

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the author's institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/authorsrights>

Contents lists available at [ScienceDirect](#)

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

The forecastability quotient reconsidered

Everette Shaw Gardner Jr.^{a,*}, Yavuz Acar^b^a University of Houston - Bauer College of Business, 334 Melcher Hall, Houston, TX 77204-6021, United States^b Bogazici University - Department of Management, Istanbul, Turkey

ARTICLE INFO

Keywords:

Error measures
Evaluating forecasts
Exponential smoothing
Inventory forecasting
Time series

ABSTRACT

Using a large sample of time series, Hill et al. (2015) developed a procedure that aims to predict whether a series is “forecastable”; that is, whether the standard deviation of the time series will later prove to be larger than that of the forecast errors. Their analysis is based on forecasting using Holt’s method of exponential smoothing. We show that Holt’s method is the wrong one to use for their time series, and we present a number of other corrections and objections to their analysis.

© 2016 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

Hill, Zhang, and Burch (2015) proposed a measure called a “forecastability quotient” that estimates “...the economic advantage of using a time phased reorder point system (TPOP) with time series forecasting rather than a simple reorder point system in an independent demand inventory management context”. In plain language, their version of the TPOP uses the forecasts from a time series method to determine when to reorder stock, while their simple reorder point system uses a moving average forecast (with the length or type of moving average unspecified).

Hill et al. define the forecastability quotient (Q) as the standard deviation of the time series divided by the standard deviation of the forecast errors. If this ratio is more than one, the authors conclude that forecasting should be done using a time series method. If the ratio is less than one, the authors reject all time series methods and conclude that forecasting should be done using a moving average. Although research has long since established that a method’s ex ante fit often has little to do with its ex post

accuracy, Hill et al. claim that Q in the holdout sample can be predicted as a function of Q in the fitting sample, together with the parameters in Holt’s additive trend method of exponential smoothing (Holt, 2004). We have a number of corrections and objections to Hill et al.’s analysis and conclusions, but our most important point is that Holt’s method is the wrong one for the data.

2. Definition of forecastability

One of the motivations for Hill et al.’s paper appears to be their misunderstanding of Granger and Newbold’s (1976) work on the forecasting of transformed time series. Hill et al. state that Granger and Newbold define forecastability as the variance of demand divided by the variance of the forecast error. This is wrong. Granger and Newbold actually define it as the variance of the optimal forecast divided by the unconditional variance of the time series.

3. Data

Hill et al.’s original data set contained 1811 time series, each with 24 months of data (available from <http://dx.doi.org/10.1016/j.ijforecast.2014.10.006>). They state that they removed all series with zero demand in 10% or more

* Corresponding author.

E-mail address: egardner@uh.edu (E.S. Gardner Jr.).

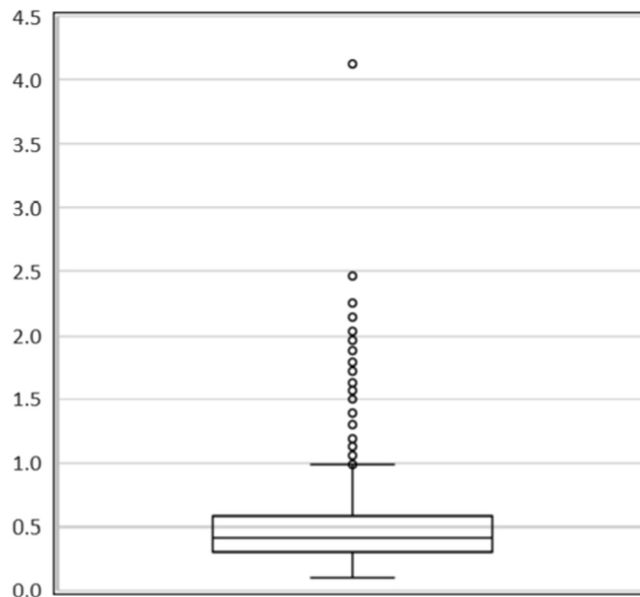


Fig. 1. Box plot of the coefficients of variation (computed over all 24 periods) for the corrected sample of 1629 time series.

of the observations (presumably all series with at least three months of zero demand), leaving 1672 series for analysis. However, we reviewed these 1672 series and found 43 additional series with at least three months of zero demand, and therefore removed these too, leaving 1629 series for re-examination.

The authors used 12 periods as a fit or estimation sample and 12 periods as a holdout sample. They claim that they deseasonalized the time series, but there is not enough data to do so unless the fit and holdout samples are combined, and even then the results would be dubious. We could see no evidence of seasonal patterns in these series, so we ignored seasonal adjustment in our re-examination of the forecasting results.

Hill et al. reported the average coefficient of variation (CV) for the fit sample only, which was 0.469, but this value is based on all 1672 series, including the 43 with at least three months of zero demand. We recomputed the CV for the fit sample after excluding the 43 series, and obtained a significantly smaller average of 0.453. For all 24 periods, we obtained an average CV of 0.479. A box plot of the CVs for all 24 periods is shown in Fig. 1. The distribution is skewed right, with numerous outliers and level shifts in about 200 of the series.

4. Choice of forecasting method

In the abstract of their paper, Hill et al. claimed that they forecasted the time series using “double exponential smoothing with a damped trend”. However, this is not what they actually did. Since Brown (1963), double exponential smoothing (DES) has been defined in the literature as a single-parameter additive trend method. In fact, Hill et al. actually used Holt’s two-parameter additive trend method, and they did not damp the trend, despite what the abstract says. We note that the parameters of the Holt method can be constrained to make it equivalent

to DES (Gardner, 1985), but it does not appear that the authors imposed such constraints.

Hill et al. justify their choice of the Holt method by stating that: “In an extensive empirical study, Makridakis and Hibon (2000) showed that DES is as good as or better than any of the other forecasting methods”. This statement is false. A number of other forecasting methods were more accurate than Holt’s in the M3 competition, both overall and at each individual forecast horizon. For example, the damped trend method of exponential smoothing was more accurate than Holt’s method, a finding that is consistent with a large body of empirical research (Gardner, 2006). It follows that there is no reason to use the Holt method to provide forecasts for any inventory system. If the number of parameters in the forecasting method is a concern, it is possible to use discounted least squares to formulate a two-parameter damped additive trend method, as in Gardner (1985).

We also question Hill et al.’s recommendation to use a moving average as an alternative to a time series method. While Hill et al. do not give empirical results for a moving average, nor do they specify the length or type of average, we assume that they refer to a simple moving average. For a variety of reasons, it is generally accepted that simple exponential smoothing (SES) should be more accurate than an unweighted moving average of any kind. For empirical evidence, see the comparisons of SES and simple moving averages in the M-competition (Makridakis et al., 1982), which appears to be the last time that moving averages were used in forecasting research.

5. Empirical results

Hill et al. presented extensive model-fitting results for the Holt method, but they gave only the Q value for the holdout sample. We attempted to replicate their results and also tested the damped trend, SES, and the

Table 1

Forecast accuracy comparisons for the corrected sample of 1629 series.

Sample	Periods	Metric	Hill et al.	Holt	Damped	SES	Naive
Fit	1–12	Q	0.87	0.97	1.05	1.01	0.71
		MAPE	0.39	0.36	0.34	0.36	0.44
		MASE	0.90	0.75	0.68	0.72	
Holdout	13–24	Q	0.79	0.84	0.97	0.97	0.70
		MAPE		0.45	0.42	0.38	0.44
		MASE		1.07	0.96	0.84	

Notes: Q is the ratio of the standard deviation of the time series to that of the forecast errors, so larger values of Q are better. The Hill et al. results are for their version of the Holt method. Hill et al. did not present MAPE or MASE results for the holdout sample.

naïve method. Comparisons of Q values, mean absolute percentage errors (MAPE), and mean absolute scaled errors (MASE) are given in Table 1. The MASE scales the errors by the fitted mean absolute error from the naïve (random walk) forecasting method (Hyndman & Koehler, 2006). The Hill et al. values are for their version of the Holt method. Note that higher Q values are better. Following Hill et al., APEs that were greater than one were reset to one.

Our Holt results are significantly better than those of Hill et al. Part of the difference may be due to the fitting procedure. We used the Excel Solver to find parameters, but Hill et al. used a grid search, even though their calculations were done in Excel. One of the referees for this paper noted that the objective function for Holt's method is bumpy, and that a grid search can sometimes give better results than a nonlinear optimization. We agree, and consider that it makes sense to test both approaches in case of doubt.

Hill et al. call their parameters optimal, but we doubt that this is the case. Their average level and trend parameters were 0.377 and 0.508. Although these are invertible parameters, our experience is that it is highly unusual to obtain a trend parameter that is this much larger than the level parameter.

As expected, the damped trend was more accurate than Holt in every comparison. We also expected the damped trend to be more accurate than SES, as it was in the M-competitions, but SES was clearly the best method. SES produced a MASE of 0.84 in the holdout sample, compared to 0.96 for the damped trend and 1.07 for Holt. In noisy time series such as these, the fitted damping parameter is usually quite small, making the damped trend forecasts about the same as SES. However, the fitting sample is only 12 periods, which is much too short to allow the damped trend method to adapt to the data.

6. Predicting Q in the holdout sample

Hill et al. present an equation for predicting Q in the holdout sample as a function of Q in the fitting sample and the level and trend parameters of the Holt method. The authors argue that these relationships in their equation are intuitive, but we do not see how parameter values can be used to predict the performance of a forecasting method. However, there is no point in discussing this equation further here, because Holt's method is the wrong one for the time series.

7. Conclusions

Contrary to Hill et al., Granger and Newbold did not define forecastability as the ratio of the demand variance to the forecast error variance, the Holt method is not double exponential smoothing, the Holt method was not the most accurate method in the M3 Competition, and the best forecasting method by a wide margin for these time series is SES, not the Holt method.

We question the need for a forecastability quotient in the first place. The correct way to deal with a time series that is difficult to forecast is to compare the accuracy to a naïve benchmark using the MASE, which is easy to interpret. Values of the MASE greater than one indicate that the forecasts are worse, on average, than fitted one-step-ahead forecasts from the naïve method, and this idea is easy to extend to simulated errors in a test or holdout sample. If no forecasting method that can beat the naïve method can be found, then the naïve method is the best choice.

Fildes and Kingsman (2011) demonstrated that the selection of a forecasting method is a critical determinant of inventory costs, but the question of how to select the best method for inventory forecasting is far from settled. For further discussion, see Gardner (2006), Snyder, Koehler, and Ord (2002), Syntetos, Boylan, and Disney (2009) and Syntetos, Nikolouopoulos, and Boylan (2010). We believe that the best simple procedure is to select the method that yields the smallest MASE on a test set. When it is possible to build a model of the inventory system, we recommend method selection based on aggregate operational performance measures such as costs, efficiency, or customer service, as discussed in Acar and Gardner (2012) and Gardner (1990).

References

- Acar, Y., & Gardner, E. S. (2012). Forecasting method selection in a global supply chain. *International Journal of Forecasting*, 28, 842–848.
- Brown, R. G. (1963). *Smoothing, forecasting and prediction of discrete time series*. Englewood Cliffs, NJ: Prentice-Hall.
- Fildes, R., & Kingsman, B. (2011). Incorporating demand uncertainty and forecast error in supply chain planning models. *Journal of the Operational Research Society*, 62, 483–500.
- Gardner, E. S. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting*, 4, 1–28.
- Gardner, E. S. (1990). Evaluating forecast performance in an inventory control system. *Management Science*, 36, 490–499.
- Gardner, E. S. (2006). Exponential smoothing: The state of the art—Part II. *International Journal of Forecasting*, 22, 637–666.
- Granger, C. W., & Newbold, P. (1976). Forecasting transformed series. *Journal of the Royal Statistical Society: Series B*, 38, 189–203.

- Hill, A., Zhang, W., & Burch, G. (2015). Forecasting the forecastability quotient for inventory management. *International Journal of Forecasting*, 31, 651–663.
- Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20, 5–10.
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22, 679–688.
- Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, R., & Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, 1, 111–153.
- Makridakis, S., & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*, 16, 451–476.
- Snyder, R. D., Koehler, A. B., & Ord, J. K. (2002). Forecasting for inventory control with exponential smoothing. *International Journal of Forecasting*, 18(1), 5–18.
- Syntetos, A. A., Boylan, J. E., & Disney, S. M. (2009). Forecasting for inventory planning: a 50-year review. *Journal of the Operational Research Society*, 60, S149–S160.
- Syntetos, A. A., Nikolouopoulos, K., & Boylan, J. E. (2010). Judging the judges through accuracy-implication metrics: the case of inventory forecasting. *International Journal of Forecasting*, 26, 134–143.

Everette Shaw Gardner Jr., is Professor and Director of the Honors Program in the Bauer College of Business at the University of Houston, where he teaches Honors colloquia in forecasting and supply chain management. His research interests include exponential smoothing and supply chain modeling.

Yavuz Acar is Assistant Professor in the Department of Management at Bogazici University in Istanbul, where he teaches supply chain management. His research interests include mathematical programming, supply chain modeling, and demand forecasting.