Forecast Accuracy and Model Choice

Keith Ord

Georgetown University

GEORGETOWN

McDONOUGH

School of Business

UNIVERSITY

Topics to be Discussed

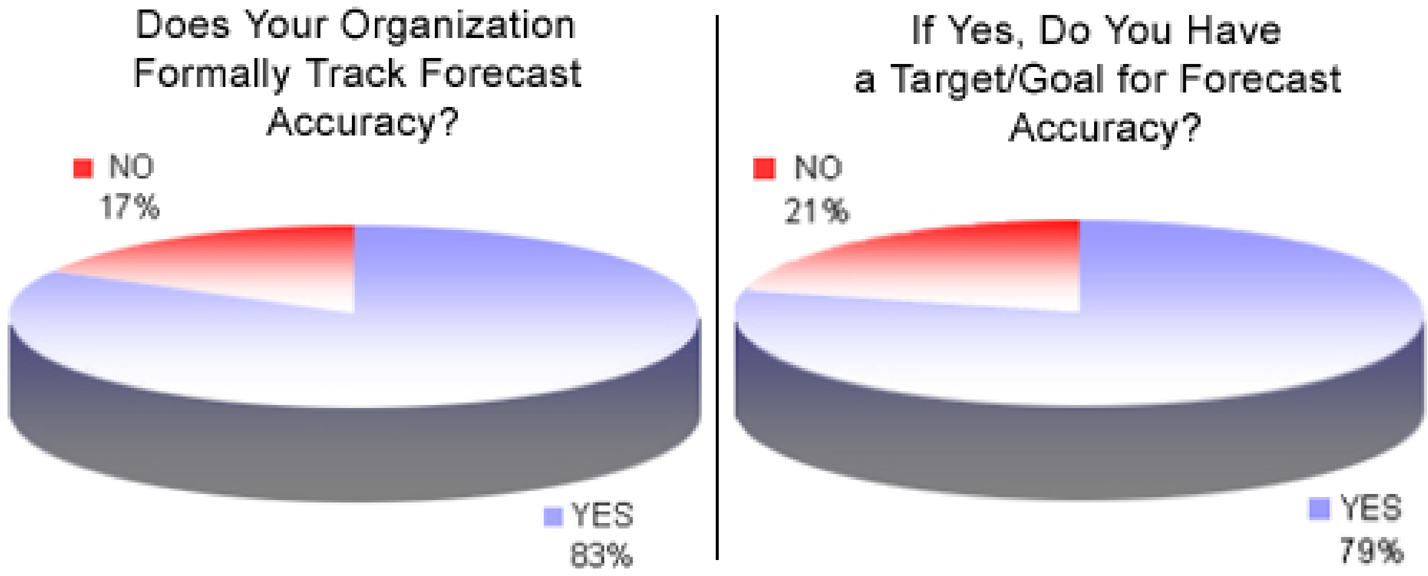
- Accuracy for activities with only Yes/No outcomes
- Forecast calibration: bias in transportation forecasts
- Measures of forecast performance
- Model choice

Some of the Issues

- Does your organization formally track forecast accuracy?
- If yes, do you have a target/goal for forecast accuracy?
- If you have a target/goal for forecast accuracy, how is it set?
- What accuracy measure do you use?

GEORGETOWN UNIVERSITY McDONOUGH School of Business

Source: Forecast Pro Newsletter, August 2009



Will it Rain?

- It rains in the DC area about 1 day in 4.
- Accuracy Criterion: Maximize the percent correct
- What should be the forecast?
- Answer: never predict rain
  - Produces 75% correct answers
  - Any other forecast produces a lower percentage correct

GEORGETOWN
McDONOUGH
School of Business
UNIVERSITY

Lack of Proper Calibration

- The criterion is not properly calibrated because it does not encourage an appropriate answer
- Ask: What is the probability of rain?
- Let Y=1 if rain; Y=0 if no rain
- Forecast P(R) = probability of rain

GEORGETOWN
McDONOUGH School of Business
UNIVERSITY

A Solution

- Use Brier's Score Function and seek minimum:

$$S = \sum [y - P(R)]^2$$

- Example true P =0.7:
  - E(S|P=0.7)=0.21
  - E(S|P=0.0)=0.70
  - E(S|P=1.0)=0.30
- Ready extension to multinomial case

GEORGETOWN
McDONOUGH
School of Business
UNIVERSITY

Baseball games played on August 10, 2009							
Predictions in Wall Street Journal, 8/10/2009							
RESULTS							
Visitor	**Home**	**P(V win)**	**P(H win)**	**Actual**		Brier	Brier-0.5
Chicago White Sox	Seattle Mariners	0.56	0.44	H	1	0.194	0.25
Detroit Tigers	Boston Red Sox	0.41	0.59	H	1	0.348	0.25
Oakland Athletics	Baltimore Orioles	0.28	0.72	V	0	0.078	0.25
Toronto Blue Jays	New York Yankees	0.40	0.60	V	0	0.160	0.25
Chicago Cubs	Colorado Rockies	0.35	0.65	H	1	0.423	0.25
Cincinnati Reds	Saint Louis Cardinals	0.33	0.67	H	1	0.449	0.25
Houston Astros	Florida Marlins	0.58	0.42	H	1	0.176	0.25
					Mean=	0.261	0.25

Measures of Bias for Quantitative Variables

- Let Y = Actual, F = Forecast
- BIAS: $$B = \sum (Y - F)$$
- PERCENT BIAS:

$$PB = 100 \sum (Y - F) / Y$$

- COMMON PRACTICE (e.g. Flyvbjerg, 2005)

$$PB = 100 \sum (Y - F) / F$$

GEORGETOWN UNIVERSITY McDONOUGH School of Business

A Few Comments

- Forecasts of future traffic flows for new transportation projects in Europe tend to:
  - Overestimate rail traffic
  - Underestimate road traffic
  - See Flyvbjerg et al., 2006; Welde & Odeck, 2009
- Is the USA any different? Is the forecasting system properly calibrated or do biased forecasts produce extra (funding) benefits?

GEORGETOWN UNIVERSITY McDONOUGH School of Business

A possible solution

- Reference class forecasting: build an historical data set of somewhat similar projects with actual outcomes and calibrate forecasts using a regression model
  - How to choose the reference set?
  - Use actual outcomes, first year or ramp-up effect?

GEORGETOWN
McDONOUGH
School of Business
UNIVERSITY

Kahneman's Story
(from Flyvbjerg et al, 2006)

- Team of academics and teachers working on a curriculum project; each was asked how long the project would take
- Answers ranged from 18 to 30 months
- Team was then asked "Think of a similar past project; how long did it take to complete?"
- Answers ranged from 7 to 10 years
- OUTCOME: Project was completed 8 years later!

GEORGETOWN UNIVERSITY McDONOUGH School of Business

Variability Measures for Quantitative Variables

- Let Y = Actual, F = Forecast; *m* forecasts, either cross-sectional or time-series
- (Forecast) Mean Square Error:

$$MSE = \sum (Y - F)^2 / m$$

- (Forecast) Mean Absolute Error:

$$MAE = \sum | Y - F | / m$$

- These measures are scale-dependent

GEORGETOWN
McDONOUGH
School of Business
UNIVERSITY

Variability Measures for Quantitative Variables, II

- Remove scale dependence by looking at relative errors
- (Forecast) Mean Absolute Error:

$$MAPE = 100\sum |Y - F| / Y$$

- Requires positive data
  - Net profits
  - Rare events

GEORGETOWN UNIVERSITY McDONOUGH School of Business

Variability Measures for Quantitative Variables, III

- For time series data, use the (Forecast) Mean Absolute Scaled Error:

$$MASE = \frac{\sum | Y_t - F_t |}{\sum | Y_t - Y_{t-1} |}$$

- Require MASE < 1 if method is to do better than a random walk (RW)
- For cross-sectional data, replace RW by a suitable “naïve” model
- For particular applications, other functions such as cost may be more appropriate

GEORGETOWN UNIVERSITY McDONOUGH School of Business

Model Choice: Prediction Validation [PVAL]

- Suppose we have *n+m* observations (cross-sectional or time series)
- Develop/estimate models using *n* observations and then compute the accuracy measures using the other *m* observations
- For time series, the hold-out sample must be at the end of the series; for cross-sectional data, cross-validation is possible, holding out multiple sets of *m*, or alternatively "leave-one-out"

GEORGETOWN
McDONOUGH
School of Business
UNIVERSITY

Model Choice: Information Criteria

- The general form of information criteria is:

$$IC = \log(s^2) + Kq(n)/n$$

- Here *K*= # parameters in the model and *q(n)* is a penalty function:
  - AIC (Akaike): $q(n) = 2$
  - BIC (Schwartz): $q(n) = \log(n)$, *etc.*
- Penalty included to avoid over-parametrization
- Asymptotically, AIC minimizes forecast error, BIC selects the correct model with probability approaching 1.

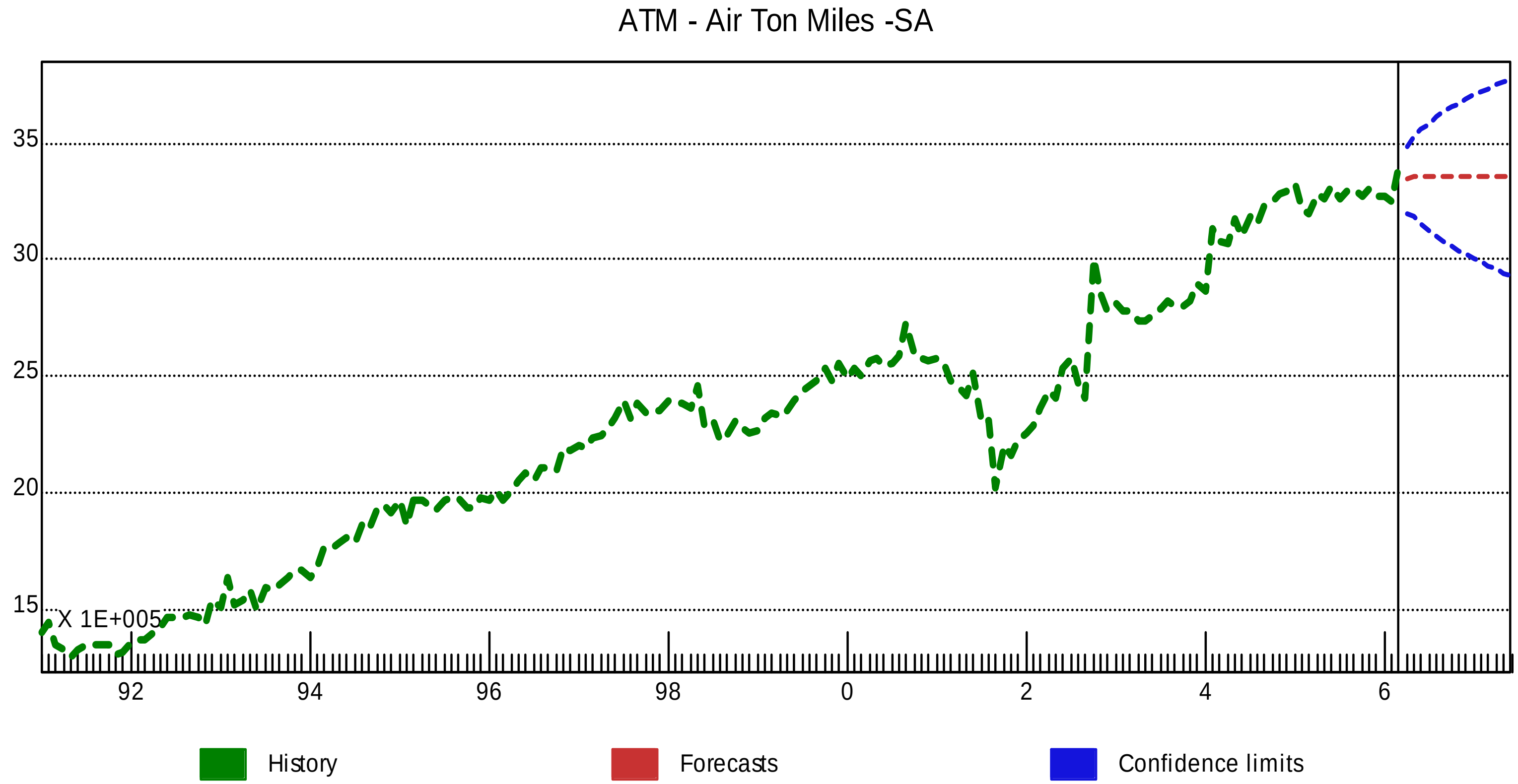
GEORGETOWN UNIVERSITY McDONOUGH School of Business

Model Choice

- AIC tends to work better than BIC for forecasting purposes (still a matter for debate)
- PVAL is widely used in practice, but recent studies have suggested that AIC works better
- For details, see Hyndman et al. (2008, chapter 7), who examine the M3 data and another large data set.

GEORGETOWN UNIVERSITY McDONOUGH School of Business

Air Ton Miles – Seasonally Adjusted



GEORGETOWN UNIVERSITY
McDONOUGH School of Business

Summary Statistics

Analysis performed using Forecast Pro

AIR TON MILES -SA					
Sample size		183			
Mean		2278535			
Std. deviation		575422			
		ARIMA		HOLT	
Adj. R-square		0.98		0.98	
Ljung-Box(18)		24.5	P=0.86	30	P=0.96
BIC		74054		73838	
RMSE		73008		71766	
Durbin-Watson		2.01		2.05	
MAPE		2.07%		2.01%	
MAE		46432		45183	
Out-of-sample MAE		43478		47657	
12 observations, rolling					

GEORGETOWN UNIVERSITY McDONOUGH School of Business

Conclusions

- Choose accuracy measures that reflect both bias and variability
- Accuracy measures should be properly calibrated relative to planning objectives
- Accuracy measures should reflect the appropriate forecasting /planning horizon
- Model choice may be based upon information criteria OR out-of-sample testing: both approaches have their advocates

References

- Flyvbjerg, B. (2005) *Transportation Research, Part A,* 39, 522 – 530
- Flyvbjerg, B., Holm, M.K.S. and Buhl, S.L. (2006) *Transportation Reviews*, 26, 1 -24.
- Hyndman, R.J., Koehler, A.B., Ord, J.K. and Snyder, R.D. (2008) *Forecasting with Exponential Smoothing*. Springer: New York
- Welde, M. and Odeck, J. (2009) *Do planners get it right?* Paper presented at the International Transport Economics Conference at the University of Minnesota in June.

GEORGETOWN McDONOUGH School of Business UNIVERSITY