Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 26 (2010) 134–143
international journal of forecasting
www.elsevier.com/locate/ijforecast

Judging the judges through accuracy-implication metrics: The case of inventory forecasting

Aris A. Syntetos[a,*], Konstantinos Nikolopoulos[b,1], John E. Boylan[c,2]

[a] *Centre for Operational Research and Applied Statistics, Salford Business School, University of Salford, Maxwell Building, Manchester M5 4WT, UK*
[b] *Manchester Business School, University of Manchester, Booth Street West, Manchester M15 6PB, UK*
[c] *School of Business and Management, Buckinghamshire New University, Newland Park, Gorelands Lane, Bucks HP8 4AD, UK*

Abstract

A number of research projects have demonstrated that the efficiency of inventory systems does not relate directly to demand forecasting performance, as measured by standard forecasting accuracy measures. When a forecasting method is used as an input to an inventory system, it should therefore always be evaluated with respect to its consequences for stock control through accuracy implications metrics, in addition to its performance on the standard accuracy measures. In this paper we address the issue of judgementally adjusting statistical forecasts for 'fast' demand items, and the implications of such interventions in terms of both forecast accuracy and stock control, with the latter being measured through inventory volumes and service levels achieved. We do so using an empirical dataset from the pharmaceutical industry. Our study allows insights to be gained into the combined forecasting and inventory performance of judgemental estimates. It also aims to advance the practice of forecasting competitions by arguing for the consideration of additional (stock control) metrics when such exercises take place in an inventory context.

Keywords: Sales forecasting; Judgemental forecasting; Adjusting forecasts

1. Introduction

In many organisations, the size and complexity of the demand forecasting task at the individual Stock Keeping Unit (SKU) level necessitates the use of statistical methods, such as exponential smoothing, instead of judgemental methods. However, in many cases these statistical forecasts will be subject to judgemental adjustments by managers (Fildes & Goodwin, 2007; Sanders & Manrodt, 1994). This is true for both 'fast' and 'slow' demand items. Some academic research has been carried out to ascertain the effectiveness of these adjustments with re-

* Corresponding author. Tel.: +44 161 295 5804.
E-mail addresses: a.syntetos@salford.ac.uk (A.A. Syntetos), kostas.nikolopoulos@mbs.ac.uk (K. Nikolopoulos), john.boylan@bucks.ac.uk (J.E. Boylan).
[1] Tel.: +44 161 306 8996.
[2] Tel.: +44 1494 60 5130.

0169-2070/$ - see front matter
doi:10.1016/j.ijforecast.2009.05.016

spect to forecast accuracy (e.g., Fildes, Goodwin, Lawrence, & Nikolopoulos, 2009 and Mathews & Diamantopoulos, 1992). However, the stock control implications of these human judgements have not received sufficient attention. We have recently undertaken a project (Syntetos, Nikolopoulos, Boylan, Fildes, & Goodwin, 2009) where we explore the inventory performance of adjustments when they are applied to statistical forecasts of intermittent demand. In this paper, we extend this analysis to the case of fast demand items by means of experimentation with an empirical dataset. We argue that forecasting performance in an inventory context should always be evaluated with respect to its implications for stock control, through accuracy implications metrics, in addition to considering forecast accuracy measures.

Similar arguments have been advanced in the academic literature in the contexts of different forecasting applications. In the area of economics and finance, for example, Timmermann and Granger (2004) highlighted the need to evaluate forecast results using utility functions. Often the predictive approach that is best based on a given accuracy metric will not be the one that outperforms competitors if utility measures are employed, such as financial outcomes, inventories, customer satisfaction, or socio-economic benefits.

The remainder of our paper is structured as follows: in the next section we discuss the importance of assessing the stock control performance of estimators in an inventory context. In Section 3 we provide details of the empirical dataset available for the purposes of our investigation and of the structure of our simulation experiment. Section 4 presents the empirical results and their interpretation, followed in Section 5 by the conclusions drawn from this work, along with some suggestions for further research in this area.

2. Accuracy implication metrics

Solutions to stock control problems with stochastic demand involve forecasting the level of demand and its variability, among other things. A number of research projects have demonstrated that the efficiency of stock control systems does not relate directly to demand forecasting performance, as measured by

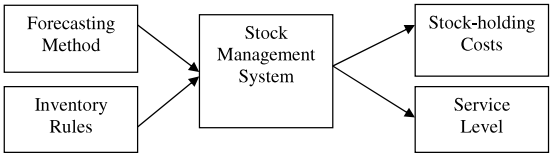

Fig. 1. Inventory system performance measurement.

standard forecasting accuracy measures. That is to say, the fact that estimator x performs better than estimator y in terms of forecasting accuracy (however this is measured) does not necessarily imply that this superiority will be reflected in the stock control implications of using these estimators. In that respect, it has become apparent that forecast accuracy must be distinguished from the empirical utility (stock control performance) of the estimators being used (e.g., Boylan, 2007, Gardner, 1990, Mahmoud & Pegels, 1989 and Syntetos & Boylan, 2008).

Boylan and Syntetos (2006) commented on the importance of capturing the combined forecasting-stock control operation through metrics relating to 'service level' and inventory costs (accuracy implication metrics); i.e., by considering what is important from a practitioner's perspective (please refer to Fig. 1).

Stock-holding costs are relatively straightforward to interpret. They are generally calculated as a percentage of the value of inventory investment, where the percentage takes into account such factors as the cost of capital, insurance, warehousing and obsolescence costs. 'Service level' is generally interpreted as 'off the shelf availability', but the way in which it is measured varies. Two common measures are defined as follows (Silver, Pyke, & Peterson, 1998):

- The fraction of replenishment cycles in which the total demand can be delivered from stock (known as P_1). This is equivalent to a specified probability of no stock-outs during a replenishment cycle, and this criterion is also known as the 'Cycle Service Level'.
- The fraction of the total demand that can be delivered from stock (known as P_2). This is also called the 'fill rate'.

In addition to these standard measures, other suggestions have been made. Gardner and Dannenbring