Forecasting

in Economics, Business, Finance and Beyond

CH 10: Forecast Evaluation

Which Forecasting Method Should You Use

Gather the historical data of what you want to forecast

Divide data into initiation set and evaluation set

Use the first set to develop the models

Use the second set to evaluate

Compare measures of forecast accuracy for each model

2

How can we compare across forecasting models?

We need a metric that provides estimation of accuracy

Forecast Error

Errors can be:

1. biased (consistent)
2. random

Forecast error = Difference between actual and forecasted value (also known as *residual*)

3

Measuring Accuracy: MFE

MFE = Mean Forecast Error (Bias)

It is the average error in the observations

$$MFE = \frac{\sum_{i=1}^{n}(A_i - F_i)}{n} = \frac{\sum_{i=1}^{n} e_i}{n}$$

1. A more positive or negative MFE implies worse performance; the forecast is biased.

4

Measuring Accuracy: MAD

MAD = Mean Absolute Deviation

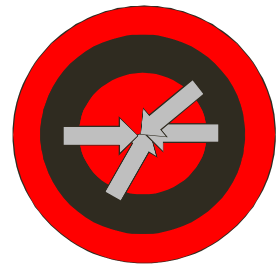
It is the average absolute error in the observations

$$MAD = \frac{\sum_{i=1}^{n} |A_i - F_i|}{n} = \frac{\sum_{i=1}^{n} |e_i|}{n}$$

1. Higher MAD implies worse performance.
2. If errors are normally distributed, then σ_ε=1.25MAD

5

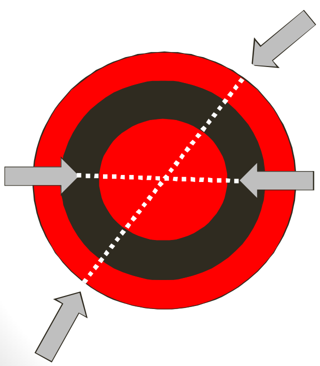
MFE & MAD: A Dartboard Analogy



Low MFE & MAD:

The forecast errors are small & unbiased

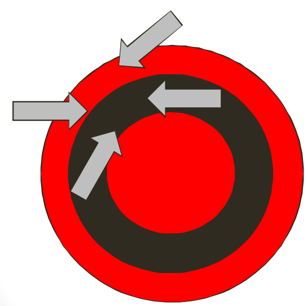
An Analogy (cont'd)



Low MFE but high MAD:

On average, the arrows hit the bullseye (so much for averages!)

An Analogy (cont'd)



High MFE & MAD:

The forecasts are inaccurate & biased

Mean Squared Error of Prediction

Mean Squared Error of Prediction (MSE_P)

$$MSE_P = \frac{\sum_{i=1}^{n}(A_i - F_i)^2}{n} = \frac{\sum_{i=1}^{n} e_i^2}{n}$$

Root Mean Squared Error of Prediction ($RMSE_P$): fixes units

$$RMSE_P = \sqrt{MSE_P}$$

9

MSE_P vs. MAD

Note that the process of squaring of each error gives you a much wider range of numbers.

The greater range gives you a more sensitive measure of the error rate, which is especially useful if the absolute error numbers are relatively close together and reduction of errors is important.

Measuring the extent of deviation helps determine the need to improve forecasting or rely on safety stock to meet customer service objectives.

10

MSE_P vs. MAD: example

Month	Demand	Forecast: Exponential Smoothing (0.4)	Error	Absolute Error	Squared Error
January	32				
February	26				
March	12				
April	5				
May	4	14	-10	10	100
June	3	10	-7	7	49
July	2	7.2	-5.2	5.2	27.04
August	5	5.12	-0.12	0.12	0.01
September	10	5.07	4.93	4.93	24.3
October	15	7.04	7.96	7.96	63.36
November	25	10.22	14.78	14.78	218.45
December	32	16.13	15.87	15.87	251.86
		Totals	**21.22**	**65.86**	**734.02**
		MAD		**8.23**	
		MSE_P			**91.75**
		$RMSE_P$			**9.58**

11

Better than MAD: Mean Absolute Percentage Error (MAPE)

There is a drawback to the MAD calculation, in that it is an absolute number that is not meaningful unless compared to the forecast.

MAPE is a useful variant of the MAD calculation because it shows the ratio, or percentage, of the absolute errors to the actual demand for a given number of periods.

$$MAPE(\%) = \frac{100}{n}\sum_{i=1}^{n}\left(\frac{A_i - F_i}{A_i}\right)$$

12

Better than MAD: Mean Absolute Percentage Error (MAPE)

Note that the result is expressed as a percentage.

Exception rules for review can be applied to any stock keeping unit or product family that has a MAPE above a certain percentage value.

Percentage – based error measurements such as MAPE allow the magnitude of error to be clearly seen without needing detailed knowledge of the product or family, whereas when an absolute error in units (or an error in $ amount) is provided, it requires knowing what is considered normal for the product or product family.

MAPE: example

Month	Demand	Forecast: Exponential Smoothing (0.4)	Error	Absolute Error	APE
January	32				
February	26				
March	12				
April	5				
May	4	14	-10	10	250
June	3	10	-7	7	233.33
July	2	7.2	-5.2	5.2	260
August	5	5.12	-0.12	0.12	2.4
September	10	5.07	4.93	4.93	49.3
October	15	7.04	7.96	7.96	53.07
November	25	10.22	14.78	14.78	59.12
December	32	16.13	15.87	15.87	49.59
		Totals	21.22	65.86	956.81
		MAD		8.23	
		MAPE (%)			119.6

Key Point

Forecasts must be measured for accuracy!

The most common means of doing so is by measuring either the mean absolute deviation or the standard deviation ($RMSE_P$) of the forecast error

Measuring Accuracy: Tracking signal

The tracking signal is a measure of how often our estimations have been above or below the actual value. It is used to decide when to re-evaluate using a model.

Running Sum of Forecast Errors (RSFE): update after each observation

$$RSFE = \sum_{i=1}^{n} e_i$$

Tracking Signal (TS): update after each observation

$$TS = \frac{RSFE}{MAD}$$

Measuring Accuracy: Tracking signal

Positive tracking signal:

- most of the time actual values are above our forecasted values

Negative tracking signal:

- most of the time actual values are below our forecasted values

If TS > 4 or < -4, *investigate!*

17

Example: bottled water at Kroger

Month	Actual	Forecast
Jan	1,325	1,370
Feb	1,353	1,361
Mar	1,305	1,359
Apr	1,275	1,349
May	1,210	1,334
Jun	1,195	1,309

Exponential Smoothing
(α = 0.2)

Month	Actual	Forecast
Jan	1,325	1370
Feb	1,353	1306
Mar	1,305	1334
Apr	1,275	1290
May	1,210	1251
Jun	1,195	1175

Forecasting with trend
(α = 0.8)
(δ = 0.5)

Question: Which one is better?

18

Bottled water at Kroger: compare MAD and TS

	MAD	TS
Exponential Smoothing	70	- 6.0
Forecast Including Trend	33	- 2.0

We observe that FIT performs a lot better than ES

Conclusion: Probably there is trend in the data which Exponential smoothing cannot capture

19

Choosing a Forecasting Algorithm

This is a difficult question! Real data does not follow any model, so smallest MSE forecasts may not in fact have smallest MSE_P.

Some general advice can however be given. First identify what **measure of forecast error** is most appropriate for the particular situation at hand. One can use mean squared error, mean absolute error, one-step error, 12-step error, etc. Assuming enough (historical) data is available, we can then proceed as follows:

- Omit the last k observations from the series, to obtain a reduced data set called the **training set**.
- Use a variety of algorithms and forecasting techniques to predict the next k observations for the training set.

20

Choosing a Forecasting Algorithm

- Now compare the predictions to the actual realized values (the **test set**), using an appropriate criterion such as **root mean squared error of prediction ($RMSE_P$)**
- Use the forecasting technique/algorithm that gave the smallest value of $RMSE_P$ for the test set, and use it on the original data set (training + test set) to obtain the desired out-of-sample forecasts.

Example: The Accidental Deaths Data

The file DEATHSF.TSM contains the original series plus the next 6 realized values $Y_{73},\dots,Y_{78}$. Using DEATHS.TSM, we obtain $P_{72}Y_{73},\dots,P_{72}Y_{78}$ via each of the following methods (and compute corresponding RMSE's):

Forecasting Method	RMSE
HW	1143
SARIMA model	583
Subset MA(13)	501
SHW	401
ARAR	253

The ARAR algorithm does substantially better than the others for this data.

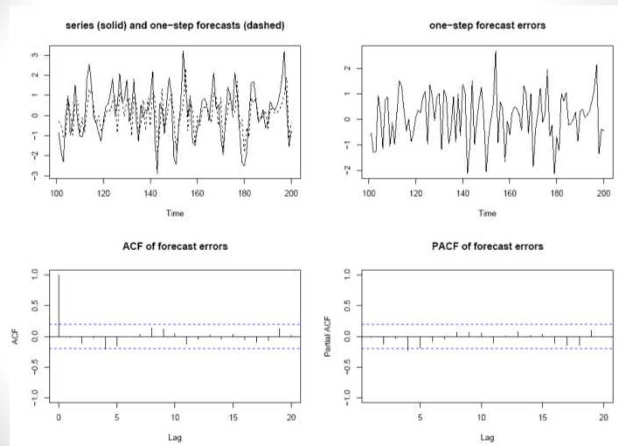
Forecast Monitoring

If the original model fitted to the series up to time *n* is to be used for ongoing prediction as new data comes in, it may prove useful to monitor the one-step forecast errors for evidence that this model is no longer appropriate. That is, for t = n+1, n+2, …, we monitor the series: $\widehat{Z}_t = X_t - \widehat{X}_t$

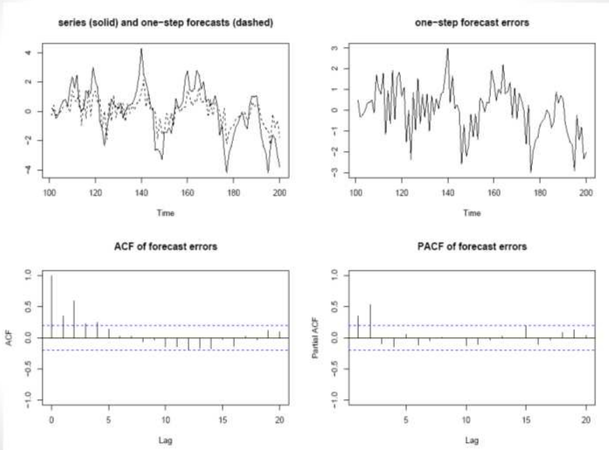
As long as the original model is still appropriate, the series $\widehat{Z}_t$ should exhibit the characteristics of a WN sequence. Thus one can monitor the sample ACF and PACF of this developing series for signs of trouble, i.e. *autocorrelation*.

Example: Observations for t = 1, …, 100 were simulated from an MA(1) model with $\theta = 0.9$. Consider what happens in the following two scenarios corresponding to the arrival of new data for t=101,…,200, stemming from two different models.

Case 1: New data continues to follow the same MA(1) model



Case 2: New data switches to an AR(1) model with $\phi = 0.9$



25

If you are interested….

I STRONGLY recommend that you read Chapters 10 – 14 of the text!

Other recommended reading:

Evaluation of forecasting techniques and forecast errors: *With focus on intermittent demand*

By Peter Wallström

Forecasting with Confidence

From KPMG International, a **global** network of professional firms providing Audit, Tax and Advisory services