THE IRRELEVANCE OF TESTS FOR BIAS IN SERIES OF MACROECONOMIC FORECASTS

Roy H. Webb


The standard test for bias in a series of forecasts begins by estimating coefficients in the following equation:

\[ A_t = \alpha + \beta r_{t-1}P_t + \epsilon_t \]

where \( A_t \) is the actual value at time \( t \) of the variable predicted, \( r_{t-1}P_t \) is the prediction made at time \( t-1 \) for the value at time \( t \), \( \alpha \) and \( \beta \) are coefficients estimated by least squares, and \( \epsilon_t \) is an error term that is assumed to be from a series of independent and identically distributed normal random variables with zero mean. An F-test can then be used to test the joint hypothesis that \( \alpha = 0 \) and \( \beta = 1 \). If that hypothesis is rejected, the standard interpretation is that the series of forecasts is biased.

Most of the authors apparently believe that by examining the statistical bias of those time series, they are testing an important component of the new classical economics, the hypothesis of rational expectations. As Hafer put it, “Because [wealth-maximizing] agents presumably will not make forecasts that are continually wrong in the same direction, rational forecasts should be statistically unbiased.”

The assertion that bias is not consistent with rational expectations is examined below. First, two meanings of the term rational expectations will be presented. Several difficulties with interpreting the test for bias are discussed next; several limit the relevance of the test for the more important definition of rational expectations. Even if that test is interpreted as applying to the less important definition (a technical requirement adopted for analytical convenience), it is argued that the authors cited above have failed to consider several possible explanations for their results that do not contradict rational expectations. It is concluded that tests of macroeconomic predictions for statistical bias have not yielded useful information about the rationality of expectations.

Rational Expectations

The term rational expectations has become widely used; different authors, however, may attach different meanings to the term. This paper will focus on two ideas, one that is a general principle and the other a highly restricted form of the first. The general principle is that the actions of optimizing individuals lead to an absence of rents in equilibrium (in other words, profitable opportunities will be exploited). An important implication is that costly information will be used efficiently. This “informational efficiency” idea is crucial to economists often labeled as rational-expectations analysts. The restricted form of informational efficiency that is often used is “certainty equivalence,” which implies that a representative individual’s optimally predicted value of an economic magnitude can be identified with the mathematical expectation of a specific linear function that correctly describes the operation of the economy. The assumption of certainty equivalence helps economists

---

1 An even larger literature exists in finance, where tests for bias have been stimulated by the efficient markets hypothesis.

2 Hafer, page 3.
build theoretical models that are mathematically tractable; as the passages below indicate, however, it is not a critical idea for economists who have pioneered the use of rational expectations.

The [rational expectations] hypothesis asserts three things: (1) Information is scarce, and the economic system generally does not waste it. (2) The way expectations are formed depends specifically on the structure of the relevant system describing the economy. (3) A "public prediction," in the sense of Grunberg and Modigliani, will have no substantial effect on the operation of the economic system (unless it is based on inside information). . . . For purposes of analysis, we shall use a specialised form of the hypothesis. In particular, we assume: (1) The random disturbances are normally distributed. (2) Certainty equivalents exist for the variables to be predicted. (3) The equations of the system, including the expectations formulation, are linear. Muth [1961], p. 348. (Emphasis added.)

But it has been only a matter of analytical convenience and not of necessity that equilibrium models have used . . . the assumption that agents have already learned the probability distributions they face. [It] can be abandoned, albeit as a cost in terms of the simplicity of the model. Lucas and Sargent [1979], p. 13.

[The rational expectations] approach says that if people do not observe something directly—such as the current price level—then they form the best possible estimate of this variable, given the information that they possess. In other words people make efficient use of their limited data, so as not to commit avoidable errors. Barro [1984], pp. 468-69.

As Muth noted, the strong requirement of certainty equivalence was adopted for analytical convenience. Without certainty equivalence it is not necessary that optimal predictions are mathematical expectations. Also, note that the authors stress that individuals make the best possible use of information they possess—not that individuals have perfect information.

SEVEN REASONS WHY OPTIMAL FORECASTS CAN SHOW BIAS

This section explains why a series of predictions could appear biased even though they were originally prepared optimally. Many of the explanations have the common thread of asymmetric information. In some cases, the ex post reviewer uses more information than was actually available to forecasters when the forecasts were made. In others, the process of reviewing forecasts ignores relevant data that was available to forecasters. Failure to properly account for either of those informational asymmetries limits the relevance of tests for bias. The first four reasons below question the relevance of tests for bias as a test for both informational efficiency and certainty equivalence; the last three only apply to the strict requirements of certainty equivalence.

1. Unequal Data Availability: Real-Time Forecasts versus Ex Post Evaluation

In many cases economists have tested for biased predictions by comparing recorded forecasts with the latest available data. The data on which the forecasts were based, however, have often been revised substantially by the agencies that compile and report the data. In fact, it is possible that the ex post bias found in forecasts could be due to the reviewer having access to data revisions that were unavailable to real-time forecasters (that is, those who actually issued forecasts before the fact).³

Lupoletti and Webb [1986] noted that preliminary data on the rate of change of the GNP implicit price deflator were at one time biased predictors of the final data released. Since most findings of biased forecasts or surveys of expectations refer to the inflation rate, the biased original inflation data could explain many biased forecasts without contradicting their rationality.

To see whether early reports of the percentage change in the implicit price deflator were biased, consider:

\[ A_t = \alpha + \beta P_{t-1} + \epsilon_t \]

where \( A_t \) is the actual value at time \( t \) of the percentage change in the implicit price deflator from the previous quarter, \( P_t \) is the first data officially released for that percentage change, \( \alpha \) and \( \beta \) are

³ It is implicitly assumed that forecasters attempt to predict the true value that is estimated in official reports, and that successive revisions are usually closer to the true value than initial reports. The first assumption may not always be valid; consider a bond trader who is concerned about market changes in the first few minutes following a preliminary report.

⁴ The actual data are index numbers, based on 1982 = 100. They reflect all revisions through early 1987.

⁵ Approximately fifteen days after the beginning of each calendar quarter \( q \), the Commerce Department released its preliminary estimate of the implicit price deflator for the previous quarter \( q \). At that time a forecaster would also have had a value for the deflator two quarters earlier \( q-2 \), which would have been revised twice since its preliminary release. At the beginning of quarter \( q+1 \), therefore, the first official estimate of the change in the deflator between quarter \( q \) and quarter \( q-1 \) becomes available. It is that first official estimate that is used as the early series \( P_t \) in this section.
coefficients estimated by least squares, and \( \epsilon \) is an error term that is assumed to be from a series of independent and identically distributed normal random variables with zero mean. If the preliminary value \( P \), is an unbiased predictor of the latest revised value \( A \), then the estimate of the coefficient \( \alpha \) should be 0 and the estimate of \( \beta \) should be 1.

Table I contains regression results for equation (2) over the 1970s. The hypothesis of no bias is decisively rejected by a conventional F-test. Forecasters in the 1970s, therefore, should not be assumed to have had unbiased data on which to base their forecasts, given the subsequent revisions in the implicit deflator.

The implicit deflator is not the only measure of prices that has been studied. Leonard and Solt [1986] noted that the consumer price index diverged from other measures of consumer prices before 1983 due to the CPI's treatment of mortgage interest payments (which has been criticized by many analysts). They found that survey data which other authors had found to be biased were unbiased when compared against a better estimate of consumer prices.

The problem of biased initial data that is later revised is not confined to prices. Mork [1987] found that early releases of real GNP growth from 1968 through 1984 were biased. Since it is widely believed that real GNP is the best single statistic for describing the economy's performance, Mork's finding is particularly disturbing. Certainly many economists' expectations of other variables would be affected by the reported growth rate of real GNP.

Zarnowitz [1982] has not only found evidence of biased initial data releases for many time series, but also found "extraordinary divergences" among various data series describing real economic activity in 1973-74. Since most studies of forecasts or expectations include that period, confusion at that time could have a strong impact on the results of ex post studies.

None of the studies cited in the introduction attempt to determine the extent to which their results might be due to bias in the data available to forecasters at the time forecasts were prepared. Pearce [1984] and Zarnowitz [1985], however, do mention the problem of data available to forecasters.

2. Difficulty of Improving Real-Time Forecasts

It may seem that a biased series of forecasts would indicate that forecasters did ignore an easy method of improving forecasts: simply removing that bias. That, at least, is apparently the assumption of most of the articles cited.

Now suppose that a series of forecasts was found to be biased—that is, after the coefficients in equation (1) were estimated, the joint hypothesis of \( \alpha = 0 \) and \( \beta = 1 \) was rejected. As Theil [1966] has noted, a more accurate series of forecasts \( P' \) could then be constructed by adjusting the series \( P \):

\[
P' = \hat{\alpha} + \hat{\beta}P,
\]

(3)

where \( \hat{\alpha} \) and \( \hat{\beta} \) are estimates of the coefficients \( \alpha \) and \( \beta \) from equation (1). For example, if the predicted series was expressed in percentage points and a forecaster was on average one percentage point too high (\( \hat{\alpha} = -1 \)), then the adjusted forecast would subtract one percentage point from that forecaster's prediction. This would be an almost costless way of improving forecasts. Failing to use it would therefore seem to waste information.

The flaw in that argument is that it assumes that the coefficients of equation (3) were known to forecasters at the time of forecast. In fact, those coefficients could have been estimated only after the forecasts were issued. Now if the coefficients were stable over time, one could reasonably impute their knowledge to a forecaster, since after a few years the forecaster could have recognized the bias and estimated the coefficients. But if the coefficients were to change over time, then using historic data to estimate them would not necessarily improve forecasts, since estimates of \( \alpha \) and \( \beta \) would no longer be relevant.

---

**Table I**

**REGRESSION RESULTS: TESTS FOR BIAS IN PRELIMINARY DATA**

\[
A = 2.95 + 0.67 P,
\]

\[
(0.65) \quad (0.09)
\]

<table>
<thead>
<tr>
<th>Time span: 70:1 to 79:4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R^2 = .57 )</td>
</tr>
<tr>
<td>( DW = 2.18 )</td>
</tr>
<tr>
<td>F-statistic (for ( \alpha = 0 ) and ( \beta = 1 )) = 11.9</td>
</tr>
<tr>
<td>( F_{0.1,2.38} = 5.26 )</td>
</tr>
</tbody>
</table>

Notes: \( A \) is the actual inflation rate, measured with latest data. \( P \) is the inflation rate, based on the preliminary data release. Standard errors are in parentheses.
In other words, the analyst studying a sample of forecasts has more information than did the forecaster when the forecasts were prepared. Just because using some of the additional information could improve forecasts does not show that forecasters ignored potentially valuable data.

For example, Table II contains regression results for equation (1), using quarterly growth rates of the implicit price deflator from 1970 to 1984 as the actual series and the published forecast of Wharton Econometric Forecasting Associates as the forecast series. An F-test shows statistical bias; moreover, the residuals were significantly autocorrelated. Could Wharton have produced more accurate forecasts by using the Theil adjustment mechanism shown by equation (3)?

To answer that question, the first 25 observations were used to estimate equation (1), and the estimated coefficients were used to adjust the forecast as in equation (3). Next, one observation was added, equation (1) was reestimated, and the next quarter’s forecast was adjusted. That process was repeated until 32 adjusted forecasts were obtained. The adjusted forecasts had a slightly larger root mean squared error (1.565) than did Wharton’s published series (1.538). It therefore appears that Wharton did not waste the information from their past forecast errors even though a sample of its historic forecasts now appear biased.

Wharton’s inflation forecasts therefore provide a counterexample to the idea that a retrospective finding of bias proves that information was wasted. Of course, an author might still be able to show that other information—that was available to Wharton when its forecasts were prepared—could have improved its forecasts. The point is, that author would have to specifically identify the useful information that was wasted. A simple test for bias does not identify that information; moreover, any process that identified wasted information would probably make a test for bias superfluous. None of the authors cited in the introduction specifically identify the wasted information that could account for findings of bias.

### Table II

<table>
<thead>
<tr>
<th>REGRESSION RESULTS: INFLATION FORECASTS FROM WHARTON ECONOMETRICS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_t = 3.23 + 0.60 \cdot P_{t-1} + 0.37 \cdot u_{t-1}$</td>
</tr>
<tr>
<td>($0.94$) ($0.14$) ($0.13$)</td>
</tr>
</tbody>
</table>

**Notes:** $A$ is the actual inflation rate, measured with latest data. $P$ is the Wharton Econometrics forecast for the inflation rate, prepared at the end of the previous quarter. Standard errors are in parentheses.

3. **Average versus Marginal**

In many cases it is marginal behavior that determines economic outcomes. Studies of surveys of expectations, however, often focus on average behavior. The relevance of such studies was questioned by Mishkin [1981, p.295]:

> Not all market participants have to be rational in order for a market to display rational expectations. The behavior of a market is not necessarily the same as the behavior of the average individual. As long as unexploited profit opportunities are eliminated by some participants in a market . . . then the market will behave as though expectations are rational despite irrational participants in that market.

Mishkin tested the same survey data that were found to be inconsistent with rationality by Friedman. By focusing on marginal behavior, he found the data to be consistent with rationality. To explain Friedman’s results, he suggested that the survey data did not accurately describe actual behavior in the bond market.

Other studies of expectations also focus on average behavior. None of the authors cited in the introduction who study surveys of expectations examine whether their conclusions would change if they examined marginal behavior.
4. Uninformed Opinions of Irrelevant Aggregates

Although most economists could state an opinion for the future time path of macroeconomic variables, not all would be willing to bet money on their predictions. Equivalently, it is a trivial matter to put a number on a survey form; if an important decision were to be based on the data, however, careful and thoughtful analysis would probably precede any forecast.

Surveys of expectations do not necessarily measure solid analyses or even informed opinions that affect real decisions. Instead, it is quite possible that they contain relatively uninformed opinions of persons who will not make important decisions based on their expectations and accordingly have little incentive to acquire costly information. The relevance of any findings of bias in such surveys is questionable.

5. Nonstationary Data

Suppose that a data series $z$ is generated by the following random walk:

$$
(4) \quad z_t = z_{t-1} + \epsilon_t,
$$

where $\epsilon_t$ is from a series of independent and identically distributed normal random variables with zero mean. The mathematical expectation of $z_t$ at time $t-1$ is therefore $z_{t-1}$. But a sample of such forecasts could be found to be statistically biased if the coefficients of equation (1) were estimated, assuming that $A_t = z_t$ and $e_t P_t = z_{t-1}$. That is, equation (1) is misspecified if the actual data-generating process is given by (4). Significance tests from a misspecified equation can of course be misleading.

A Monte Carlo study illustrates that point. Equation (4) was used to generate 101 observations of $z_t$, where $z_t = 0$ and values of $\epsilon_t$ were randomly drawn from a normal distribution with a zero mean and a unit variance. Using $z_{t-1}$ as the forecast for $z_t$, since $E_{t-1}[z_t] = z_{t-1}$, equation (1) was estimated and an F-test performed for bias. The procedure was then repeated 999 times, thereby testing 1000 random walks of 100 observations each for bias. By construction there was no bias; yet in 189 cases, the hypothesis of no bias was rejected at the 5 percent level, and in an additional 139 cases it was rejected at the 10 percent level. That is, investigators would have found bias in many instances due to the inappropriate choice of a test statistic.

Once the possibility of nonstationary data is recognized, the burden of proof should be on the author to demonstrate that F-tests are valid. For example, Schwert [1987] discusses procedures that could be used to test time series for stationarity. In addition, Nelson and Plosser [1982] and Schwert have presented evidence that many macroeconomic time series appear to be empirically indistinguishable from random walks. Yet none of the authors cited in the introduction test for stationarity in the actual data series employed. That is especially troubling for those authors that examined predicted stock prices, since stock indexes are widely believed to follow random walks.

6. Peso Problems

If an unlikely event would make a dramatic impact on predicted outcomes, that event's likelihood can affect optimal forecasts, even if the event did not occur during a particular interval. In effect, the forecast contains a risk premium for the unlikely yet dramatic event. For example, if Russian investors in 1916 assigned a positive probability to a Bolshevik Revolution, stock prices of Russian firms in 1916 might appear lower than could be explained by observable factors such as earnings, dividends, and interest rates. In hindsight, such a forecast appears eminently rational. Krasker [1980] has noted that such peso problems can invalidate usual tests of efficiency in the foreign exchange market.

In studying macroeconomic forecasts, a particularly important event to consider is the possibility of a major policy regime change. The acknowledged possibility of a regime change could account for statistical bias over almost any specific interval. For example, downward-biased forecasts of inflation could be due to a positive probability placed on the Federal Reserve's adopting a monetary policy emphasizing price stability. Even if such a policy were not adopted during a particular time period, a forecaster's subjective probability of such a policy being adopted may have been correct.

---

8 To see why equation (1) would be misspecified, note that it assumes the existence of a fixed constant term, which is not consistent with the assumed random walk. A random walk can often drift far from the origin without ever crossing the origin in a fixed sample. In that case, a regression equation such as (1) will find a significant constant term and slope coefficient different from unity. Those findings, however, have no meaning for the future behavior of the random walk.

9 Indeed, even with stationary data that is highly autocorrelated, Mankiw and Shapiro [1986] have shown that conventional F-tests will reject true models too frequently.

10 This is labeled a peso problem due to a lengthy period when the Mexican peso traded in forward markets at a rate below the fixed spot rate, due to the widespread belief that a devaluation of the peso would eventually occur.
Although it is not possible to completely rule out peso problems, it is feasible to see whether plausible anticipated policy changes could account for findings of statistical bias. None of the authors cited in the introduction makes the attempt.

7. Representative Individual's Utility Function

Zellner [1986] has noted that, for many utility functions, an individual's optimal forecast can be biased. In particular, by accepting some bias it may be possible to lower the standard error of a point predictor and thereby lower the mean squared error of a series of forecasts. Also, if the loss of utility from an over-prediction does not precisely equal the loss of utility from an under-prediction of the same magnitude, then an unbiased forecast may not maximize utility. Zellner provides a specific example to illustrate the latter point. Stockman [1987] derives a loss function for forecast errors from an agent's exact decision problem, finding that in general such loss functions will not value over- and underpredictions equally.

None of the studies cited in the introduction provide evidence that a representative individual's utility function is maximized by an unbiased forecast. That key point is simply assumed.

CONCLUSION

Many authors have tested for bias in surveys of macroeconomic expectations or time series of forecasts. Although the authors believed they were testing the rationality of expectations, there are many reasons why they could have found bias. Seven reasons are listed above that are seldom examined, that are likely to affect the results of conventional tests, and that have little relevance to important economic questions. Some of the reasons are due to the reviewer using information that was not available to forecasters. Others are due to the reviewer not using relevant information that was available to forecasters. Since any finding of bias could be due to at least one of the reasons given above, the relevance of such tests is questionable.

The convenient assumption of certainty equivalence can be appropriately tested, once careful attention is given to data available to real-time forecasters. The fundamental idea of informational efficiency is much harder to test. It has not been, and almost certainly cannot be, properly examined by simple tests for biased expectations or forecasts.

References


---

**Economic Review Index**

**Volume 73, 1987**

**Federal Reserve Bank of Richmond**

**January/February**

- Intraday Credit: Risk, Value, and Pricing
- Repurchase and Reverse Repurchase Agreements
- The Agricultural Outlook for 1987
  . . . Financial Turnaround Unlikely

**March/April**

- The Theory of Multiple Expansion of Deposits: What It Is and Whence It Came
- The Equilibrium Approach to Exchange Rates
- IPC or Total Deposits? There is a Difference

**May/June**

- Swiss and United States Monetary Policy: Has Monetarism Failed?
- The Effect of Exchange Rate Variation on U.S. Textile and Apparel Imports
- Cost Dispersion and the Measurement of Economies in Banking
- An Overview of Agricultural Policy . . . Past, Present, and Future

**July/August**

- Henry Thornton: Seminal Monetary Theorist and Father of the Modern Central Bank
- Classical and Neoclassical Roots of the Theory of Optimum Tariffs
- A Review of Bank Performance in the Fifth District, 1986

**September/October**

- The Fed's Anti-Inflationary Strategy: Is It Adequate?
- The Case for Rules in the Conduct of Monetary Policy: A Concrete Example
- The Australian Money Market and the Operations of the Reserve Bank of Australia: A Comparative Analysis

**November/December**

- The Irrelevance of Tests for Bias in Series of Macroeconomic Forecasts
- Money Growth Volatility and High Nominal interest Rates

**FEDERAL RESERVE BANK OF RICHMOND**

---

David L. Mengle, David B. Humphrey, and Bruce J. Summers
Stephen A. Lumpkin
Raymond E. Owens
Thomas M. Humphrey
Alan C. Stockman
Donald L. Welker
Georg Rich
Christine Chmura
David B. Humphrey
Raymond E. Owens
Robert L. Hetzel
Thomas M. Humphrey
John R. Walter and David L. Mengle
Robert P. Black
Bennett T. McCallum
Michael Dotsey
Roy H. Webb
Yash Mehra