Granger, C.W.J. "The Typical Spectral Shape of an Economic Variable." Econometrica 34 (January 1966): 150-161.

Granger, C.W.J. and Joyeux, R. "An Introduction to Long-Memory Time Series Models and Fractional Differencing." Journal of Time Series Analysis 1 (1980): 15-39.

Granger, C.W.J. and Rees, H.J.B. "Spectral Analysis of the Term Structure of Interest Rates," Review of Economic Studies 35 (January 1968): 67-76.

Leroy, S.F. and Porter, R.D. "The Present Value Relation: Tests Based Upon Implied Variance Bounds," Econometrica 49 (May 1981): 555-74.

Mandelbrot, B.B. "Statistical Methodology for Nonperiodic Cycles: From the Covariance to R/S Analysis." Annals of Economic and Social Measurement 1 (July 1972): 259-290.

Marsh, T. and Merton, R. "Dividend Variability and Variance Bounds Tests for the Rationality of Stock Market Prices." Discussion paper presented at the National Bureau of Economic Research Program on Economic Fuctuations Research Meeting, October 1984.

Michener, R.W. "Variance Bounds in a Simple Model of Asset Pricing." Journal of Political Economy 90 (February 1982): 166-75.

Nelson, C.R. and Plosser, C.I. "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications." Journal of Monetary Economics 10 (September 1982): 139-162.

Peters, S.C. and Freedman, D.A. "Using the Bootstrap to Evaluate Forecasting Equations." Technical Report No. 20, Department of Statistics, University of California, Berkely, October 1984.

Porter-Hudak, S. "Long-Term Memory Modelling-A Simplified Spectral Approach." Unpublished Ph.D. dissertation, University of Wisconsin, 1982.

Roll, R. The Behavior of Interest Rates, New York: Basic Books, 1971.

Schenker, N. "Qualms about Bootstrap Confidence Intervals." Journal of the American Statistical Association 80 (June 1985): 360-361.

Shea, G.S. "Pitfalls in Smoothing Interest Rate Term Structure Data: Equilibrium Models and Spline Approximations." Journal of Financial and Quantitative Analysis 19 (September 1984): 253-269.

Shea, G.S. "The Japanese Term Structure of Interest Rates," Unpublished Ph.D. Dissertation, University of Washington, 1982.

Shiller, R.J. "The Volatility of Long-Term Interest Rates and Expectations Models of the Term Structure," Journal of Political Economy 87 (December 1979): 1190-1219.

Shiller, R.J. and Grossman, S.J. "The Determinants of the Variability of Stock Market Prices." American Economic Review - AEA Papers and Proceedings 71 (May 1981): 222-227.

Singleton, K.J. "Expectations Models of the Term Structure and Implied Variance Bounds," Journal of Political Economy 88 (December 1980): 1159-76.

Whittle, P. Prediction and Regulation by Linear Least-Square Methods. Second Edition Revised. Minneapolis: University of Minnesota Press, 1983.