

NBER TECHNICAL PAPER SERIES

DO WE REJECT TOO OFTEN?  
SMALL SAMPLE PROPERTIES OF TESTS  
OF RATIONAL EXPECTATIONS MODELS

N. Gregory Mankiw

Matthew D. Shapiro

Technical Working Paper No. 51

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
June 1985

We are grateful to Donald Andrews, James Heckman, Adrian Pagan, Peter Phillips, and James Poterba for helpful discussions. The research reported here is part of the NBER's research program in Economic Fluctuations. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

Do We Reject Too Often?  
Small Sample Properties of Tests  
of Rational Expectations Models

ABSTRACT

We examine the small sample properties of tests of rational expectations models. We show using Monte Carlo experiments that the asymptotic distribution of test statistics can be extremely misleading when the time series examined are highly autoregressive. In particular, a practitioner relying on the asymptotic distribution will reject true models too frequently. We also show that this problem is especially severe with detrended data. We present correct small sample critical values for our canonical problem.

N. Gregory Mankiw  
Department of Economics  
Harvard University  
Littauer Center  
Cambridge, MA 02138

Matthew D. Shapiro  
Department of Economics  
28 Hillhouse Road  
Yale University  
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,

$$(1) \quad E_{t-1} Y_t = \phi_0.$$

We can equivalently write

$$(1') \quad Y_t = \phi_0 + \nu_t \quad \text{where} \quad E_{t-1} \nu_t = 0.$$

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

$$(2) \quad Y_t = \phi_0 + \phi_1 X_{t-1} + \nu_t$$

using ordinary least squares. According to equation (1), the coefficient  $\phi_1$  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:

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

Under the null hypothesis,  $X_{t-1}$  and  $\nu_t$  are uncorrelated. The model, however, does not preclude a contemporaneous correlation between  $\epsilon_t$  and  $\nu_t$ . 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

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

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  $\nu_{t-1}, \nu_{t-2}, \dots$ . 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.<sup>1</sup>

In this paper we examine the conditions under which the asymptotic theory leads to incorrect inference for samples of typical size. In Mankiw and Shapiro [1984], we examine this issue in the context of a specific application: tests of the permanent income hypothesis. We study here with Monte Carlo experiments the canonical problem described above. We show that when both the contemporaneous correlation,  $\rho$ , and the autoregressive parameter,  $\theta$ , are close to unity, the asymptotic test of the null hypothesis that  $\phi_1 = 0$  leads to a rejection too often.<sup>2</sup> 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.

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 ( $\theta \leq 0.9$ ), one should be wary of the asymptotic distribution. That is, if the forecasting variable is highly 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 practitioner relying on the asymptotic distribution will usually reject a true null hypothesis. We also provide correct critical values for detrended data.

Section 2 describes the Monte Carlo experiment, while Section 3 presents the results for stationary (non-detrended) data. Section 4 consider the problem when the data are detrended. Section 5 offers some concluding observations.

## 2. Monte Carlo Experiment

The critical parameters for each Monte Carlo experiment are  $\rho$  and  $\theta$ . Given these parameter values, a series of  $N$  innovations,  $\nu_t$  and  $\epsilon_t$ , are generated from a bivariate normal with variances equal to one and covariance

equal to  $\rho$ . The variable  $Y_t$  is set equal to  $\nu_t$ . The variable  $X_t$  is generated from equation (3) using the innovations  $\epsilon_t$ . 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, which allows us to estimate the distribution of the statistic. We present below the fraction of the time a practitioner using the asymptotic 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  $\tau$  is the t-statistic, then the five percent critical value is the number  $\Psi$  such that  $\text{Prob}(|\tau| \geq \Psi) = .05$  under the null hypothesis.<sup>3</sup>

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  $\epsilon_t$  is doubled, then all values of  $X_t$  are doubled. While this would alter the estimate of  $\phi_1$  in equation (2), it would not change the t-statistic. Second, allowing the constant  $\phi_0$  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  $\rho$ ,  $\theta$ , and the sample size  $N$ .

### 3. Results

Table 1 presents the results for sample sizes of  $N = 50$  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 asymptotic distribution. For values of  $\rho$  and  $\theta$  close to unity, the actual size is far greater than five percent. A practitioner using the asymptotic 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 asymptotic distribution theory.

The tables also present the correct critical value for a five percent test. If  $\rho$  and  $\theta$  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.

#### 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

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

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  $\alpha$  and  $\beta$  must be estimated.

Nelson and Plosser [1982] and Nelson and Kang [1981,1984] demonstrate that the assumption that  $X_t$  is stationary is itself not innocuous. Severe

biases can result if  $X_t$  in fact has a unit root. We assume here, however, that  $X_t$  is stationary 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).<sup>4</sup>

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  $\alpha$  and  $\beta$  must be estimated: if they were known,  $X_t$  could be observed and the problem would be identical to the one above.

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  $\alpha$  and  $\beta$ . Hence, we can set  $\alpha = \beta = 0$  without loss of generality.

Table 2 presents the results of the Monte Carlo experiment for sample sizes of 50 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  $\theta$  and  $\rho$  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  $\theta$  and  $\rho$  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 stationary 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 unwarranted 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]).

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

Notes

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 econometrician is using OLS to regress one serially correlated series on another independent serially correlated series. Because that model is misspecified (a lagged dependent variable is incorrectly omitted), the problem of spurious regression remains asymptotically. In contrast, we consider incorrect inferences about a correctly specified model caused by unwarranted reliance on distributions that are correct only asymptotically.

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. Note that this is not equivalent to a 2.5 percent one-tailed critical value, since the distribution of  $\tau$  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,  $\tau^2$ ).

4. This detrending procedure produces the same coefficient on detrended  $Z_t$  as the procedure of regressing  $Y_t$  on  $Z_t$  and a time trend together.

Table 1: Non-detrended Stationary Data

Top number is the percent rejections using the nominal five percent critical value of 2.0. Bottom number is the correct critical value for a five percent two-tailed test.

		<u>N = 50</u>						
		<u> ρ  =</u>	1.0	0.9	0.8	0.7	0.5	0.0
θ =	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

		<u>N = 200</u>						
		<u> ρ  =</u>	1.0	0.9	0.8	0.7	0.5	0.0
θ =	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 2: Detrended Data

Top number is the percent rejections using the nominal five percent critical value of 2.0. Bottom number is the correct critical value for a five percent two-tailed test.

N = 50

	$ \rho  = 1.0$	0.9	0.8	0.7	0.5	0.0
$\theta = 0.999$	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

N = 200

	$ \rho  = 1.0$	0.9	0.8	0.7	0.5	0.0
$\theta = 0.999$	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.
- 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.
- 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, 1984, Pitfalls in the Use of Time as an Explanatory Variable in Regression, Journal of Business and Economic Statistics 2, 73-82.
- 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.
- 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.