The Validity of Forecasting

Dr Garrick R. Small (garrick.small@uts.edu.au) and
Raymond Wong (Raymond.S.Wong@uts.edu.au)

University of Technology, Sydney,
Australia

August 2001
Abstract

Considerable interest has been shown over recent decades in the application of quantitative methods to the forecasting of key property variables. The dominant method to emerge has been multiple regression analysis. This paper considers the various approaches that have been taken to test forecast validity in an attempt to study the validity of forecasts using regression and other methods. The paper reviews the literature on tests of forecast validity and considers methodological issues that may be behind the apparent poor performance of most published forecasts.

The paper concludes with recommendations for future directions in forecasting research. In particular, it is suggested that simpler, easier to understand, forecasts may more desirable for many reasons, especially as they appear to perform at least as well as complex approaches. It would also appear that greater emphasis on tests of forecast validity should be encouraged as a prerequisite for the publication of any forecasting studies.

Keywords:

Property forecasting; forecast validity; regression; financial forecasting;

INTRODUCTION

Many published research on forecasting ignores formal evaluation procedures and simply presents possible but untested approaches. In other cases, even valuation is done, it tends to be narrow and with the intention of advocating a particular method.

Few authors discuss the principles for evaluating forecasting methods (Cox and Loomis 2001). Most discuss error measures such as MSE, RMSE and MAPE, and suggest readers to use multiple measures to test for accuracy.

What is an appropriate measure for assessing forecast accuracy? We observe that many property researchers have paid too little attention to forecast accuracy or they have chosen the measures arbitrarily, with the result that flawed error measures are widely used.

This paper aims to discuss the validity of forecasting along the theme of “positive economics” proposed by Milton Friedman (1953), as opposed to elaborating on forecasting methods which focus on particular techniques used. We also highlight some of the more popular accuracy measures, pinpoint their deficiencies, and argue that in the absence of a consensus measure any effort to search for the ‘best’ forecasting method will be in vain.
FORECASTING METHODOLOGY

What is forecasting?

According to Armstrong (2001), forecasting is the prediction or estimate of an actual value in a future time period (for time series) or for another situation (for cross-sectional data). Forecasting is concerned with what the future will look like, rather than what it should look like. The latter is the job of planning, using forecasts as inputs.

While a forecast is often required in decision making, it is possible to avoid forecasting completely. Forecasts may not be needed in situations where we can control the future (e.g., we need not to forecast the weather inside our house); we can respond to changes rapidly (as in driving an automobile); and future risks that can be hedged or minimised by insurance, diversification or contingency planning.

Forecast Horizon

The accuracy of forecast models can be measured over two distinct horizons: the ex post and the ex ante forecast periods (see Figure 1 above). In an ex post forecast, observations on both the dependent (forecast outcome) and independent (explanatory) variables are known with certainty during the forecast period (Pindyck and Rubinfeld. 1998; Armstrong 2001). Thus, ex post forecasts can be checked against existing actual data to assess how well the model would do if the best possible forecasts were made of the causal variables.
An *ex ante* forecast provides values of the dependent variable beyond the estimation period and it does not use actual values of explanatory variables from later periods. This term, often used interchangeably with *unconditional forecast*, is what we think of as a forecast. It provides the best measure of a model’s ability to forecast and offers a benchmark for comparison.

**Forecasting Models versus Measurement Models**

A *model* is a representation of the real world by summarising the theory relevant to a situation under consideration for empirical measurement and test with facts and data. A *forecasting model* is a formal statement about variables and relationships among variables in the production of forecasts. A *measurement model* is used to obtain estimation of parameters (the “true” values of an unknown population) from data. A forecasting model may draw upon a variety of measurement models for estimates of key parameters.

Some authors do not draw a distinction between the two types of models but are concerned with model fitting. There is a misconception that those measurement models with a high $R^2$ (coefficient of determination in regression analyses) are considered valid and accurate for forecasting purposes.

**Mathematical assumptions in forecasting**

Each forecasting method is based on assumptions about the behaviour of the variables involved and precision with which the mathematical relations adopted actually map them.

All forecasting is typically based on the assumption that the future is going to be like the past. This assumption has been reliably applied to the natural sciences, but has proven highly problematic in the social sciences. Empirical studies in economics appear to have done at least as much to contradict received theory than confirm it. For example, (Jones 1976) surveyed the empirical rejection of the marginalist theory of pricing that underlies the market supply curve. The empirical definition of the demand curve was been first challenged by (Working 1927) In the property area there appears to be little empirical support for the theoretical belief that interest rates negatively affect property prices (Oluwoye and Small 2001). The economic history of the last century contains several examples of the future not conforming to the experience of the past. The occurrence of stag-flation, economic stagnation simultaneous with monetary inflation, which erupted in the last quarter of the twentieth century was something that economic theory and experience believed impossible.

Simple time series forecasts carry the burden of flaws in this assumption. In addition, time series forecasts believe that past trends will extrapolate uniformly into the future. Despite usually being tolerably reliable, this assumption fails to predict turning points and turning points are the critical event in successful investment speculation.
The most common analytical tool used for forecasting is multiple regression analysis. In addition to the foregoing, this requires extensive assumptions about the linearity and causal interrelationships between variables. The dependent variable should not be a function of their own previous values (autocorrelation), but this assumption contradicts time series approaches and the expectation of cyclic behaviour. The independent variables should be orthogonal, that is, they should not interact with each other as indicated by multicollinearity. Financial and economic variables are believed by theory to be highly interactive and this is commonly evident in quantitative studies. Regression has a difficulty with turning points that is usually overcome by assuming that the cyclic behaviour originates in one or more of the independent variables and relying on lagging them to align the phasing with the independent variable. This assumes constant phase relations between variables. Finally, regression assumes usually assumes linear causality between the dependent variable and the dependent variable. Reflection on the actual causal mechanisms suggests that this is unlikely in fact.

Neural networks appear to supply a more flexible and promising approach because they do not make many of the assumptions that bedevil regression. They appear to have the capacity to accommodate more complex interactions between the variables, however it will be seen that they have not yet produced convincing results. Part of the reservation concerning their use is that they abandon any meaningful assumptions about what connects the variables. This is consistent with Friedman’s (1953) suggestion that positive economics abandon attempts to understand what actually causes the various interactions that compose economic theory and concentrate only on models that predict the future tolerably well.

The problem with this approach is that it can never be trusted, because the actual mechanics that causally connect the variables may be totally misunderstood. In some cases, connections may be reliably observed for some time, only to be radically overturned due to some unconsidered effect. For example, I may find that every time I place a key in a slot in an automobile, it moves. This may lead to a theory connecting the effort of key-turning to the propulsion of the automobile. I may add other factors to my theory including the additional energy supplied by pressing pedals and working the gear lever. The theory may work very well until I run out of fuel.

Work in weather prediction has offered valuable insights into the understanding and prediction of financial variables. The weather changes constantly between certain bounds. If local climate is considered to be the combination of several variables, such as temperature, pressure, wind speed and direction, it has been found that for a region, each variable tends to be a function of the values of the previous values of itself and the others. These can be modelled by sets of simple mathematical equations. This appears to parallel the situation in finance where current financial decisions are human decisions that are made in the light of available financial data. That is, financial decisions produce the more or less closed set of financial and economic variables and these decisions are functions of previous values of the variable that comprise the closed set. If this is true, then insights from weather prediction should apply to financial forecasting. The observations about the way that financial variables are produced describe the very conditions that give rise to chaotic behaviour (Gleick 1987).
The difficulty with chaotic systems is that while the causal relations can be understood fairly simply, it is very difficult to use this understanding for forecasting. A prohibitively large amount of data is required. (Newell and Matysiak 1997) investigated property data but was not able to find conclusive evidence of chaotic behaviour. His result did not discount the likelihood of a chaotic system, but rather that the data was not sufficiently fine-grained to reveal it definitively. Newell did supply an interesting example of a simple mathematical equation that could produce chaotic results despite being mathematically definitive. The equation $y_{t} = a \cdot y_{t-1} - a \cdot y_{t-1}^2$ may return various linear and chaotic results depending on the value of the parameter $a$ and the initial value of $t$. With $a=3.86$ and the initial value of $y=0.5$, the pattern shown in fig 2 is produced. The unsettling thing about the pattern is that it produces a regular, pseudo-cyclic trace for about four cycles which then disappears for a number of periods before reappearing. While this trace suggests that the cyclic behaviour returns, exploration of the function beyond the time shown in the graph reveals that the traces continually mutates, sometimes quite suddenly. Interested readers would be well advised to experiment with this function.

What can be learned from chaos is that it is an accurate explanation of the actual phenomena being investigated, past trends do not map into the future, despite being capable of precise mathematical analogy. For the property investor, this means that if property variables are existentially chaotic, which appears rationally likely, any deterministic forecasting tool, such as regression, may prove dangerously misleading.

![Chaotic trace: $a=3.86$, $y_0=0.5$](image)

If data quality can be improved and tools for matching chaotic patterns made more powerful, this approach to forecasting may produces useful results. Given the problem of obtaining data, it may be some time before this is realised. Overall, chaos theory should be included in property education, at least to make property graduates aware that it does provide a promising explanation for the fundamental causal mechanisms that drive financial systems. Armed with such an awareness, practitioners may be better positioned to prudently
attempt forecasting using the simpler, though less reliable, tools that are commonly employed. Chaos will not be considered further in this paper. The validity of forecasting will always be in question so long as researchers have no convincing theory to explain why the variables behave as they do. To the extent that some forecasting tools may have moderate predictive power, they will continue to provide an attractive way to circumvent this logical necessity. However, it should always be remembered that they are necessarily unreliable and must rest wholly on their positive achievements. The historical performance of forecasting tools is therefore most important. A review of the attempts to study forecast accuracy is therefore necessary.

ACCURACY MEASUREMENTS

“What is an appropriate measure for assessing the accuracy of a forecast?” is perhaps one of the most contentious issues among researchers over the last fifty years.

The quality (accuracy) of a model can be estimated by examining the inputs (assumptions) to the model, or by comparing the outputs (forecasts) from the model. Friedman (1953) claims that testing outputs is the only useful approach to evaluating forecasting methods. Nagel (1963) criticised Friedman’s position as unreasonable. Machlup (1955) goes to the other extremes by implying that the testing of input is the only worthwhile way to test models. We think it is more reasonable to test both inputs, for improvement of a model, and outputs, for selection of the best model.

Forecasting accuracy is regarded as an “optimist’s term for forecast errors” by Armstrong (2001). A forecast error, on the other hand, represents the difference between the forecast value and the actual value.

Armstrong (2001) and Ahlburg (2001) suggest that the following factors should be taken into consideration when selecting a measurement of accuracy:

Prediction intervals: the errors should be obtained from a test that closely resembles the actual forecasting situation
The error term should not be overly influenced by outliers
The term should be independent of scale
The error measures should be sensitive to changes in the model being tested
Reliability
Validity

In the measurement of accuracy, we should not overlook the practical aspects from a forecast user’s perspective. For example, the ability of a model (leading indicator) in predicting the timing for a turning point is sometimes considered more important than forecasting the magnitude of an expected change in trend. Also, the extent of downside risks is of more interests than the scale of upside gains.
The Validity of Forecasting  2002 PRRESS

Garrick Small & Raymond Wong

Coefficient of Determination ($R^2$)

Regression, the analysis of relationships among variables, is probably the most frequently used analytical tools in econometric models developed for property forecasting purposes. The Least-Squares Method (or OLS, Ordinary Least Square method), based on the Gauss-Markov Theorem and a set of assumptions (e.g. normality, homoscedastisity, no autocorrelation, nonstochastic) is most commonly used. In its simplest form, a linear regression is represented by:

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \epsilon \]

where \( Y \) = value of the dependent variable
\( \beta_0 \) = intercept or constant
\( \beta_1 \ldots \beta_n \) = slope of the regression line
\( X_1 \) = given value of the independent variable
\( \epsilon \) = observed error or residual

The Coefficient of Determination ($R^2$) is the square of the correlation between $Y$ (the forecast variable) and $\hat{Y}$ (the estimated $Y$ value based on the set of explanatory variables). It denotes the explanatory power of $X$ for variations in $Y$. An adjusted $R^2$ is the coefficient of determination adjusted for degrees of freedom.

In spite of its limitations, occasional accidents and incidental pollution, Stigler (1981) claims that the OLS method is the automobile of modern statistical analysis known to and valued by all. Unlike neural networks, which are regarded as a black-box forecasting technique, the same results in terms of $R^2$, constant and slope can always be replicated from a given data set.

However, as Milton Friedman concludes, after learning it in a hard way when his newly designed alloys based on regression analysis ruptured within less than four hours of use, statistical measures of a model’s ability to fit historical data provide little evidence about its ability to predict with new data (Friedman and Schwartz 1991). A long list of research (Ferber 1956; Schupack 1962; Mayer 1975; Elliott and Baier 1979) shows that the fit of a model to time series data is a poor way of assessing predictive validity. Pant and Starbuck (1990) found a modest relationship of only 0.2 to 0.49 between fit and forecast accuracy using the 1,001 time series of the M-competition.

Figure 3
Armstrong (2001) also suggests that $R^2$ and adjusted $R^2$ should not be used for forecast accuracy comparisons. As shown in Figure 2 above, it is possible that the actual and forecast values of a model are moving in tandem in terms of both direction and magnitude, thus showing a perfect $R^2$, but they are quite separately apart from each other.

**Mean Squared Errors (MSE)**

The mean squared error is an accuracy measure computed by squaring the individual error for each item in a data set and then finding the average or mean value of the sum of those squares (Makridakis, Wheelwright et al. 1998):

$$MSE = \frac{1}{n} \sum_{t=1}^{n} e^2_t$$

where

- $MSE = \text{mean squared error}$
- $n = \text{time periods}$
- $e^2_t = \text{forecast error}$

The MSE is having the advantage of being easier to handle mathematically. By using Excel’s Solver, it is quite easy to adjust forecasting parameters to come up with a “best fit” to historical data, i.e. finding the optimal weighting factors by minimising the MSE values.

However, since a MSE is in absolute value instead of in percentage, it does not facilitate comparison across different time series of various time intervals. It also gives greater weight to large errors than to smaller ones because the errors are squared (magnified) before being summed. Since the ability of a forecasting method to detect large errors is often regarded as one the most important criteria, the MSE method has been popular for years. However, Armstrong (2001) argues that the MSE should not be used for forecast comparisons because it is not independent of scale and it is unreliable when compared to other measures.
Mean Absolute Percentage Error (MAPE)

The mean absolute percentage error is the mean or average of the sum of all of the percentage errors for a given data set taken without regard to sign so as to avoid the problem of positive and negative values cancelling one another (Makridakis, Wheelwright et al. 1998):

\[
PE_t = \left( \frac{Y_t - F_t}{Y_t} \right) \times 100
\]

\[
MAPE = \frac{1}{n} \sum_{t=1}^{n} |PE_t|
\]

where 
- PE = percentage error
- \(Y_t\) = actual observation for time period \(t\)
- \(F_t\) = forecast for the same period
- MAPE = mean absolute percentage error
- \(n\) = time periods

The MAPE measure is less sensitive to outlier distortions and allows for a direct comparison between different forecasting methods. However, it has a bias favouring underestimates, i.e. a forecast of zero can never have more than 100% MAPE but there is no limit to errors on overestimates.

Theil’s U-statistic

The U-statistic developed by Theil (1966) is an accuracy measure that emphasises the importance of large errors (as in MSE) as well as providing a relative basis for comparison with naïve forecasting methods. Makridakis et al. (1998) have simplified Theil’s equation to the form shown below:

\[
U = \sqrt{\frac{\sum_{t=1}^{n-1} \left( \frac{F_{t+1} - Y_{t+1}}{Y_t} \right)^2}{\sum_{t=1}^{n-1} \left( \frac{Y_{t+1} - Y_t}{Y_t} \right)^2}}
\]
where $U = \text{Theil’s U-statistic}$

$F = \text{forecast}$

$Y = \text{observation}$

Theil’s U-statistic can be interpreted as dividing the RMSE (Root Mean Square Error, or square root of the MSE) of the proposed forecasting method by the RMSE of a no-change (naïve, $U=1$) model. If $U$ is equal to 1, it means that the proposed model is as good as the naïve model. If $U$ is greater than 1, there is no point in using the proposed forecasting model since a naïve method would produce better results. It is worthwhile to consider using the proposed model only when $U$ is smaller than 1 (the smaller the better), indicating that more accurate forecasts than a no-change model can be obtained.

**Popularity of Accuracy Criteria**

During the 1981 International Symposium on Forecasting, Carbone and Armstrong (1982) asked a group of academics and practitioners for the most preferred forecast criteria and accuracy measures. The most favoured criteria were accuracy, ease of interpretation and implementation. The most popular measures among the experts were MSE and RMSE while Theil’s U and $R^2$ were the least favoured. The reasons given for the unpopularity of the U-statistic included difficulty to understand, little need for it and other measures could do the job more simply.

It is worthwhile to note that the Carbone and Armstrong survey was done some 20 years ago and attitude nowadays may have changed. Even Armstrong (2001, pp 460) himself is now a strong opponent against using MSE and RMSE due to their “poor reliability and validity”. He proposes using “Percent Better” and “Relative Absolute Errors” as alternative accuracy measures, which will be subject to the test of time and are not elaborated in this paper. On the other hand, until further research reveals otherwise, we tend to agree with authors such as Ahlborg (2001) that no single accuracy measure is appropriate for all situations.

**EMPIRICAL EVIDENCE**

Over the last 20 years, notably the M-competition organised by Spyros Markridakis et al (1992; 1993 and 2000), many studies have compared the post-sample forecasting accuracies of all major forecasting methods using real-life business, economic, financial, demographic and other data sets. In the absence of a generally accepted accuracy measure, as noted in the foregoing sections, it is almost impossible to select the ‘best’ forecasting method with the most accurate results. Some key findings of past studies are summarised as follows.

*Simple versus complex methods:* Generally speaking, the post-sample accuracy of simple methods is at least as good as that of complex or statistically sophisticated methods.
(Armstrong 1984; Schnarrs 1984; Armstrong 1985; Makridakis, Wheelwright et al. 1998; Allen and Fildes 2001). Simplicity in an econometric model refers to a small number of causal variables and that its parameters are linear. For extrapolations, a deseasonalised random walk (naïve) model can outperform Box-Jenkins and Bayesian forecasting methodologies. Therefore, it appears that there is no reason, other than to impress clients or forecast users, to use a more complex method for forecasting unless its use is supported by substantial evidence of accuracy improvements. Simpler methods, on the other hand, aid decision makers’ understanding and implementation, reduce the likelihood of mistakes, and are less expensive.

A recent survey (Higgins 2000) of property analysts in Australia reveals that they tend to use simple rather than complex methods in forecasting: 14% of them rely on judgement (based on market knowledge); 21% use time series (trends, moving averages); and 65% adopt a casual approach (economic modelling).

Econometric methods: After comparing empirical studies published in the literature, Armstrong (1985) concluded that the accuracy of econometric methods was not significantly better than time series methods.

Multivariate models: The performance of multivariate models such as ARIMA, Vector autoregressive is not found to be more accurate (Rise and Tsotheim 1984; McNees 1986).

Non-linear models may provide a better fit to historical data but do not necessarily improve the forecast accuracy (De Gooijer and Kumar 1992). An over-fitted model will force the regression to go through every point, i.e. the $R^2$ is high but the error is forced to zero. It does not correspond to real world situations where a purely random component (chance variable) that cannot be foreseen and predicted always exists.

Adaptive methods: The general belief that adaptive models can learn and therefore will forecast more accurately is not supported by empirical evidence (Gardner and Dannenbring 1980).

CONCLUSION

This paper in fact raises more questions than it answers. There is a methodological tension between the demands of rigorous science for causal explanation and the practical necessity to be able to provide useful forecasts in the face of limited data and inappropriate analytic tools. This tension results in the need for careful study of forecast accuracy, and the practical selection of forecast methods that balance simplicity with effectiveness. It highlights certain areas for further research in respect of first finding an appropriate yardstick (accuracy measure) before continuing to look for the ‘best’ forecasting method. We argue that a good statistical model should be parsimonious, i.e. it should use as few mathematical terms as possible to describe real world situations.
The credibility of forecasting will be strengthened if its users find it simple, easy to understand, and can be implemented with minimal cost and time. Once the trust in forecasting is established, firms are more willing to devote further resources in either performing the necessary forecasts in-house or buying them from outside research houses. As researchers, we stand to gain more from demystifying the process of forecasting than being perceived as charlatans.

REFERENCES


*Extra citations:*