COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 743

DO WE REJECT TOO OFTEN?

SMALL SAMPLE BIAS IN TESTS OF RATIONAL EXPECTATIONS MODELS

by

N. Gregory Mankiw

and

Matthew D. Shapiro

April 1985

Do We Reject Too Often?

Small Sample Bias in Tests of Rational Expectations Models

N. Gregory Mankiw*

Massachusetts Institute of Technology

Matthew D. Shapiro*

Yale University

March 1985

Abstract

We examine the small sample properties of tests of rational expectations models. We show using Monte Carlo experiments that these tests can be extremely biased toward rejection for sample sizes typical in applied research. These biases are important when the time series examined are highly autoregressive. We also show that these tests are even more biased with detrended data. We present correct small sample critical values for our canonical problem.

Keywords: Macroeconomics; Rational Expectations; Small Sample Bias; Monte Carlo.

*We are grateful to Donald Andrews, James Heckman, Adrian Pagan, Peter Phillips, and James Poterba for helpful discussions.

address: Matthew D. Shapiro
Cowles Foundation for Research in Economics
2125 Yale Station
New Haven, CT 06520

1. The Issue

Economic models with rational expectations often imply that the expectation of some variable, Y_t, conditional on information available at time t-1 is a constant. That is,

$$(1) \qquad E_{t-1} Y_t = \Phi_0.$$

We can equivalently write

$$(1') \qquad Y_t = \Phi_0 + \nu_t \qquad \text{where } E_{t-1} \nu_t = 0.$$

Such models play a central role in much recent empirical macroeconomics. For example, the permanent income hypothesis has this implication, where Y_t is the change in consumption (Hall [1978], Flavin [1981]). The expectations theory of the term structure can be represented as in equation (1), where Y_t is the difference between the holding return on a long-term bond and the one-period bill rate (Shiller [1979], Jones and Roley [1983]). The hypothesis that the real interest rate is constant takes this form, where Y_t is the *ex post* real interest rate (Mishkin [1981]). These three examples only begin to catalog the models that imply such an orthogonality condition.

The standard test of the model (1) is to regress the realization of the variable Y_t on lagged information. (See Abel and Mishkin [1983].) That is, we might estimate

$$(2) \qquad Y_t = \Phi_0 + \Phi_1 X_{t-1} + \nu_t$$

using ordinary least squares. According to equation (1), the coefficient Φ_1 equals zero. We evaluate the model by statistically testing the null hypothesis H_0: $\Phi_1 = 0$.

Suppose the variable X_t follows a first-order autoregressive process:

$$(3) \qquad X_t = \theta X_{t-1} + \varepsilon_t.$$

Under the null hypothesis, X_{t-1} and ν_t are uncorrelated. The model, however, does not preclude a contemporaneous correlation between ε_t and ν_t. Indeed, in many cases, the underlying theory implies such a contemporaneous correlation. (For example, if Y_t is the change in consumption and X_t is income, then Hall's version of the permanent income hypothesis implies that Y_t is perfectly correlated with the innovation in X_t.) Suppose

$$(4) \qquad \text{corr}(\varepsilon_{t+j},\nu_t) = \begin{cases} \rho & \text{if } j = 0 \\ 0 & \text{otherwise} \end{cases}$$

In this case, the right-hand side variable in the regression (2), while contemporaneously uncorrelated with the residual, is not uncorrelated with it at all leads and lags. In particular, X_{t-1} is correlated with $\nu_{t-1}, \nu_{t-2},\ldots.$ Therefore, the Gauss-Markov Theorem does not apply. The justification of ordinary least squares estimation of (2) and the subsequent hypothesis testing relies on asymptotic distribution theory.[1]

In this paper we examine the conditions under which the asymptotic theory leads to incorrect inference for samples of typical size. In Mankiw

---

[1]Our problem differs from the well-known spurious regression problem (Granger and Newbold [1974]; Nakamura, Nakamura, and Orcutt [1976]). In the spurious regression problem, the econometrican is using OLS to regress one serially correlated series on another independent serially correlated

and Shapiro [1984], we examine this issue in the context of a specific application: tests of the permanent income hypothesis. Small sample bias in related contexts is also discussed in Hoffman and Schmidt [1981], Huizinga and Mishkin [1984], Flavin [1984], and Hoffman, Low and Schlagenhauf [1984].

We study here with Monte Carlo experiments the canonical problem described above. We show that when both the contemporaneous correlation, ρ, and the autoregressive parameter, Θ, are close to unity, the asymptotic test of the null hypothesis that $\Phi_1 = 0$ is leads to a rejection too often.[2] For a sample size of 100, a test with a nominal size of five percent actually has a size of twenty-eight percent. In other words, the significance level of a rejection is overstated by a factor of five. Thus, if practitioners rely on the asymptotic distributions of the test statistics, they will reject true models much too frequently.[3]

Our results provide guidance to those testing orthogonality conditions such as equation (1). In particular, unless the serial correlation of the forecasting variable, X_t, is small ($\Theta < 0.9$), one should be wary of the asymptotic distribution. That is, if the forecasting variable is highly

---

series. Because that model is misspecified (a lagged dependent variable is incorrectly omitted), the problem of spurious regression remains asymptotically. In constrast, we consider incorrect inferences about a correctly specified model caused by unwarranted reliance on distributions that are correct only asympototically.

[2]These results are related to those of Dickey and Fuller [1979,1981] and Evans and Savin [1981,1984], who show that standard critical values are inadequate in the presence of unit roots. All our examples, however, assume stationarity. Hence, the sole issue is small sample bias.

[3]Another question regarding such orthogonality tests is whether they are powerful against plausible alternative hypotheses. For discussions of this question, see Shiller and Perron [1984] and Summers [1982]. In this paper we consider only the properties of the test under the null hypothesis.

autocorrelated, the rejection of the null hypothesis may require a stricter critical value than implied by the asymptotic distribution. We provide the correct critical values for the canonical problem.

We also examine the use of detrended data for testing orthogonality conditions. We show that finite sample bias is even greater for detrended data. A nominal five percent test can have an actual size of over fifty percent. In this case, a practicioner relying on the asympototic distribution will usually reject a true null hypothesis. We also provide correct critical values for detrended data.

Section II describes the Monte Carlo experiment, while Section III presents the results for stationary (non-detrended) data. Section IV consider the problem when the data are detrended. Section V offers some concluding observations.

2. Monte Carlo Experiment

The critical parameters for each Monte Carlo experiment are ρ and Θ. Given these parameter values, a series of N innovations, ν_t and ε_t, are generated from a bivariate normal with variances equal to one and covariance equal to ρ. The variable Y_t is set equal to ν_t. The variable X_t is generated from equation (3) using the innovations ε_t. The initial value X_0 is chosen randomly from the stationary distribution for X, which is the univariate normal with mean equal to zero and variance equal to $1/(1 - \Theta^2)$.

Once the data are generated, we estimate equation (2) and record the value of the t-statistic. We replicate this procedure 1000 times, allowing us to estimate the distribution of the statistic. We present below the fraction of the time a practitioner using the asympototic distribution will reject the true null hypothesis based on the conventional five percent critical value. We also present the true critical value required for a five percent test. That is, if τ is the t-statistic, then the five percent critical value is the number Ψ such that $\text{Prob}(|\tau| > \Psi) = .05$ under the null hypothesis. Note that this is not equivalent to a 2.5 percent one-tailed critical value, since the distribution of τ is not symmetric in small samples. Our procedure, however, leads to inferences identical to those had we examined the $\chi^2(1)$ statistic (that is, τ^2).

The results of these experiments are more general than they might at first appear. First, changing the variance of any of the variables does not alter the value of the test statistic. For example, if the standard deviation of ε_t is doubled, then all values of X_t are doubled. While this would alter the estimate of Φ_1 in equation (2), it would not change the t-statistic. Second, allowing the constant Φ_0 to be non-zero or including a constant in equation (3) does not change the value of the t-statistic. Either constant would increase all the values of X or Y by a constant. Since our regression includes a constant term, there would be no effect on the t-test for the slope coefficient. Thus, each experiment is fully defined by ρ, Θ, and the sample size N.

3. Results

Tables 1, 2, and 3 present the results for sample sizes of N = 50, N = 100, and N = 200, which are in the range of sample sizes typically found in applied macroeconomic research. The top number in each cell is the true size of test based on a critical value of 2.0, that is, a test of size five percent relying on the asympototic distribution. For values of ρ and Θ close to unity, the actual size is far greater than five percent. A practitioner using the asympototic distribution will reject a true null hypothesis more than five percent of the time. Any given rejection is far less significant than one would be led to believe from asympototic distribution theory.

The tables also present the correct critical value for a five percent test. If ρ and Θ are close to unity, the critical value of the t-statistic is closer to 3.0 than to the usual 2.0. A valid rejection at a five percent significance level requires a much more conservative critical value.

Figure 1 presents the distribution of the test statistic for $\Theta = 0.99$, $\rho = 0.9$, $N = 100$. Figure 1 also presents the asymptotic distribution, which is of course standard normal. We find that the true distribution of the statistic is very different from the asymptotic distribution. In particular, the true distribution does not have zero mean. This figure vividly illustrates the inaccuracy of the asymptotic distribution as an approximation to the actual distribution of the test statistic.

4. The Use of Detrended Data

Often in applied work the time series of interest are not stationary. It is standard to model time series such as real GNP or industrial production as stationary around a deterministic trend. That is, one might postulate that

(5) $$Z_t = \alpha + \beta \text{ Time} + X_t$$

where Z_t is the observed economic series and X_t is some stationary stochastic process. While Z_t is directly observable, X_t is not, since the parameters α and β must be estimated.

Nelson and Plosser [1982] and Nelson and Kang [1981,1983] demonstrate that the assumption that Z_t is stationary around a deterministic trend is itself not innocuous. Severe biases can result if Z_t is in fact non-stationary. We assume here, however, that equation (5) is the correct model. Our goal is to concentrate on the problem of small sample bias.

A standard procedure is to "detrend" Z_t by taking the residuals from an OLS regression of Z_t on a time trend and to use the detrended series for empirical testing. That is, we use the detrended series to test the null hypothesis that $\Phi_1 = 0$ in equation (2). By the Frisch-Waugh [1933] theorem, this procedure is numerically equivalent to regressing Y_t on both Z_t and a time trend. In other words, since the time trend is orthogonal to the detrended series by construction, estimating the time trend together with the parameter Φ_1 is equivalent to first estimating the time trend and then estimating Φ_1 using the detrended series.

In this section we show that the bias discussed in the previous section is particularly pronounced with detrended data. This increase in bias arises because the parameters α and β must be estimated: if they were known, X_t could be observed and the problem would be identical to the one above. In most applied work, however, the trend must be estimated, increasing the severity of the small sample bias.

To study the bias when using detrended data, we perform the same Monte Carlo experiment as above except that X_t is now first detrended. It might appear that we are assuming that $\alpha = \beta = 0$. It is straightforward to show, however, that the detrended series of Z_t is independent of α and β. Hence, we can set $\alpha = \beta = 0$ without loss of generality.

Tables 4, 5, and 6 present the results of the Monte Carlo experiment for sample sizes of 50, 100, and 200. Again, the top number in each cell is the percent rejections based on a "five percent" test. The bias is even larger than before. If Θ and ρ are close to unity, the actual significance level is roughly fifty percent. Thus, using the asymptotic distribution, a practitioner will reject the true model as frequently as not.

Again, the bottom number of each cell is the correct critical value for a five percent two-tailed test. We find that much larger critical values are required with detrended data. In particular, for Θ and ρ close to unity, the required critical value is 3.5 rather than the asymptotic 2.0.

5. Conclusions

Standard tests of orthogonality can reject too often when applied to strongly autoregressive series. This bias is particularly severe with detrended data. The Monte Carlo results presented in this paper should be useful for practitioners testing rational expectations models. The critical values presented can help ensure that rejections of models are not attributable to unwarrented reliance on asymptotic distribution theory.

The problems we discuss are probably prevalent in research using standard macroeconomic time series. Nelson and Plosser [1982] show that for many time series, one cannot reject the existence of a unit root. Hence, even if one maintains the assumption that the series are stationary around a trend, one must allow the possibility that the series are strongly autoregressive. It is not sufficient to rely upon estimated autoregressive parameters, since these are biased toward zero (Hurwicz [1950], Sawa [1978]). It therefore seems prudent for a researcher to consider the worst case parameter values before rejecting a null hypothesis with confidence.

Our findings suggest that other tests of orthogonality may be severely biased as well. Tests of Granger [1969] causality, pioneered by Sims [1972] and others, are essentially orthogonality tests. Unlike the class of problems considered here, Granger causality tests allow lags of the left-hand side variable to enter on the right-hand side of the regression.[4] Nonetheless, we believe that these tests are also biased in samples of typical size, a conjecture we have confirmed by limited Monte Carlo experimentation. Our results suggest that if the series are highly autocorrelated and the innovations are contemporaneously correlated, as in many applications, reliance on the asymptotic distribution yields too many rejections of Granger non-causality.

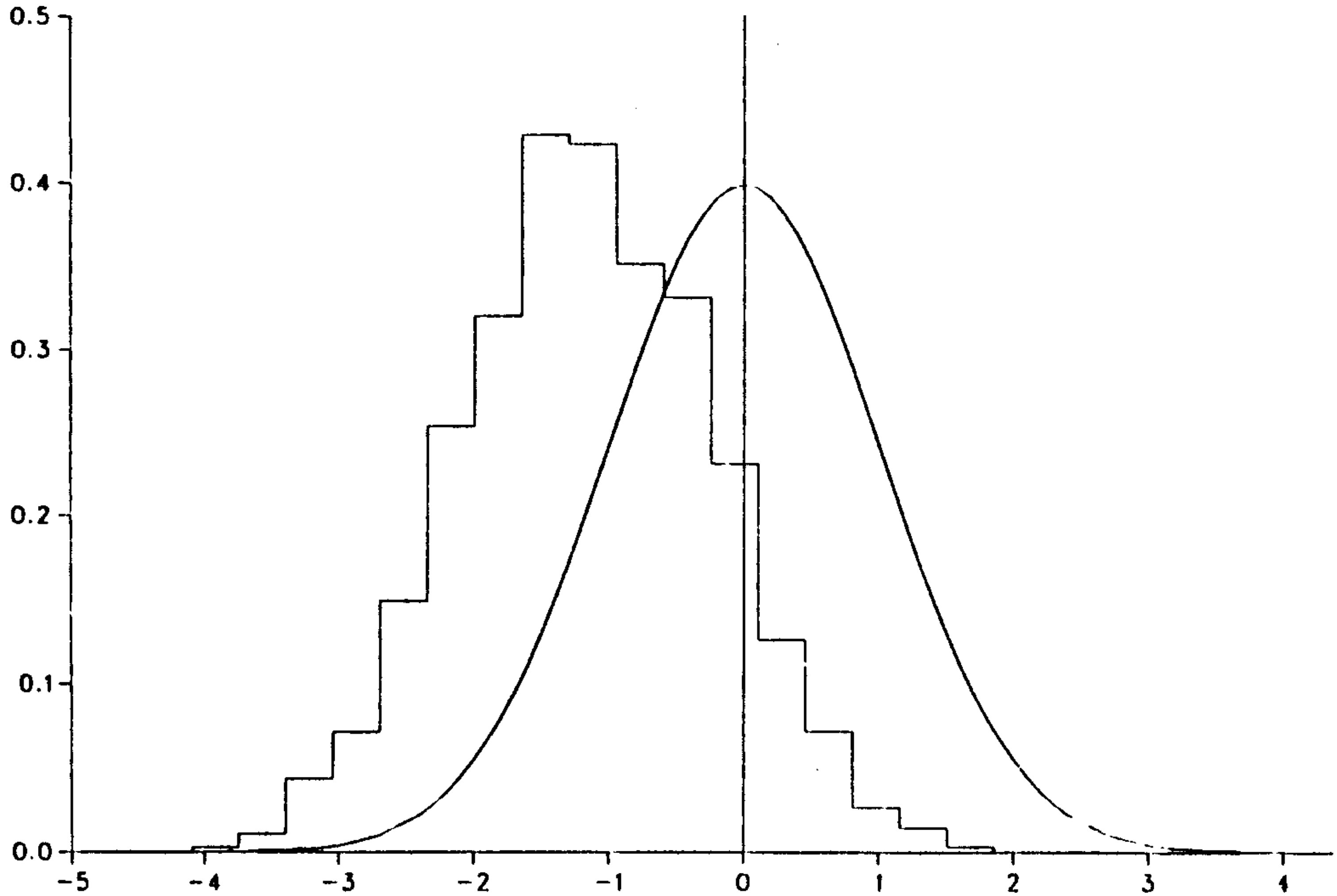
Our results are also relevant to tests of certain structural models. Hansen and Singleton [1982] propose a general strategy for estimating and

---

[4]Geweke [1981] and Nelson and Schwert [1982] consider the finite sample properties--primarily the power--of tests of Granger causality.

testing rational expectations models that is based on orthogonality conditions such as our equation (1). Their test of over-identifying restrictions is essentially the orthogonality test in our equation (2). Our results indicate that these tests, which rely on asymptotic theory, may also be biased toward rejection in samples of typical size.

Figure 1



The distribution of the t-statistic for $\Theta = 0.99$, $\rho = 0.9$, and $N = 100$, compared to the asymptotic distribution (the standard normal).

Table 1: Results of Monte Carlo Experiment

Non-detrended Stationary Data

N = 50

Top Number = Percent rejections using the nominal five percent critical value of 2.0.

Bottom Number = Correct critical value for a five percent two-tailed test.

θ = \ $\lvert\rho\rvert$ =	1.0	0.9	0.8	0.7	0.5	0.0
0.999	30 3.0	24 2.8	20 2.7	16 2.5	11 2.3	7 2.1
0.99	26 2.9	20 2.6	15 2.5	13 2.4	10 2.3	7 2.1
0.98	22 2.8	17 2.6	15 2.4	11 2.3	8 2.3	7 2.1
0.95	17 2.6	12 2.4	10 2.3	8 2.2	7 2.1	6 2.1
0.9	12 2.4	9 2.2	8 2.2	6 2.1	6 2.1	6 2.1
0.0	5 2.0	6 2.0	6 2.0	6 2.0	5 2.0	5 2.0

Table 2: Results of Monte Carlo Experiment

Non-detrended Stationary Data

$N = 100$

Top Number = Percent rejections using the nominal five percent critical value of 2.0.

Bottom Number = Correct critical value for a five percent two-tailed test.

θ = \ $\lvert\rho\rvert$ =	1.0	0.9	0.8	0.7	0.5	0.0
0.999	28 2.8	24 2.8	20 2.6	16 2.5	11 2.4	6 2.1
0.99	22 2.7	20 2.6	16 2.5	13 2.4	8 2.2	5 2.0
0.98	19 2.6	16 2.5	13 2.4	10 2.3	7 2.1	6 2.0
0.95	12 2.4	11 2.4	9 2.2	8 2.1	6 2.0	5 2.0
0.9	10 2.3	9 2.2	8 2.1	6 2.1	6 2.0	5 1.9
0.0	5 2.0	6 2.0	5 2.0	5 1.9	5 1.9	5 1.9

Table 3: Results of Monte Carlo Experiment

Non-detrended Stationary Data

N = 200

Top Number = Percent rejections using the nominal five percent critical value of 2.0.

Bottom Number = Correct critical value for a five percent two-tailed test.

θ = / $\|\rho\|$ =	1.0	0.9	0.8	0.7	0.5	0.0
0.999	29 2.9	23 2.8	20 2.7	16 2.6	10 2.4	5 2.0
0.99	18 2.6	15 2.5	13 2.4	11 2.4	8 2.3	4 1.9
0.98	13 2.5	10 2.3	9 2.3	9 2.2	7 2.1	5 1.9
0.95	9 2.2	7 2.2	7 2.1	6 2.0	6 2.0	5 2.0
0.9	7 2.1	6 2.0	6 2.0	5 2.0	6 2.0	6 2.1
0.0	5 2.0	4 1.9	4 1.9	5 2.0	5 2.0	5 2.0

Table 4: Results of Monte Carlo Experiment

Detrended Data

N = 50

Top Number = Percent rejections using the nominal five percent critical value of 2.0.

Bottom Number = Correct critical value for a five percent two-tailed test.

θ = \ $\|\rho\|$ =	1.0	0.9	0.8	0.7	0.5	0.0
0.999	60 3.5	45 3.2	36 3.0	28 3.0	16 2.7	6 2.0
0.99	54 3.4	40 3.2	33 3.0	27 2.9	15 2.6	6 2.0
0.98	50 3.3	37 3.1	30 3.0	24 2.8	14 2.6	5 2.0
0.95	38 3.1	30 3.0	25 2.8	19 2.7	12 2.6	6 2.0
0.9	28 2.9	22 2.8	19 2.7	14 2.6	10 2.3	6 2.0
0.0	6 2.1	7 2.1	7 2.1	6 2.0	5 2.0	6 2.0

Table 5: Results of Monte Carlo Experiment

Detrended Data

N = 100

Top Number = Percent rejections using the nominal five percent critical value of 2.0.

Bottom Number = Correct critical value for a five percent two-tailed test.

θ = \ $\|\rho\|$ =	1.0	0.9	0.8	0.7	0.5	0.0
0.999	58 3.5	48 3.3	39 3.2	30 2.9	16 2.6	6 2.0
0.99	49 3.4	39 3.2	32 3.0	25 2.8	15 2.5	5 2.0
0.98	41 3.3	34 3.1	27 2.9	21 2.7	13 2.4	5 2.0
0.95	26 2.9	21 2.8	17 2.6	15 2.5	10 2.2	6 2.0
0.9	18 2.7	15 2.5	13 2.4	11 2.3	7 2.1	5 2.0
0.0	5 2.0	6 2.0	5 2.0	5 2.0	5 1.9	4 1.9

Table 6: Results of Monte Carlo Experiment

Detrended Data

N = 200

Top Number = Percent rejections using the nominal five percent critical value of 2.0.

Bottom Number = Correct critical value for a five percent two-tailed test.

0 =	$\lvert\rho\rvert$ = 1.0	0.9	0.8	0.7	0.5	0.0
0.999	61 3.3	48 3.3	38 3.1	29 3.0	18 2.8	5 2.0
0.99	41 3.1	32 3.0	27 2.9	21 2.7	13 2.5	5 2.0
0.98	29 2.9	24 2.8	20 2.7	17 2.6	11 2.4	6 2.0
0.95	17 2.6	14 2.5	12 2.4	11 2.3	7 2.2	6 2.1
0.9	10 2.3	9 2.3	8 2.2	8 2.2	6 2.1	7 2.1
0.0	5 2.0	5 1.9	4 1.9	4 1.9	5 1.9	5 2.0

References

Abel, Andrew B. and Frederic S. Mishkin, 1983, An Integrated View of Tests of Rationality, Market Efficiency and the Short-run Neutrality of Money, Journal of Monetary Economics 11, 3-24.

Dickey, D. A. and W. A. Fuller, 1981, Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root, Econometrica 49, 1057-1072.

Dickey D. A. and W. A. Fuller, 1979, Distribution of the Estimators for Autoregressive Time Series with a Unit Root, Journal of the American Statistical Association 74, 427-431.

G. B. A. Evans, and N. E. Savin, 1981, Testing for Unit Roots: 1, Econometrica 49, 753-799.

G. B. A. Evans, and N. E. Savin, 1984, Testing for Unit Roots: 2, Econometrica 52, 1241-1269.

Flavin, Marjorie A., 1981, The Adjustment of Consumption to Changing Expectations about Future Income, Journal of Political Economy 89, 974-1009.

Flavin, Marjorie A., 1984, Time Series Evidence on the Expectations Hypothesis of the Term Structure, Carnegie-Rochester Conference on Public Policy 20, 211-238.

Frisch, Ragnar, and Frederick V. Waugh, 1933, Partial Time Regressions as Compared with Individual Trends, Econometrica 1, 387-401.

Geweke, John, 1981, A Comparison of Tests of the Independence of Two Covariance Stationary Time Series, Journal of the American Statistical Association 76, 363-373.

Granger, C.W.J., 1969, Investigating Causal Relations by Econometric Models and Cross-Spectral Methods, Econometrica 37, 424-438.

Granger, C.W.J., and Paul Newbold, 1974, Spurious Regression in Econometrics, Journal of Econometrics 2, 111-120.

Hall, Robert E., 1978, The Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence, Journal of Political Economy 86, 971-987.

Hansen, Lars Peter, and Kenneth J. Singleton, 1982, Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models, Econometrica 50, 1269-1286.

Hoffman, Dennis L., and Peter Schmidt, 1981, Testing the Restrictions Implied by the Rational Expectations Hypothesis, Journal of Econometrics 15, 265-288.

Hoffman, Dennis L., Stuart A. Low, and Don E. Schlagenhauf, 1984, Tests of Rationality, Neutrality, and Market Efficiency: A Monte Carlo Analysis of Alternative Test Statistics, Journal of Monetary Economics 14, 339-364.

Huizinga, John, and Frederic S. Miskkin, 1984, Inflation and Real Interest Rates on Assets with Different Risk Characteristics, Journal of Finance 39, 699-711.

Hurwicz, L., 1950, Least Squares Bias in Time Series, in T.C. Koopmans, ed., Statistical Inference in Dynamic Economic Models, New York: Wiley.

Jones, David S. and V. Vance Roley, 1983, Rational Expectations and the Expectations Model of the Term Structure, Journal of Monetary Economics 12, 453-465.

Mankiw, N. Gregory, and Matthew D. Shapiro, 1984, Trends, Random Walks, and Tests of the Permanent Income Hypothesis, Journal of Monetary Economics, forthcoming.

Mishkin, Frederic S., 1981, The Real Interest Rate: An Empirical Investigation, Carnegie-Rochester Conference on Public Policy 15, 151-200.

Nakamura, Alice O., Masao Nakamura, and Guy H. Orcutt, 1976, Testing for Relationships Between Time Series, Journal of the American Statistical Association 71, 214-222.

Nelson, Charles R., and Heejoon Kang, 1981, Spurious Periodicity in Inappropriately Detrended Time Series, Econometrica 49, 741-751.

Nelson, Charles R., and Heejoon Kang, 1983, Pitfalls in the Use of Time as an Explanatory Variable in Regression, NBER Technical Working Paper No. 30.

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

Nelson, Charles R., and and G. William Schwert, 1982, Tests for Predictive Relationships Between Time Series Variabes: A Monte Carlo Examination, Journal of the American Statistical Association 77, 11-18.

Sawa, T., 1978, The Exact Moments of the Least Squares Estimator for the Autogressive Model, Journal of Econometrics 8, 159-172.

Shiller, Robert J., 1979, The Volatility of Long-Term Interest Rates and Expectations Models of the Term Structure, Journal of Political Economy 82, 1190-1219.

Shiller, Robert J. and Pierre Perron, 1984, Testing the Random Walk Hypothesis: Power versus Frequency of Observation, Yale University, mimeo.

Sims, Christopher A., 1972, Money, Income and Causality, American Economic Review 62, 540-552.

Summers, Lawrence H., 1982, Do We Really Know That Financial Markets Are Efficient?, NBER Working Paper No. 994.