Four Useful Ways to Measure Forecast Error

Dr. Tom Willemain PhD Co-Founder & Senior Vice President Smart Software, Inc.

Why bother?

1. Forecast error hurts your business: lost profits/excess inventory/unhappy customers.
2. While some error is inevitable, you can't ignore it, and there are ways to reduce it.
3. You can't reduce it if you pretend it's not a problem and neglect to measure it.
4. Measurement uses summary metrics that assess your performance and track it over time.

How to compute forecast error

Actual demand through September

ITEM	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
SKU1	42	54	34	71	68	54	62	34	62			
SKU2	43	59	65	42	57	60	51	40	52			
SKU3	40	51	32	54	52	69	36	60	45			

Forecasts for October-December made using data through September

ITEM	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
SKU1	42	54	34	71	68	54	62	34	62	53	53	53
SKU2	43	59	65	42	57	60	51	40	52	52	52	52
SKU3	40	51	32	54	52	69	36	60	45	49	49	49

Actual demand in October-December

ITEM	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
SKU1	42	54	34	71	68	54	62	34	62	51	48	45
SKU2	43	59	65	42	57	60	51	46	40	60	59	56
SKU3	40	51	32	54	52	69	36	60	45	30	38	41

Four general types of error metrics

1. Scale-dependent error
2. Percentage error
3. Relative error
4. Scale-free error

Remark: Scale-dependent metrics are expressed in the units of the forecasted variable. The other three are expresses as percentages.

1. Scale-dependent error metrics

- Mean Absolute Error (MAE) aka Mean Absolute Deviation (MAD)
- Median Absolute Error (MdAE)
- Root Mean Square Error (RMSE)

- These metrics express the error in the original units of the data.
  - Ex: units, cases, barrels, kilograms, dollars, liters, etc.
- Since forecasts can be too high or too low, the signs of the errors will be either positive or negative, allowing for unwanted cancellations.
  - Ex: You don't want errors of +50 and -50 to cancel and show "no error".
- To deal with the cancellation problem, these metrics take away negative signs by either squaring or using absolute value.

Calculation of scale-dependent error metrics

Time	Actual	Forecast	Error	Absolute Error	Squared Error
Tomorrow	50	48	2	2	4
Next Day	45	62	-17	17	289
Day After	37	28	8	8	64
			MAE	9	
			MdAE	8	
			RMSE		10.9

2. Percentage error metrics

- Mean Absolute Percentage Error (MAPE)

- This metric expresses the size of the error as a percentage of the actual value of the forecasted variable.
- The advantage of this approach is that it immediately makes clear whether the error is a big deal or not.
- Ex: Suppose the MAE is 100 units. Is a typical error of 100 units horrible? ok? great?
- The answer depends on the size of the variable being forecasted. If the actual value is 100, then a MAE = 100 is as big as the thing being forecasted. But if the actual value is 10,000, then a MAE = 100 shows great accuracy, since the MAPE is only 1% of the actual.

Calculation of percentage error metric MAPE

Time	Actual	Forecast	Error	Absolute Error	Squared Error	Absolute % Error
Tomorrow	50	48	2	2	4	4%
Next Day	45	62	-17	17	289	38%
Day After	37	28	8	8	64	22%
			MAE	9		
			MdAE	8		
			RMSE		10.9	
			MAPE			21%

3. Relative error metric

- Median Relative Absolute Error (MdRAE)

- Relative to what? To a benchmark forecast.
- What benchmark? Usually, the “naïve” forecast.
- What is the naïve forecast? Next forecast value = last actual value.
- Why use the naïve forecast? Because if you can’t beat that, you are in tough shape.

Calculation of relative error metric

Time	Actual	Forecast	Error	Absolute Error	Squared Error	Absolute % Error	Naïve Forecast*	Absolute Error	Relative Error
Tomorrow	50	48	2	2	4	4%	43	7	29%
Next Day	45	62	-17	17	289	38%	50	5	340%
Day After	37	28	8	8	64	22%	45	8	100%
							*today=43		
			MAE	9					
			MdAE	8					
			RMSE		10.9				
			MAPE			21%			
			MdRAE						100%

4. Scale-free error metric

- Median Relative Scaled Error (MdRSE)

- This metric expresses the absolute forecast error as a percentage of the natural level of randomness (volatility) in the data.
- The volatility is measured by the average size of the change in the forecasted variable from one time period to the next.
  - (This is the same as the error made by the naïve forecast.)
- How does this metric differ from the MdRAE above?
  - They do both use the naïve forecast, but this metric uses errors in forecasting the demand history, while the MdRAE uses errors in forecasting future values.
  - This matters because there are usually many more history values than there are forecasts.
  - In turn, that matters because this metric would “blow up” if all the data were zero, which is less likely when using the demand history.

Calculation of relative scaled-error metric

Time	Actual	Forecast	Error	Absolute Error	Squared Error	Absolute % Error	Naïve Forecast*	Absolute Error	Relative Error	Average Change	Relative Error
Tomorrow	50	48	2	2	4	4%	43	7	350%	6	33%
Next Day	45	62	-17	17	289	38%	50	5	29%	6	283%
Day After	37	28	8	8	64	22%	45	8	100%	6	133%
							*today=43			*assume these	
			MAE	9							
			MdAE	8							
			RMSE		10.9						
			MAPE			21%					
			MdRAE						100%		
			MdSAE								133%

The special problem of intermittent demand

	January	February	March	April	May	June	July	August	September
Unit demand	0	3	1	0	0	5	0	2	2

- “Intermittent” demand has many zero demands mixed in with random non-zero demands.
- MAPE gets ruined when errors are divided by zero.
- MdRAE can also get ruined.
- MdSAE is less likely to get ruined.

Recap and remarks

- Forecast metrics are necessary aids for monitoring and improving forecast accuracy.
- There are two major classes of metrics: absolute and relative.
- Absolute measures (MAE, MdAE, RMSE) are natural choices when assessing forecasts of one item.
- Relative measures (MAPE, MdRAE, MdSAE) are useful when comparing accuracy across items or between alternative forecasts of the same item or assessing accuracy relative to the natural variability of an item.
- Intermittent demand presents divide-by-zero problems which favor MdSAE over MAPE.
- When assessing forecasts of multiple items, it often makes sense to use weighted averages, weighting items differently by volume or revenue.

Useful reference

- R. J. Hyndman. “Another look at forecast-accuracy metrics for intermittent demand.” *Foresight*, June 2006, pp. 43-46.

- Note: *Foresight* is the practitioner-oriented publication of the International Institute of Forecasters (IIF). You might consider joining IIF and subscribing to *Foresight* as a good way to keep up with advances in forecasting: *https://forecasters.org*.