

406347 UK Ökonometrische Prognose
SS 2004
University of Vienna
Department of Economics

EVALUATING FORECAST ACCURACY

0001078 Elisabeth Woschnagg
9501689 Jana Cipan

CONTENT

1	INTRODUCTION	2
1.1	MODEL-BUILDING PROCEDURE USING THE SAME DATA SET	2
2	MODEL UNCERTAINTY	3
2.1	DATA DRIVEN INFERENCE	4
2.2	SOME RESEARCH FINDINGS	4
2.2.1	<i>Inferential biases – narrow prediction intervals</i>	4
2.3	WAYS OF GETTING MORE REALISTIC ESTIMATES OF PREDICTION ERROR	5
2.3.1	<i>Bayesian model averaging</i>	6
3	COMPARING OUTCOME AND PREDICTIONS	7
3.1	MINIMUM MSFE	7
3.1.1	<i>Allowing for parameter uncertainty</i>	8
3.1.2	<i>MSFEs for scalar multi-step systems</i>	9
3.2	TESTING FOR UNBIASEDNESS	10
3.3	DATA SPLITTING	11
3.4	RIVAL FORECASTS	12
4	RELEVANT MEASURES IN FORECAST EVALUATION	12
4.1	STAND-ALONE ACCURACY MEASURES	13
4.1.1	<i>RMSE</i>	13
4.1.2	<i>Decomposition of the MSE</i>	13
4.1.3	<i>MAE</i>	14
4.1.4	<i>MAPE</i>	14
4.2	RELATIVE MEASURES	15
4.2.1	<i>The standardized root mean-squared error</i>	15
4.3	STABILITY TO LINEAR TRANSFORMATION	16
	REFERENCES	17

1 Introduction

The history of the evaluation of forecast accuracy goes along with that of time-series analysis. The first tests for forecasting models were developed in 1939 by Tinbergen, in response to Keynes, who stated that theories must be confirmed if the data and statistical methods are employed correctly.

A crucial criticism is the Lucas Critique. It states that the future development is influenced by forecasts, because expectations are self-fulfilling. This produces a circuit and raises the questions how forecasts should take into account the fact of self-fulfilling prophecies in time-series forecasting. This theory implies that forecasts are informational input for the data generating process (DGP), and that they are invalidated by agents reacting to them. Therefore, forecasts are susceptible to biasedness. Opponents of the Lucas critique claim that forecasts are no probability-based techniques that point into the future, but rather extrapolative patterns. In any statistical problem, the three main sources of uncertainty are:

- Uncertainty about the estimates of the model parameters, assuming that the structure of the model is known
- Uncertainty about the data, for example unexplained random variations of the observed variable or measurement errors.
- Model uncertainty: Uncertainty about the structure of the model, for example because the model is misspecified a priori or because the assumption that the model parameters are fixed is wrong. Standard analysis ignores model uncertainty.

1.1 Model-building procedure using the same data set

The Box-Jenkins model-building procedure suggests the following proceeding for econometric data analysis. Model specification means formulating a sensible time-series model that is a plausible approximation and predicts future data with adequate precision.

After assuming a sensible time-series model to be correct, the model is fitted. Parameters which are unobservable are estimated, and strategies are used such as excluding, down-weighting and adjusting outliers or transformations to achieve normality or constant residual variance. Biases in the model-fitting process carry on to predictions. For example, parameters may be biased using the same set of data for estimation and for model selection.

The third step is to check for the model's rationality – also referred to as calibration. The shorter the interval is for which the prediction is made, the more input variables for model selection and fitting one gets. Still, uncertainty may be reflected better by choosing wider intervals for prediction.

An issue that is criticized with emphasis by Chatfield is that problems arise from formulating, fitting and testing a model using the same set of data. The least squares theory does so. Biases arising from this procedure are called model-selection biases.

2 Model uncertainty

The standard time series literature assumes that there is a true model for a given time series and that this model is known before it is fitted to the data. After fitting the 'true' model to the data the same model is then used for forecasting. Regarding econometric models, the model may be misspecified a priori or the parameters assumed to be fixed when they change through time. In addition, in time series forecasting the uncertainty problem arises because the model is defined, fitted and tested using the same set of data.

Chatfield (2000) departs from the assumption that a true model does not exist. The main task would then be to find a model that provides an adequate approximation to the given data. One model, that seems to fit the underlying data best, may be selected as a 'winner' although the other models give a fit very close to the one selected. The properties of an estimator may depend, not only on the selected model, but also on the selection process.

2.1 *Data driven inference*

Data dredging is the general process of selecting a model from a large set of candidate models, and then used for inference and forecasting¹. An example for data-dredging is the search for calendar effects in stock market behavior. Analysts may be able to discover some regularity when looking at financial series over different time periods. But this regularity may fail to generalize to other time periods or to other similar variables. The analyst tests this effect because it has been spotted in the data and it may prove significant. This type of data-driven inference is likely to produce spurious results and suspect forecasts. Based on such results rules for investing in the stock market may be recommended, such as ‘Sell on the first day of the month’ or ‘Buy on Friday’.

Testing for the presence of unit roots, autocorrelated residuals, presence of break points, indicates model uncertainty. The more models are seen as potential candidates of a true model and the more testing is carried out on the same data used to fit the model, the more inference will be biased. In evaluating the forecasts the question arises how this uncertainty about the model will affect the estimates of the forecast accuracy.

2.2 *Some research findings*

If a time-series model is selected by minimizing the within-sample prediction mean square error (PMSE) then the Optimism Principle applies. The fitting of a model gives optimistic results in that the performance on new data is on average worse than on the original data. In particular, the fit of the best-fitting model is typically better than the resulting accuracy of out-of-sample forecasts.

2.2.1 Inferential biases – narrow prediction intervals

Empirical studies have shown that the prediction intervals are too narrow in that 95% - prediction intervals will contain less than 95% of actual future observations. If a best-fitting

¹ Another possible description is data snooping. The expression data mining has been abandoned by statisticians due to its application as knowledge discovery in very large databases by computer scientists. Chatfield (2000), ch. 8.

model is chosen from many alternatives, the residual variance, and hence the PMSE, are underestimated. The predictive variance conditional on the selected model is being taken for calculating the prediction intervals.

The example of Draper² is concerned with forecasting the price of oil. It demonstrates that conditioning on a single model or scenario can underestimate the uncertainty in forecasts. The problem was to forecast the 1986 price per barrel of crude oil from data up to 1980. 10 different econometric models and 12 scenarios (low demand elasticity or a drop in OPEC production) were considered. He found that the prediction intervals for an individual scenario did not reflect the uncertainty. For example, the 90% P.I. for the reference scenario was from 27 \$ to 51 \$. The actual price per barrel in 1986 turned out to be 11 \$. This was outside the different P.I.s that were computed.

Nowadays different computational procedures such as simulation, resampling or cross validation are applied because it is difficult to establish in theory how uncertainty about the model will affect the estimates of the forecast accuracy. Chatfield (2000) cites the example of Hjorth (1987) who simulated data from an ARMA (1,1) model, but found that the correct form of the ARMA model was identified in only 28 of 500 series. For the 472 series other alternative models were identified as correct. By allowing to select an ARMA model, the estimate of forecast accuracy – the PMSE – was optimistic compared to the true forecast accuracy, PMSE, of the ARMA (1,1) model. The average estimated prediction MSE was less than one third of the true PMSE for the model that was actually fitted.

2.3 Ways of getting more realistic estimates of prediction error

In order to get more precise estimation errors, the following guidelines can be used for overcoming the effects of model uncertainty.

In choosing a single model one can distinguish between local and global models. Models with constant values and parameters through time, such as regression models with constant coefficients are called global. Local models are models that allow parameters to adapt through

² Chatfield, 2000, ch.7.

time. Local models are often fitted by some kind of updating procedure, such as the Kalman Filter, which is easy to apply using a computer.

The model building Principle of Parsimony states that the simpler the model, the more adequate is the representation of the data. Complicated models may reduce bias but at the same time may increase variance.

Sensitivity analysis involves making small changes to the model assumptions in order to see how stable the deductions (also the forecasts) from the model are.

When several competing models seem plausible, one can make use of more than one. If the pattern of the data changes through time, e.g. changing seasonality or structural breaks, different models can be used for different parts of the data. As opposed to univariate models, multivariate models have a higher chance of detecting sudden changes. Different models can be used for different lead times. h -steps-ahead forecasts require minimizing prediction errors h -steps-ahead. An application would be if a short-term forecasting model differs from a long-term forecasting model.

2.3.1 Bayesian model averaging

Given there exists a set of plausible models, prior probabilities can be attached to them, so that the data can be used to evaluate posterior probabilities for the different models. Models with low posterior probabilities may be left out for the sake of simplicity. Empirically, the weighted sum of the predictions of the competing models has a much lower PMSE in the long run than the forecasts from the individual models. An example was Hjorth's simulation of ARMA (1,1) model³. In the example of forecasting the oil price, Draper went on by using a Bayesian model averaging approach. The 90% P.I. gave a range between 20 \$ and 92 \$ when both, model and scenario uncertainty, were taken into account. This was much more realistic than the P.I.s found for individual models and scenarios. The disadvantages are that there may be difficulties specifying prior probabilities. Posterior probabilities require computation by means of the Bayes factor, which complicates the model.

³ Section 2.2.1., p. 5.

When a time series model is formulated and fitted to the same data, then inferences and forecasts made from the fitted model will be biased and over-optimistic when the prior data-dependent model-selection process is ignored. Forecasting aims at producing predictions that are as close as possible to the actual data. The deviation of the predictions from the real data should be minimized.

Unbiasedness and efficiency are a minimum requirement for optimal or rational forecasts. Weak rationality refers to forecasts that are not systematically wrong. Strong rationality or efficiency requires that forecast errors are uncorrelated with the series or the information available at the time when the forecast was made. Rationality does not rule out serial correlation in the error process e_{T+h} . It is a necessary attribute to make forecasts useful.

3 Comparing Outcome and Predictions

Ex post forecast accuracy can be written as a function I of the actual values of a series (A_t) and the predicted values (P_t), for which $I(\cdot)$ obtains a maximum. For example:

$$I(P_t, A_t), I(A_t - P_t, A_t) = I(e_t, A_t), C(e_t)$$

where

$$e_t = A_t - P_t$$

C are the costs of a forecast error, which only depend on e_t , and which is a quadratic function in order to guarantee mathematical tractability. Real-world cost functions of forecast errors are rarely available. The minimum mean-square forecast-error (MMSFE) minimizes

$$E[e_t^2] \text{ in a sample}$$

$$MSFE_h = \frac{1}{H} \sum_{h=1}^H e_{T+h}^2$$

In further discussion we will come back to this model.

3.1 Minimum MSFE

Taking formula $E[e_t^2]$ from the beginning of chapter 3, we can reformulate:

$$MSFE = E[e_{T+h}^2] = (\mu_a - \mu_p)^2 + \sigma_a^2 + \sigma_p^2 - 2r\sigma_a\sigma_p$$

and derive from this equation the characteristics of the minimum MSFE predictor given by the solutions of the first-order conditions:

$$\frac{\partial MSFE}{\partial \mu_p} = -2(\mu_a - \mu_p) = 0$$

$$\frac{\partial MSFE}{\partial \sigma_p} = 2(\sigma_a - r\sigma_p) = 0$$

$$\frac{\partial MSFE}{\partial r} = -2\sigma_a\sigma_p = 0$$

The second-order conditions for a minimum are satisfied only by the first two conditions. We know that our model is minimized by $\mu_a = \mu_p$ and $\sigma_p = r\sigma_a$. The standard deviation of the two series is equal only when the predictor series is perfectly correlated with the actual series ($r=1$) Substituting these gives:

$$MSFE_{\min} = \sigma_p^2 \left(\frac{1}{r^2 - 1} \right)$$

Assuming that the process is weakly stationary and has finite unconditional moments of at least second order, one can write:

$$MSFE_h = E[e_{T+h}^2] = E[(y_{T+h} + \hat{y}_{T+h})^2]$$

The MSFE combines the squared bias in forecast errors with the forecast-error variance.

$$\begin{aligned} E[e_{T+h}^2] &= (E[e_{T+h}])^2 + E[(e_{T+h} - E[e_{T+h}])^2] \\ &= (E[y_{T+h} - \hat{y}_{T+h}])^2 + V[e_{T+h}] \end{aligned}$$

3.1.1 Allowing for parameter uncertainty

Even though it is very restrictive, we assume a model such as

$$y_t = \kappa + \rho y_{t-1} + v_t \quad \text{with } e_t = v_t$$

such that the (theoretical) prediction and the (theoretical) forecast error are uncorrelated, $E[P_t e_t] = 0$.

Allowing for parameter uncertainty, such that

$$P_{T+1} = \hat{\kappa} + \hat{\rho} y_T$$

gives the expected value of correlation between prediction and forecast error:

$$E[e_{T+1} P_{T+1}] = E[(\rho - \hat{\rho}) \hat{\rho} y_T^2] + E[\hat{\rho} y_T v_{T+1}]$$

Since $E[\hat{\rho} y_T v_{T+1}] = 0$, and approximating $E[\hat{\rho}] = \rho - \frac{2\rho}{T}$

$$E[e_{T+1} P_{T+1}] = E[(\rho - \hat{\rho}) \hat{\rho} y_T^2 | y_T] = (\rho E[\hat{\rho}] - E[\hat{\rho}^2]) y_T^2$$

$$\begin{aligned} &\approx \left(\frac{2\rho^2}{T} - V[\hat{\rho}] \right) y_T^2 \\ &\approx -\frac{1}{T} (1 - 3\rho^2) y_B^2 \end{aligned}$$

The variance of the given model for $e_{T+1} = (\rho - \hat{\rho})y_T v_{T+1}$ is

$$\begin{aligned} V[e_{T+1}] &= V[\hat{\rho}]V[y_T] + V[v_{T+1}] \\ &\approx T^{-1} (1 - \rho^2) \sigma_v^2 / (1 - \rho^2) + \sigma_v^2 \\ V[e_{T+1}] &= \sigma_v^2 \left(1 + \frac{1}{T} \right) \end{aligned}$$

The variance therefore is a function of the unknown quantity σ_v^2 .

3.1.2 MSFEs for scalar multi-step systems

Multi-step MSFEs do not allow for different but isomorphic representations. A common basis for comparison is not sufficient. It will be proofed that when 1-step forecasts bring the same result for the MSFE, this is not necessarily the case for 2 (or more) period forecasts.

Consider the stationary first-order autoregression

$$y_t = \rho y_{t-1} + u_t$$

where $u_t \sim \text{IN}[0, \sigma_u^2]$, $|\rho| < 1$ and $y_0 \sim \text{N}[0, \sigma_u^2 / (1 - \rho^2)]$ so the MSFE for the conditional 1-step forecast is

$$E[(y_{T+1} - \hat{y}_{T+1})^2 | y_T] = E[u_{T+1}^2] = \sigma_u^2$$

Instead one can use the differences for forecasts. $\Delta y_t = y_t - y_{t-1}$ is forecast using $\hat{\Delta} y_{T+1} = (\rho - 1)y_T$

$$E[(\Delta y_{T+1} - \hat{\Delta} y_{T+1})^2 | y_T] = E[u_{T+1}^2] = \sigma_u^2$$

For conditional 1-step forecasts the same MSFE results.

In a two-period setting the following holds.

$$y_{t+2} = \rho y_{t+1} + u_{t+2} = \rho^2 y_t + u_{t+2} + \rho u_{t+1}$$

Using $E[y_{T+2} | y_T] = \hat{y}_{T+2} = \rho^2 y_T$ the MSFE is:

$$E[(y_{T+2} - \hat{y}_{T+2})^2 | y_T] = E[(u_{T+2} + \rho u_{T+1})^2] = (1 + \rho^2) \sigma_u^2$$

Whereas in terms of changes :

$$\Delta y_{t+2} = (\rho - 1)y_{t+1} + u_{t+2} = (\rho - 1)(\rho y_t + u_{t+1}) + u_{t+2}$$

Using the conditional expectation $E[\Delta y_{T+2} | y_T] = \hat{\Delta} y_{T+2} = (\rho - 1) \rho y_T$, the MSFE is

$$E[(\Delta y_{T+} - \hat{\Delta} y_{T+2})^2 | y_T] = E[(\rho - 1)u_{T+1} + u_{T+2}]^2 = \sigma_u^2 (1 + (\rho - 1)^2)$$

This term is larger than $(1 + \rho^2)\sigma_u^2$ if $\rho < 1/2$ and larger than it if $\rho > 1/2$.

Continuing this comparison with calculating the variances, in the 1-step setting we get:

$$V[e_{T+1} | y_T] = \sigma_n^2 + V[\hat{\rho}] y_T^2$$

and for the two-step forecast the variance is

$$V[e_{T+1} | y_T] = E[(\rho^2 - \hat{\rho}^2)^2] y_T^2 + (1 + \rho^2) \sigma_u^2$$

In the next step the variances are calculated. Assuming parameter uncertainty we get to a similar result.

3.2 Testing for unbiasedness

Tests of unbiasedness are often based on a regression equation of the form

$$A_{T+h} = \alpha + \beta P_{T+h|T} + \epsilon_{T+h}$$

The null hypothesis $\alpha = 0$ and $\beta = 1$ entails unbiasedness.

$$E[A_{T+h}] = \alpha + \beta E[P_{T+h}]$$

The requirement for unbiasedness is $E[A_{T+h} - P_{T+h}] = 0$.

This test checks for efficiency at the same time, because it proves uncorrelation between the forecasts and their errors.

One can define forecast efficiency as the condition that $\beta = 1$, so that the residual variance in the regression equals the variance of the forecast error.

If $\alpha = (1 - \beta) E[P_{T+h}]$ then we have

$$e_{T+h} = A_{T+h} - P_{T+h} = \alpha + (\beta - 1) P_{T+h} + \epsilon_{T+h}$$

then

$$V[e_{T+h}] = (\beta - 1)^2 V[P_{T+h}] + V[\epsilon_{T+h}] + 2(\beta - 1) C[P_{T+h}, \epsilon_{T+h}]$$

So that $\beta = 1$ implies $V[e_{T+h}] = V[\epsilon_{T+h}]$ whatever the value of α . If $\alpha = 0$ then

$$MSFE = E[e_{T+h}^2] = V[e_{T+h}] + V[\epsilon_{T+h}]$$

Under these conditions the forecast error and predictor are uncorrelated:

$$E[P_{T+h} e_{T+h}] = \alpha E[P_{T+h}] + (\beta - 1) E[P_{T+h}] + E[P_{T+h} \epsilon_{T+h}]$$

When the data are non-stationary, integrated series⁴, then a further requirement is that the actual and predicted series are cointegrated, which means that the linear combination of two non-stationary variables is stationary. Otherwise the forecast would have unbounded variance. If we come back to the model $y_t = \kappa + \rho y_{t-1} + v_t$ from the beginning of chapter 3, and assuming $\kappa = 0$ and $\rho = 1$ gives us a random walk for y_t . It is intuitive that this will satisfy the conditions for forecast efficiency. Getting consistent with the notation of the test of unbiasedness, it follows that

$$y_{T+h} = \alpha + \beta y_{T+h-j} + \epsilon_{T+h}$$

where

$$y_{T+h} = y_{T+h-j} + \sum_{s=0}^j v_{T+h-s}$$

so that

$$y_{T+h-j} = \alpha + \beta y_{T+h-j} + (\epsilon_{T+h} - \sum_{s=0}^j v_{T+h-s})$$

3.3 Data splitting

Data splitting means checking out-of-sample forecast accuracy. The set of data is divided into two parts: the construction (or calibration) sample and the validation (or hold-out) sample. The construction sample is used to choose and fit a model. A prediction over the following period is made, where the forecast data can be compared with the validation sample. The data during time 1 to N are collected, and a point in time between 1 and N is chosen, which partitions the sample into the construction sample (N – m) and the validation sample m. In the next step the construction sample is used to find and fit a model, which predicts the values in the validation sample. At N-m the value for N-m+j is predicted where j goes from 1 to m. This produces a prediction for all values between N-m+1 and N. Two problems are linked to data splitting: firstly the question how data should be split, and secondly the fact that again the same set of data is used for defining fitting and checking a model. There are no guidelines of how to split the data. The drawback is efficiency loss because the model is fitted only to part of the time-series. But it is widely used for checking over-optimistic forecasts. Analysts usually hold back 10% of the data but this has no theoretical meaning.

⁴ Integrated: non-stationary series can be differentiated so as to make them stationary. The undifferenced series is the integrated series.

3.4 *Rival forecasts*

Simple time series models and stationary predictors are commonly used for forecasting, despite their limited validity. According to Clements and Hendry, stationary series, autoregressive-distributed lag models or vector autoregressive representation (VARs) are preferred over univariate time-series models. They mention four models. Clements and Hendry judge the no-change rule of Theil to be the most naive forecast. The second is the Box- Jenkins model, described as the best available univariate extrapolative model. Referring to Mincer and Zarnowitz, the third method is defining relative efficiency (RE) as the ratio of the MSFEs and the rival model (a benchmark model that is extrapolative in that it is based on the variable's own history). Forth, Salmon and Wallis show that structural econometric models can be approximated by finite-order ARMA models.

There are cases in which there is no unambiguous superior model. Models can dominate each on different measures, or dominance may alter with forecast horizons or the choice of variables.

4 **Relevant measures in forecast evaluation**

Fair (1986) discusses the three most common measures of predictive accuracy, namely RMSE, MAE, and Theil's inequality coefficient U. These measures have been used to evaluate the accuracy of ex post as well as ex ante forecasts. An ex post forecast is one in which the actual values of the exogenous variables are used; an ex ante forecast is one in which guessed values of the exogenous variables are used. There are two main reasons why ex ante forecasting comparisons are of little interest. First, the ex ante forecasts are based on guessed rather than actual values of the exogenous variable(s). Thus the error cannot be separated into the part due to bad guesses and the part due to other factors. The second reason, according to Fair (1986), is the use of subjective add factors in forecasts. 'It is thus the accuracy of the forecasting performance of the model builders rather than of the models that are being examined.'⁵

⁵ Fair (1986), ch. 33, p.1985.

Zhuo Chen and Yuhong Yang (2004) distinguish between stand-alone and relative accuracy measures. Stand-alone accuracy measures are those that can be obtained without additional reference forecasts. They are usually associated with a certain loss function, e.g. accuracy measures that are based on quadratic and absolute loss functions. Accuracy measures based on the mean square error criterion, especially MSE itself, have been the most preferred measures in evaluating forecasts for a single series.

4.1 Stand-alone accuracy measures

4.1.1 RMSE

The RMSE depends on the scale of the dependent variable. It should be used as relative measure to compare forecasts for the same series across different models. The smaller the error, the better the forecasting ability of that model according to the RMSE criterion.

$$RMSE = \sqrt{\frac{1}{h+1} \sum_{t=s}^{s+h} (\hat{x}_{t-1}(1) - x_t)^2}$$

One problem associated with the use of the RMSE or similar measures is the fact that the forecast error variance vary across time. It can vary because of nonlinearities in the model and because of variation in exogenous variables (if included in the model). Fair (1986) states that no rigorous statistical interpretation can be put on the RMSEs because they are not estimates of any parameter in the model⁶.

4.1.2 Decomposition of the MSE

The MSE mean squared forecast error can be decomposed as

$$\sum (\hat{x}_{t-1}(1) - x_t)^2 / h = (\bar{\hat{x}} - \bar{x})^2 + (s_{\hat{x}} - s_x)^2 + 2(1-r)s_{\hat{x}}s_x$$

⁶ Although they are in a loose sense estimates of the averages of the variances across time, Fair (1986), ch. 33, p.1987.

$\bar{\hat{x}}$ and \bar{x} are the means of the forecasted variable and of the true variable respectively. $s_{\hat{x}}$ and s_x are the standard deviations of the forecasted variable and of the true variable, respectively. r is the correlation between \hat{x} and the true variable x . The proportions are defined as:

The bias proportion $\frac{(\bar{\hat{x}} - \bar{x})^2}{\sum (\hat{x}_t - x_t)^2 / h}$ measures how far the mean of the forecast is from the mean of the actual series. The variance proportion $\frac{(s_{\hat{x}} - s_x)^2}{\sum (\hat{x}_t - x_t)^2 / h}$ measures how far the variance of the forecast is from the variance of the actual series. The covariance proportion $\frac{2(1-r)s_{\hat{x}}s_x}{\sum (\hat{x}_t - x_t)^2 / h}$ measures the remaining unsystematic forecasting errors. If the forecast is 'good', the bias and the variance proportions should be small so that most of the bias should be concentrated on the covariance proportions⁷.

4.1.3 MAE

The Mean absolute error MAE is also dependent on the scale of the dependent variable but it is less sensitive to large deviations than the usual squared loss.

$$MAE = \frac{1}{h+1} \sum_{t=s}^{s+h} | \hat{x}_{t-1}(1) - x_t |$$

4.1.4 MAPE

Another popular accuracy measure is the Absolute Percentage Error (MAPE). The MAPE is scale independent. However, MAPE was criticized for the problem of asymmetry and instability when the original value is small. MAPE as accuracy measure is affected by four problems: (1) Equal errors above the actual value result in a greater APE; (2) Large percentage errors occur when the value of the original series is small; (3) Outliers may distort the comparisons in empirical studies; (4) MAPEs cannot be compared directly with naive models such as random walk.

⁷ Acc. to Pindyck & Rubinfeld (1996). Prof. Kunst disagreed with this opinion in his lecture on 16th June 2004.

$$MAPE = \frac{1}{h+1} \sum_{t=s}^{s+h} \left| \frac{\hat{x}_{t-1}(1) - x_t}{x_t} \right|$$

4.2 Relative measures

The idea of relative measures is to evaluate the performance of a forecast relative to that of a benchmark forecast. Measures may produce very big numbers due to outliers and/or inappropriate modeling, which in turn make the comparison of different forecasts not feasible or not reliable. A shock may make all forecasts perform very poorly, and stand-alone measures may put excessive weight on this period and choose a measure that is less effective in most other periods. Relative measures may eliminate the bias introduced by potential trends, seasonal components and outliers, provided that the benchmark forecast handles these issues appropriately. However, choosing the benchmark forecast is subjective and not necessarily easy. The earliest relative forecast accuracy measure seems to be Theil's U2-statistic, of which the benchmark forecast is the value of the last observation.

4.2.1 The standardized root mean-squared error

The idea of the standardized root mean-squared error – denoted by U_2 – is to facilitate comparison between different variables. This equation is referred to as the no-change rule of Theil.

$$U_1 = \frac{\sqrt{\frac{1}{h+1} \sum_{t=s}^{s+h} (\hat{x}_{t-1}(1) - x_t)^2}}{\sqrt{\frac{1}{h+1} \sum_{t=s}^{s+h} \hat{x}_{t-1}(1)^2 + \frac{1}{h+1} \sum_{t=s}^{s+h} x_t^2}}$$

The Theil coefficient is scale invariant and it lies between zero and one. If the Theil coefficient equals zero then we have a perfect fit.

4.3 Stability to linear transformation

Stability of accuracy measures is another issue worth of consideration. As Clements and Hendry (1993) suggested, stability of accuracy measures with respect to the linear transformation of the original series is an important factor.

Zhuo Chen and Yuhong Yang (2004) use a series of monthly Austria/U.S. foreign exchange rate from January 1998 to December 2001. The original series is measured as how many Austrian Schillings are equivalent to one U.S. Dollar. It is calculated as the average of daily noon buying rates for cable transfers payable in foreign currencies. The authors perform a linear transformation of the original series by subtracting the mean of the series and multiply it with 10, i.e., $y^{new} = 10 \cdot y^{original} - 10 \cdot \text{mean}(y^{original})$

There are 4 forecasts generated by a random walk, an ARIMA(1,1,0), an ARIMA(0,1,1), and a forecast generated by a model selected based on BIC criterion. Table 1 presents the change of the values produced by the accuracy measures using the last 20 points. The first two accuracy measures – MAPE and RMSE – produced very different values after the transformation since they are not scale transformation invariant. The last accuracy measure – Theil's inequality coefficient – had some minor changes. This suggested that the first two measures are generally not good for cross-series comparison of forecasting procedures since a linear transformation of the original series may change the ranking of the forecasts.

Table 1 – Comparison of different criteria after transformation

Forecast series	RW		ARIMA(1,1,0)		ARIMA(0,1,1)		ARIMA(BIC)	
	original	new	original	new	original	new	Original	new
MAPE	0,024	0,302	0,023	0,306	0,022	0,296	0,03	0,381
RMSE	0,428	4,278	0,431	4,305	0,421	4,207	0,569	5,489
Theil's U2	1	1	1,006	1,006	0,983	0,983	1,331	1,283

REFERENCES

Christopher Chatfield (2000), *Time-series forecasting*, Chapman & Hall/CRC, 2000, Boca Raton, ch. 6, 8.

Michael P. Clements and David F. Hendry (1998), *Forecasting Economic Time Series*, Cambridge Univ. Press, ch. 3, 13.

Ray C. Fair (1986), *Handbook of Econometrics*, Vol. III, ed. Z. Griliches and M.D. Intriligator, Elsevier Science Publishers BV, ch. 33.

Schlittgen, Rainer (1995), *Zeitreihenanalyse*, Rainer Schlittgen u. Bernd H. J. Streitberg, Oldenbourg.

Zhuo Chen and Yuhong Yang (2004), *Assessing Forecast Accuracy Measures* Iowa State University.