The effects of integrating management judgement into OUT levels: in or out of context?[1]

A.A. Syntetos[a #], I. Kholidasari[b] and M.M. Naim[a]

[a]Cardiff University, UK; [b]University of Bung Hatta, Indonesia

[#]Corresponding author: +44 (0)29 2087 6572, SyntetosA@cardiff.ac.uk

HIGHLIGHTS

1. It is common practice that statistically derived inventory decisions, such as Order-Up-to (OUT) levels, are judgementally adjusted;
2. To our knowledge the effect of such adjustments has not been studied empirically and this is the first endeavour to do so;
3. A database related to 1,800 SKUs from an electronics industry company is analysed;
4. The linkage between adjustments and their justification is examined;
5. Useful insights are offered to academics, practitioners and software developers.

ABSTRACT

Physical inventories constitute a significant proportion of companies' investments in today's competitive environment. The trade-off between customer service levels and inventory reserves is addressed in practice by statistical inventory software solutions; given the tremendous number of Stock Keeping Units (SKUs) that contemporary organizations deal with, such solutions are fully automated. However, empirical evidence suggests that managers habitually judgementally adjust the output of such solutions, such as replenishment orders or re-order levels. This research is concerned with the value being added, or not, when statistically derived inventory related decisions (Order-Up-To, OUT, levels in particular) are judgementally adjusted. We aim at developing our current understanding on the effects of incorporating human judgement into inventory decisions; to our knowledge such effects do not appear to have been studied empirically before and this is the first endeavour to do so. A number of research questions are examined and a simulation experiment is performed, using an extended database of approximately 1,800 SKUs from the electronics industry, in order to evaluate human judgement effects. The linkage between adjustments and their justification is also evaluated; given the apparent lack of comprehensive empirical evidence in this area, including the field of demand forecasting, this is a contribution in its own right. Insights are offered to academics, to facilitate further research in this area, practitioners, to enable more constructive intervention into statistical inventory solutions, and software developers, to consider the interface with human decision makers.

Keywords: Judgemental adjustments; Inventory management; Behavioural operations.

---

[1] All the *Appendices* of this paper are presented as supplementary material in an Electronic Companion.

1. RESEARCH MOTIVATION AND CONTRIBUTION

Physical inventories are potentially a major element of a company's investments. Companies strive to ensure high customer satisfaction, and off-the-shelf availability is almost a necessity under very many current supply chain arrangements. Commonly, forecasting and stock control are the most important elements of an inventory management system. Given the tremendous number of Stock Keeping Units (SKUs) that contemporary organizations deal with, such systems need to be fully automated. However, although such systems are indeed fully automated, what most often happens in practice is that managers intervene into the system and superimpose their own judgement on statistically derived quantities (Sanders and Manrodt, 1994; 2003; Nikolopoulos et al., 2005). That is, they judgementally adjust statistical forecasts and / or inventory quantities such as re-order points, Order-Up-To (OUT) levels, replenishment quantities, safety stocks, and so on, depending on the inventory models being employed to manage stock.

There is a growing knowledge base in the area of judgementally adjusting statistical forecasts in an inventory / supply chain context (Fildes et al., 2009; Syntetos et al., 2009). However, there has been very limited discussion on the effects of judgementally adjusting inventory related decisions, although it is common knowledge that such adjustments routinely take place in industry (Kolassa et al., 2008, Syntetos et al., 2011).

This research is concerned with the value being added, or not, when statistically derived inventory related decisions are judgementally adjusted. In particular, we look at the effects of adjusting OUT levels through an evaluation of the resulting trade-offs between inventory investments and service levels achieved. We aim at developing our understanding on the effects of incorporating human judgement into inventory control and to our knowledge this is the first study that addresses this issue empirically using a comprehensive data set made available by a company from the electronics sector. Our research relates explicitly to the areas of Operational Research (OR) and Operations Management (OM) in conjunction with behavioural aspects of decision-making, and as such it contributes to the advancement of knowledge in the field of Behavioural Operations (BO) (Hamalainen et al., 2013).

Based on the literature, we develop a number of research questions, which are examined through the analysis of an extended database of approximately 1,800 SKUs from the electronics industry. The case organization operates a periodic OUT policy and a simulation experiment is constructed and run to evaluate in a dynamic fashion the effects of judgemental adjustments on the OUT levels: that is, we compare what has actually happened to what would have happened if OUT levels had not been judgementally adjusted. The research questions relate to the possible explanatory power of the adjustments' characteristics (magnitude and direction) on the performance of the adjustments as judged by the resulting trade-offs between inventory costs and service levels achieved. With regards to that latter, we consider Cycles Service Levels (CSL) and Fill Rates (FR).

In addition, we examine the linkage between adjustments and their justification, provided in the form of text accompanying some of the adjustments made by the human decision makers. Linking the effects of adjustments to the rationale behind their proposition does not appear to have been studied before empirically in the wider area of inventory management, including the field of demand forecasting, and this part of our research is in itself a contribution.

The remainder of our paper is organized as follows: in the next section we provide the research background followed, in Section 3, by the development of our research questions. In Section 4 we provide details related to the case study organization, the database available for the purposes of our research and the process of judgemental intervention in the company under concern. In Section 5, we first perform statistical goodness-of-fit tests on the appropriateness of a number of theoretical distributions to characterize the adjustments, followed by a descriptive analysis of the justifications accompanying the adjustments. In Section 6, we address the research questions developed in our paper through a detailed simulation experiment. We conclude in Section 7 with a summary of our findings and the limitations of our work, as well as offering insights to academics, practitioners and software developers.

2. RESEARCH BACKGROUND

Because of the tremendous number of SKUs that both manufacturing and service organizations deal with, the inventory task needs to be automated. This is particularly pertinent in the electronics sector, which has highly complex products with an extremely high number of assembly configurations, sub-assemblies and components. Automation here implies fully quantitative models that can run on their own without human intervention, thus relying upon statistical generalisable principles. Such models exploit past information that is available to the system and thus may not capture contextual knowledge that managers possess. For example, decision makers may know that organisations are in the process of change, or that a product promotion is about to take place, that certain actions are being undertaken by competitors that will affect demand for a particular product, or that a manufacturing problem exists. Similarly, a variable that is difficult to measure may be missing from the model. For example, obsolescence, that is a crucial issue in the after-sales industry dealing with service parts (but also in inventory management in general given the current short life cycles of products), is rarely explicitly accounted for in stock control models.

It is for these reasons that managers adjust the output of automated systems, by altering some quantities, with the aim to yield desirable outcomes. However, it is known that managers may often superimpose their own judgement to statistically derived quantities without any *good* reasons for doing so (Olsmats et al., 1988). In such cases the adjustments may not account for any benefit, but the very process still implies that valuable and expensive managerial time has been unnecessarily spent, or they may even deteriorate performance, however performance is being measured (Sanders, 1992). In the closely related field of demand forecasting, performance would be measured through forecast accuracy (Fildes et al., 2009) and accuracy implication / inventory metrics (Syntetos et al., 2009, 2010). In demand forecasting, judgemental interventions may be often attributed to one of the following reasons:

- A mere desire for a sense of ownership of a system that is perceived as 'black box' by the operating manager (Goodwin, 2002; Onkal and Gonul, 2005). Lack of training in the field of forecasting or utilization of forecast software packages, or in house developed solutions, that are not particularly transparent (for example, in terms of the methods being used, parameter selection and optimization, etc) results in managers intervening into the system so that they perceive they have a better understanding of how the forecasts are derived.

- The misinterpretation of ‘noise’ as a systematic pattern (Harvey, 1995; Goodwin, 2005). The human tendency to ‘see’ patterns where there are not necessarily any patterns at all is reflected to, typically, small adjustments that do not account for any benefits or, even worse, they may also damage forecast accuracy.

- Political pressures to favour a particular outcome as opposed to ‘optimality’ (Goodwin, 1996). There are many scenarios where this might occur. In the sales functions, for example, where performance is measured and rewarded based on service levels achieved, managers will typically inflate the forecasts in an attempt to avoid running out of stock. Similarly, and under the pressures of limited supply for a particular material / product, forecasts may be inflated in order to ensure some priority from the suppliers.

A comprehensive knowledge base is emerging in the area of judgementally adjusting demand forecasts with important implications for the design and development of Forecast Support Systems (FSS) (Franses and Legerstee, 2011, 2013). Previous empirical studies have looked, as discussed above, both at the effects of judgemental adjustments on forecast accuracy and their inventory implications. In an inventory management context, demand forecasts are inputs to a stock control model, which achieves a certain customer service level with a particular investment in inventories. As the interactions between forecasting and stock control do not guarantee that a superior forecast accuracy performance is reflected on a better trade-off between service and inventory cost, implication metrics capture the true utility of the forecasts.

Kolassa et al. (2008) suggested that replenishment orders (and inventory related decisions in general) are more often judgementally adjusted than statistical forecasts. In such a context, Disney et al. (1997) advocated that, in practical settings, human intervention often makes things worse, which is supported by behavioural studies through the medium of the well-known ‘Beer Game’ (Disney et al., 2004). But Disney et al. (2004) did note that intervention makes things worse with ‘routine’ tasks but potentially can have benefits, if appropriate and simple protocols are put in place, for ‘exceptions’.

Further, statistical forecasts and inventory decisions may be actually both judgementally adjusted. Syntetos et al. (2011) analysed, by means of employing a System Dynamics (SD) methodology, the effects of judgement on supply chain behaviour. They highlighted two behavioural factors: i) judgemental changes to forecasts and ii) judgemental changes to orders. The second may be distinct from the first, as it is not necessarily based on any expectation of a changed demand pattern, but rather on a subjective reaction to external stimuli, e.g., changes in price or shortage of supply. However, the former may indeed affect the latter if a single person, for example, performs both adjustments. Similarly, in the context of a small organization, the judgemental adjustments of the orders may reflect a certain reaction mechanism on the part of the stock controller to known adjustment behaviours on the forecasting side. The distinction between forecasting and ordering adjustments has been neglected in the academic literature, but is very important in practice (Kolassa et al., 2008).

Currently, there are two main empirical gaps in the area of judgement in inventory management: i) studying the effects of judgementally adjusting inventory related decisions; ii) studying the effects of jointly adjusting forecasts and inventory related decisions. The first gap is addressed in this paper. The second gap is an important avenue for further research.

3. RESEARCH QUESTIONS

In this section we look in some more detail at previous work in the area of judgementally adjusting statistical forecasts in order to inform the development of a number of research questions that would be worthwhile pursuing in an empirical context in the area of stock control. Syntetos et al. (2009) studied the forecasting and inventory effects of integrating management judgement into intermittent demand forecasts. A number of research questions were first developed followed by simulation on empirical data. The research design employed in our paper follows closely that suggested and used by Syntetos et al. (2009).

Many studies have concluded that managerial interventions in the process of statistical forecasting improve forecast accuracy (Lawrence et al., 1986; Mathews and Diamantopoulos, 1986, 1990, 1992; Wolfe and Flores, 1990, Syntetos et al., 2010). Goodwin (2000a) suggested that the use of judgemental adjustments to statistical forecasts is justified when non-time series information has predictive power and this information is difficult to capture in a statistical model. This finding is supported by Goodwin (2005) and Sanders and Ritzman (2001) who argued that judgement can be valuable when the forecasters have important information about forthcoming events that is difficult to capture in a statistical model. One would expect that the benefits of judgemental adjustments that have been reported in the forecasting literature should also apply in terms of inventory decisions and OUT levels in particular, which are relevant for the case organization. So it is natural that the first important question that this research attempts to answer relates to any potential improvements resulting from judgementally adjusting OUT levels. The first research question is:

Q1. Is there any improvement in judgementally adjusting OUT levels and, if so, why?

Forecasting related research suggests that the size and sign of adjustments have some explanatory power in terms of performance. Fildes et al. (2009) found that large adjustments are more effective in improving forecast accuracy than small adjustments. As discussed in the previous section, small adjustments typically constitute reaction to 'noise'. In particular, large negative adjustments are most effective. The literature suggests the existence of bias towards making overly positive adjustments as a consequence of non-symmetric loss functions of the managers (Franses and Legerstee, 2009). Given business pressures to achieve performance via high customer service levels, or ensure priority from the suppliers, negative adjustments reflect some genuinely missing information from the statistical model.

Syntetos et al. (2009) confirmed the above findings when empirically assessing the inventory implications rather than the resulting forecast accuracy of judgemental adjustments. It is reasonable to expect that the sign (direction) and size (magnitude) of OUT level adjustments should relate to the inventory performance of such adjustments. Accordingly, the second research question of this study is:

Q2. How does the sign and size of the OUT level adjustments affect the performance of inventory systems?

Documentation of reasons behind a particular forecasting model being chosen and why adjustments of forecasts are made is important in reducing bias in relevant processes (Goodwin, 2000a). Such documented reasons could be used in determining why a forecast is potentially erroneous since the rationale for decision making on the part of the forecaster is recorded and can be evaluated. Further, it has been argued (through laboratory experiments, Goodwin, 2000b) that when users have to provide reasons for judgemental interventions, they make fewer unnecessary judgemental adjustments without being deterred from making these adjustments when they are appropriate. Despite the obvious importance of the provision of a justification, to our knowledge neither the nature nor the effects of such justifications have ever been studied empirically in the stock control literature. Hence, our third research question is:

> **Q3. Is the performance resulting from OUT level adjustments that are accompanied by a justification any different to the performance of those that are not?**

4. CASE STUDY ORGANISATION

The company that provided the database used for the purposes of this research represents the European logistics operations of a major international electronics manufacturer. The entire database relates to service parts used for supporting the final pieces of equipment, such as printers, sold in Europe. Whereas physical flows of service parts are initiated in Germany, information is controlled from the United Kingdom (UK). The organisation has manufacturing plants in Japan, Taiwan, Malaysia, China, and one small manufacturing facility in the UK.

4.1. ERP system

The organisation has implemented the Enterprise Resource Planning (ERP) software package, SAP R/3 (SAP-AG, Germany); the materials management (MM) module of such software is essentially being used to control inventories. Demand that triggers the orders can be expressed as actual orders or demand forecasts. Users are required to specify demand categories and control stock quantities periodically with the review period(s) being set manually. Decisions on replenishments are being made in terms of a *min-max* system, equivalent to the re-order point *s*, order-up-to level *S* (*s, S*) policy or versions of it. For example, in the case organisation the *S* only is required. Other stock control procedures may be implemented as well, albeit with manual specifications and inputs. The safety stock determination in SAP is also limited, since for example no 'fill rate' objectives can be defined. The software also contains forecasting functionality. Although many time-series forecasting methods are incorporated, such as moving averages as well as simple or more elaborate exponential smoothing techniques, many problems arise when dealing with spare parts since demand for such items is usually intermittent in nature, requiring different forecasting methods (such as Croston's estimator) specifically developed for such patterns (Teunter et al., 2011). SAP R/3 does not contain Croston's estimator, which is included though in SAP APO, Advanced Planner and Optimizer.

4.2. Empirical data

The database available for the purposes of this research consists of the individual demand histories of various SKUs. The demand histories have been made available to us in a weekly time-series format covering the period April 2009 to May 2011. The company classifies SKUs based on an ABC-type classification that takes into account the frequency of demand occurrence (how often demand occurs) and the 'value' of the items (price). A weighting procedure is applied to take into account these factors. In this research A and B items are only considered as C items are managed outside the system through a manual process. There are 359 A-class SKUs and 1,454 B-class SKUs, associated with a total of 1,461 and 2,958 OUT level adjustments respectively. In *Appendix A*, in the Electronic Companion of this paper, we present a descriptive analysis of the demand series for the A and B-items (Tables A1 and A2 respectively).

4.3. Inventory management

At the end of every month, a 24 week Simple Moving Average [SMA(24)] forecast is produced. This forecast is used to compute the safety stocks for every SKU by multiplying it by a safety target, expressed in terms of time requirements. This safety target equates to 8 weeks availability for A items and 12 weeks for B items. Following this, the order frequency and the lead times are also taken into account in order to calculate the order-up-to (OUT) level for every SKU. Lead times are assumed to be fixed and equal to nine weeks.

Inventory control takes place through a periodic OUT level system, which in the company is, erroneously, referred to as a re-order point (ROP) system. The OUT replenishment level is calculated at the end of every month by multiplying the SMA(24) forecast by 19 (8 weeks safety stock + 9 weeks lead time + 2 weeks order frequency adjustment) in the case of A items, and 23 (12 weeks target safety stock + 9 weeks lead time + 2 weeks order frequency adjustment) in the case of B items. The periodic nature of the system is reflected in the order frequency adjustments of two weeks. The target safety stock and order frequency for both the A and B items has been decided in an arbitrary way, and there is no explanation as to why the managers opted for those values. The inventory system described above is presented in *Appendix B* in the supplementary material accompanying this paper (Figures B1 and B2 for the A and B items respectively).

4.4. Judgemental adjustments

Initially, the OUT level is produced by the SAP system (hereafter termed as the *System OUT replenishment level*) and, when managers believe it is necessary, they may alter it by using their own judgement. This is an individual rather than group decision. There are 4 people involved into the process and all 4 of them work in the Logistics function of the company being, generally, responsible for managing inventory flows. We have no further information either on the background of those individuals or on their specific role within the Logistics function of the company. The rationale behind the adjustments should ideally be documented as contextual information that is not available via the quantitative data; this would enable senior management to have access to the motivations behind such interventions. When making the adjustments, another OUT level (this will be referred to as the *SMA-Based OUT replenishment level*) is taken into account which is the one calculated based on the descriptions provided in the previous sub-section. So, essentially managers make adjustments to the *System OUT replenishment level,* i.e. to the SAP 'black box' formula, by taking into account the *SMA-Based OUT replenishment level*. Decision makers often use statistical methods as support tools towards reaching a decision (Onkal et al., 2009; Leitner and Leopold-Wildburger, 2011). The adjusted OUT level is the manager's final decision for the end of the current month and will be used to drive replenishment decisions in the following month. This will be referred to as the *Final OUT replenishment level*. On the other hand, if managers do not make any changes to the order replenishment level, the *System OUT replenishment level* is recorded as the final decision for the current month and it constitutes the initial OUT level for the next period. The process described above is schematically presented in Figure 1.

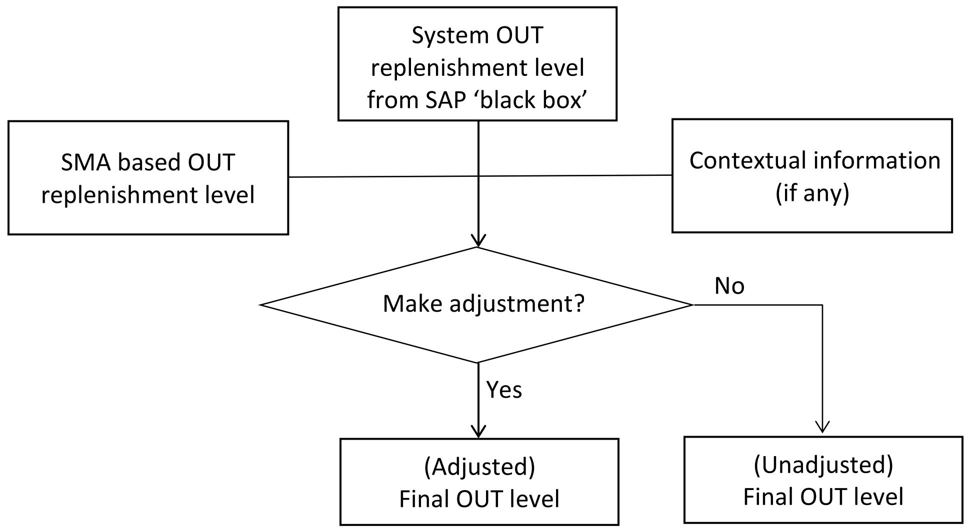

Figure 1. The process of adjustments to the OUT level

5. DESCRIPTIVE ANALYSIS

5.1. Goodness of fit tests and distributional considerations

The judgemental adjustments are first analysed in terms of the distribution of their signed size, absolute size, relative signed size and relative absolute size (relative expressions are with regards to the System OUT replenishment level), to capture collectively the characteristics of both magnitude and direction in absolute and relative terms. The goodness-of-fit of various plausible theoretical statistical distributions is analysed using the Kolmogorov-Smirnov (K-S) test. We consider distributions with no more than two parameters. The maximum number of parameters is limited to two to reflect a trade-off between goodness-of-fit and computational requirements for practical applications. The goodness-of-fit tests are conducted by deploying the *EasyFit* software package. There are nine theoretical distributions that have been considered, for both the absolute and signed cases analysed (Cauchy, Exponential, Gamma, Gumbel max & min, Logistic, Normal, Uniform, Weibull). With regards to the former, we did not restrict our analysis to distributions defined only in the positive domain. Further, zero adjustments have obviously not been considered when fitting statistical distributions. The summary of the best fitting distributions can be seen in Table 1.

Table 1. Adjustment distributions for A and B items.

	A items	B items
Signed size of adjustments	• Most adjustments (29.23%) are between -20 – -1 units • Cauchy distribution	• Most adjustments (19.47%) are between 1 – 10 units • Cauchy distribution
Absolute size of adjustments	• Most adjustments (44.49%) are between 1 – 20 units • Gamma distribution	• Most adjustments (39.72%) are between 1 – 10 units • Gamma distribution
Relative signed size of adjustments	• Most adjustments are between -25% – -30% (negative adjustments (11.84%)) and 5% – 15% (positive adjustments (7.94%)) • Cauchy distribution	• Most adjustments are between -30% – -35% (negative adjustments (8.82%)) and 35% – 40% (positive adjustments (6.39%)) • Cauchy distribution
Relative absolute size of adjustments	• Most adjustments (15.95%) are between 20% – 25% • Gamma distribution	• Most adjustments (14.30%) are between 30% – 35% • Cauchy distribution

Knowledge of particular distributions that provide a good fit to the adjustments is extremely useful towards the design of decision support systems (DSS). Since the parameters of the distributions can be calculated based on past data (past adjustments) percentiles may be specified that relate to, for example, authorization points. That is, adjustments greater than *x* amount (expressed either in signed/absolute or relative terms) need to be authorized whereas adjustments below the authorization point may be freely conducted.

5.2. Analysis of the justification of adjustments

First we have qualitatively clustered all the justifications into conceptually uniform categories, resulting in 24 such categories as given in *Appendix C*. The initial letters *D* and *I* stand for the corresponding decision to Decrease or Increase the OUT level.

In Figure 2 we indicate the frequency of the various classes of justifications for the A and B items. Excluding the adjustments associated with no justification (A=38.60%, B=77.55%), the main reason behind performing adjustments is associated with a perceived decrease (A items: 40.93%, B items: 21.47%) or increase (A items: 16.56%, B items: 45.72%) in demand. This reflects the managers' interpretation, whether correct or wrong, of the evolution of the demand series i.e. the way they were seeing the time series evolving (which, as discussed below, may or may not have matched the actual series evolution / direction), rather than information gained from customers or suppliers. Of course, we cannot exclude the possibility that some relevant information may occasionally have become available to the managers and thus influence their decisions but, without any further research data, this remains a supposition.

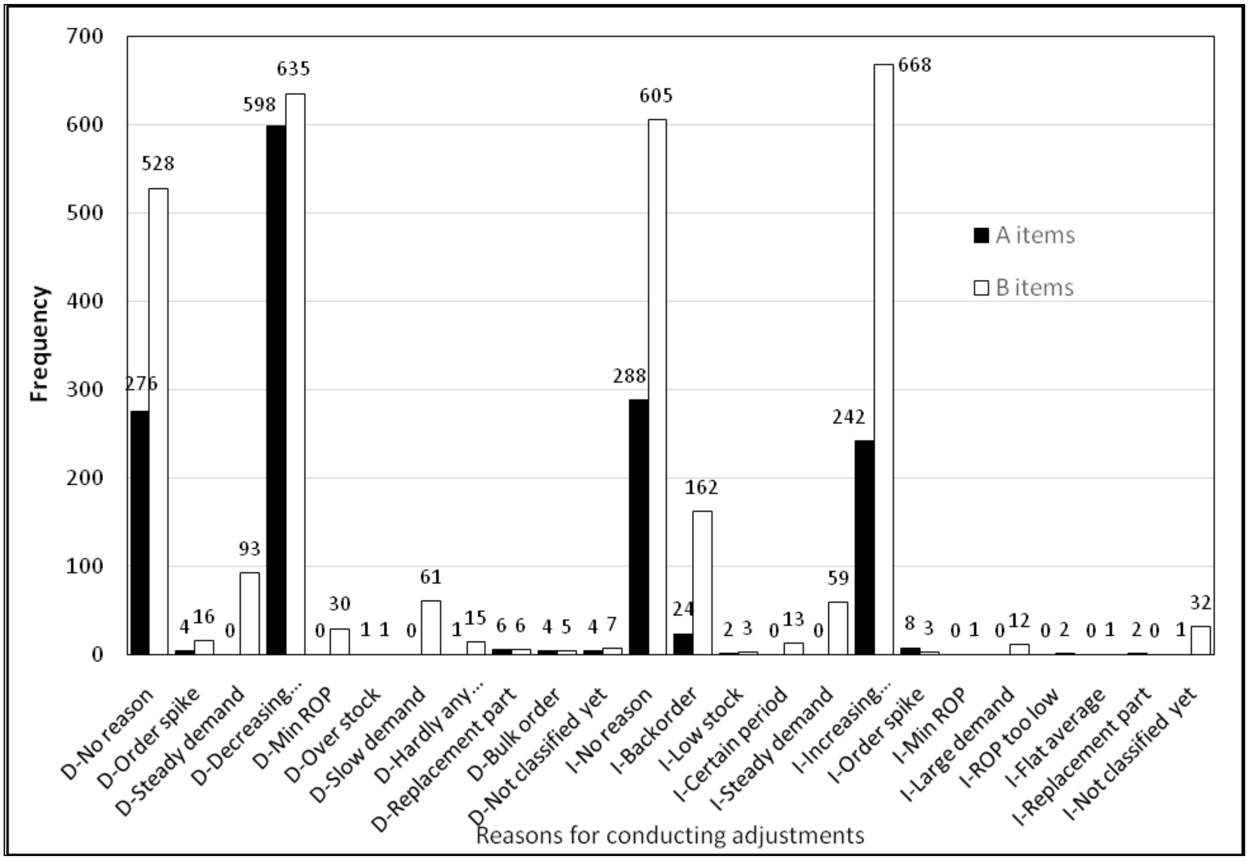

Figure 2. Frequency of adjustments per justification category for A items and B items
[The above information is also provided in a Pareto format in *Appendix C* in the EC]

A linear regression analysis was conducted using the past 24 weeks' demand data, for each point of intervention/ adjustment when such a justification has been provided, to assess whether or not non-stationary behaviour is present on the data.

1,160 cases (79.4%) in A-class SKUs reflect a consistency between what was identified by the managers and what our analysis has shown. That is, there was for example an increasing demand when such an increase was perceived by the managers. However, there was a great proportion of cases (20.6%) associated with the managers seeing a direction in the evolution of the demand series opposite to what was actually happening. There were 184 demand snapshots perceived as increasing, when in fact demand was decreasing (or was stable), and 117 cases where the opposite was the case. Similarly, for B-class SKUs, there are 2,234 (75.52%) cases where the justification and actual behaviour of the series were in accordance and 724 (24.48%) cases where a wrong direction of the demand data was perceived – in 388 cases decreasing or stable demand patterns were perceived as increasing, and in 336 cases increasing or stable demand patterns were perceived as decreasing.

The above results indicate that, in adjusting the OUT level, managers may make significant errors. Misinterpretation of the underlying properties of demand series is a common problem often being observed in forecasting related studies (Goodwin, 2005). Alternatively, it may well be the case that although managers do possess and use important contextual information, for reporting purposes, and to increase convenience, they always report the same justification! From a theoretical perspective, justification of adjustments such as that related to 'decreasing demand' should be related to forecasting. This is because the underlying structure of the series, such as that related to a trend, should be important for extrapolation purposes only. However, in the case of the company considered such a justification is offered in the context of inventory rather than forecasting. Clearly this stems from the lack of judgemental adjustments in the preceding stage of forecasting. Should that be the case most probably the perceived 'decreasing demand' would be taken into account when adjusting forecasts. This generates a number of interesting questions on the interface between adjusting at the forecasting and/or at the inventory stage. Too often forecasts are adjusted, for example, because they are confused with decisions (e.g., Fildes et al., 2009). That is, managers treat forecasting as if this were an end in itself, whereas of course forecasts always translate to a particular decision (OUT levels in our case, or more generally an inventory related decision). However, here we have an example where the decision (OUT levels) is adjusted because it is confused with a forecast.

6. SIMULATION EXPERIMENT AND ANALYSIS

6.1. The simulation model

As discussed before, there are 359 and 1,454 SKUs / demand histories, in the A and B class respectively, that are associated with adjustments. However, only 179 A-class and 228 B-class SKUs have been utilised for simulation purposes on the basis of having at least eight consecutive replenishment order observations. We appreciate that this may indeed look ad-hoc but a decision needed to be made with regards to the trade-off between sufficient data considerations and the meaningful output of the simulation experiment. Demand data series over 26 monthly periods, the prices of SKUs and the replenishment order (unadjusted and adjusted) data was used for this experiment. Lead time is equal to 9 weeks (approximated by 2 months). We consider three opportunities for replenishing stock: the System OUT replenishment level (unadjusted OUT level), the Final OUT replenishment level (adjusted OUT level), and the SMA-Based OUT replenishment level.

The code has been written in Visual Basic embedded in the Excel version. This is a dynamic experiment in the sense that we evaluate what would have happened if no judgemental adjustments had taken place. In terms of the output of the simulation experiment we record the inventory investment, cycle service level (CSL) and fill rate (FR) for each SKU. The trade-offs between inventory investment and service levels (CSLs and FRs) achieved are taken into account in reaching any conclusions. At this point we should note that we did not have access to data that could reveal stock-out costs and as such we were unable to determine the relative weights that need to be attached to the investment and stock-out costs objectives. This implies that we can only judge performance to be improved when either investment costs decline without deterioration in the service metrics or the service metrics improve without an increase in investment costs.

Inventory investment is the cost of carrying inventory volume in a given period. The inventory investment is obtained by multiplying the average of inventory volume for a particular SKU by its cost, approximated by its price.

The CSL is the probability that there will be no stock-out, whereas the fill rate is the rate (or percentage) of demand satisfied directly from stock on hand (Syntetos et al., 2010). The CSL is calculated as (percentage of periods running with no stock-outs):

CSL = 100 – the percentage of stock-outs in the simulation length

The fill rate is obtained by the following formula:

$$FillRate = \frac{\sum_{t=3}^{26} demand_t - \sum_{t=3}^{26} backorders_t}{\sum_{t=3}^{26} demand_t} x100 \qquad (1)$$

Calculations are initiated in period (t =) 3 to allow for replenishments, associated with 2-period lead times, to take place.

Two scenarios are considered for simulation purposes. In the first scenario, the stock on hand and the orders are calculated as follows:

$$\text{Stock}_t = \text{Stock}_{t-1} - \text{Demand}_t + \text{Order}_{t-2} \qquad (2)$$

$$\text{Order}_t = \text{OUT Level}_t - \text{Stock}_t \qquad (3)$$

where: t is the current time period (month). Order_{t-2} is the order placed 2 periods ago and received in the current period t (due to the lead time being 2 periods).

In the second scenario, the stock on hand is calculated as above but the order quantity is defined as:

$$\text{Order}_t = \text{OUT Level}_t - \text{OUT Level}_{t-1} + \text{Demand}_t \qquad (4)$$

The difference between them is in the calculation of the order to be placed for replenishment purposes. The second scenario is the standard one used in analytical evaluations of the OUT policy and is correct under the linear assumption, which does not always hold true. The first one is an intuitively appealing representation of the process, and consideration of this setting allows us to evaluate performance under less strict assumptions than those implied by the second scenario.

The block diagrams in Figure 3 represent, in the *z*-domain, the derivation of Orders under both scenarios. $z^{-\tau_p}$ represents a pure pipeline delay, of lead-time τ_p, between an order being placed and the Order received into Stock. $\frac{z}{z-1}$ represents the integral of incoming rate variables, Demand and the receipt of Orders after the pipeline delay to give the Stock level as calculated by equation (2). Thus, Figure 3a is the whole system visualization of Scenario 1 given by equations (2) and (3), while Figure 3b is the equivalent for Scenario 2 as per equations (2) and

(4). Thus, Scenario 1 is seen to have feedback Stock information to Orders while Scenario 2 is open loop. Both scenarios capture the type of ‘push’ / ‘pull’ production logics that are found in the electronics sector (Gonçalves et al., 2005).

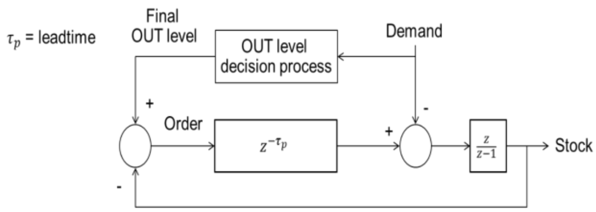

Figure 3a. Scenario1

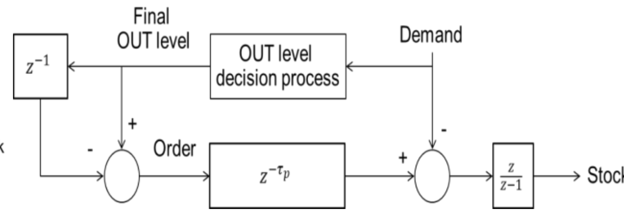

Figure 3b. Scenario 2

Figure 3. Block diagram representation of the experimental scenarios

6.2. Overall simulation results

Table 2 presents the results obtained from the two simulation scenarios considered. As can be seen from Table 2, the total inventory investment related to the judgementally adjusted orders of A items is slightly lower than that corresponding to the unadjusted ones for both scenarios. The decrease of inventory investment is 0.61% and 3.16% for scenario 1 and 2 respectively. Considering the trade-off between inventory cost and service, judgemental adjustments account for an improvement in terms of inventory investment at the expense though of an (expected) service reduction.

Table 2. Overall simulation results for A and B items

Replenishment order system	**Performance indicator**	**A items**		**B items**	
		Scenario 1	Scenario 2	Scenario 1	Scenario 2
System OUT replenishment level	Total inventory investment (€)	1,075,021	750,397	131,876	108,469
	Average CSL	0.991	0.924	0.986	0.889
	Average fill rate	0.993	0.948	0.962	0.891
Final OUT replenishment Level	Total inventory investment (€)	1,068,503	726,701	133,133	108,580
	Average CSL	0.991	0.905	0.987	0.889
	Average fill rate	0.993	0.930	0.965	0.893
SMA-based OUT replenishment Level	Total inventory investment (€)	1,036,226	685,263	135,394	109,746
	Average CSL	0.991	0.861	0.988	0.886
	Average fill rate	0.993	0.892	0.968	0.892

Turning now to the results for the B items, it can be seen from Table 2 that the Final OUT replenishment level is associated with a higher inventory investment as compared with the System OUT replenishment level for both scenarios. The increase is 0.95% and 0.10% for scenario 1 and 2 respectively. This increase results also to an increase of the service provision, though it is not particularly prevalent (less than 1%).

Sanders and Ritzman (1995) found, in the context of public warehousing operations, that the benefit attained from incorporating judgement into statistical forecasts relates to the underlying variability of the series being forecasted. Managerial intervention adds more value for greater levels of variability. Such findings were not confirmed by Syntetos et al. (2009) who looked specifically at intermittent demand series (such as the ones we analyse in this paper) from the pharmaceuticals sector. In that case, the results were mixed, favouring though slightly SKUs associated with higher demand frequency and less variability for the demand sizes – such as the A-items considered in this study (please refer to Tables A1 and A2).

Comparing the inventory investment of SMA-based OUT replenishment levels with the adjusted ones, it can be seen that the latter produce lower cost, but the difference is indeed very small: 1.67% (€2,260) in scenario 1 and 1.06% (€1,166) in scenario 2. However, and although this naturally leads to a slight decrease in the service measures under scenario 1, in scenario 2 the opposite occurs. This, in theory, indicates that adjustments lead not only to less safety stocks, as expressed through the inventory investment, but also to better service provision. It is true that the differences observed are very small but nevertheless the results favour the judgementally adjusted OUT levels.

The above findings are consistent with the results presented in most relevant studies in the forecasting field where judgemental adjustments account for (some) performance improvements, as previously highlighted in Section 3. However, the results indicate that there is less benefit resulting from judgementally adjusting stock control decisions than statistical demand forecasts. (In fact, although the magnitude of some of the adjustments may be 'large' (please refer to Table 1), the resulting effects are consistently small, suggesting *flat maxima* for the objective functions.) This finding is in agreement with the Syntetos et al. (2011) results, which showed that judgemental forecast adjustments have more prominent effects than judgemental order adjustments. Adjusting a stock decision is arguably more complex than adjusting a demand forecast since the former involves not only forecasting but the possible consequences of the decision as well. From the above discussion we provide an answer to the first **research question, Q1,** about the potential performance improvement resulting from judgementally adjusting stock control-related decisions; we find that human intervention offers a small advantage in stock control decision making.

In addition, it is worth noting that the Final OUT replenishment Level is a compromise between the System OUT replenishment level and the SMA-based OUT replenishment Level. The implication is that the managers are overall minimising risk and coming up with a trade-off solution between high inventories and service levels / fill rates. This is graphically depicted in Figure 4 where we summarize the simulation output for the B items under scenario 1. This is perhaps not too surprising given that, while maintaining customer service levels in the electronics industry is important, at the same time, with extremely short life-cycle products, the risk of obsolescence stock is high.

It is obvious that the service levels and inventory investment resulting from the Final OUT levels falls between those associated with the System and SMA-based OUT levels. Similar results have been found for the remaining combinations of classes of items and experimental scenarios. Previous work in the area of production scheduling has highlighted that human schedulers often end up making compromise solutions (McKay and Wiers, 1999; Berglunda and Karltuna, 2007). Similar findings have also been reported in the wider area of activity scheduling (e.g. Lundberg, 1988). Moreover, it is interesting to note that there is an association between the actual decisions being made (OUT levels in our case) and the outcome of these decisions (inventory investments and service levels achieved). An analysis of the adjusted OUT levels reveals that in 19.5% of the cases (across time and SKUs) the Final OUT levels fall between the SMA-based and System OUT levels. Considering also the cases where no adjustments have taken place, this percentage jumps up to 89.3%. This indicates that managers often make compromise decisions and, our simulation suggests that, those decisions also result in compromise outcomes.

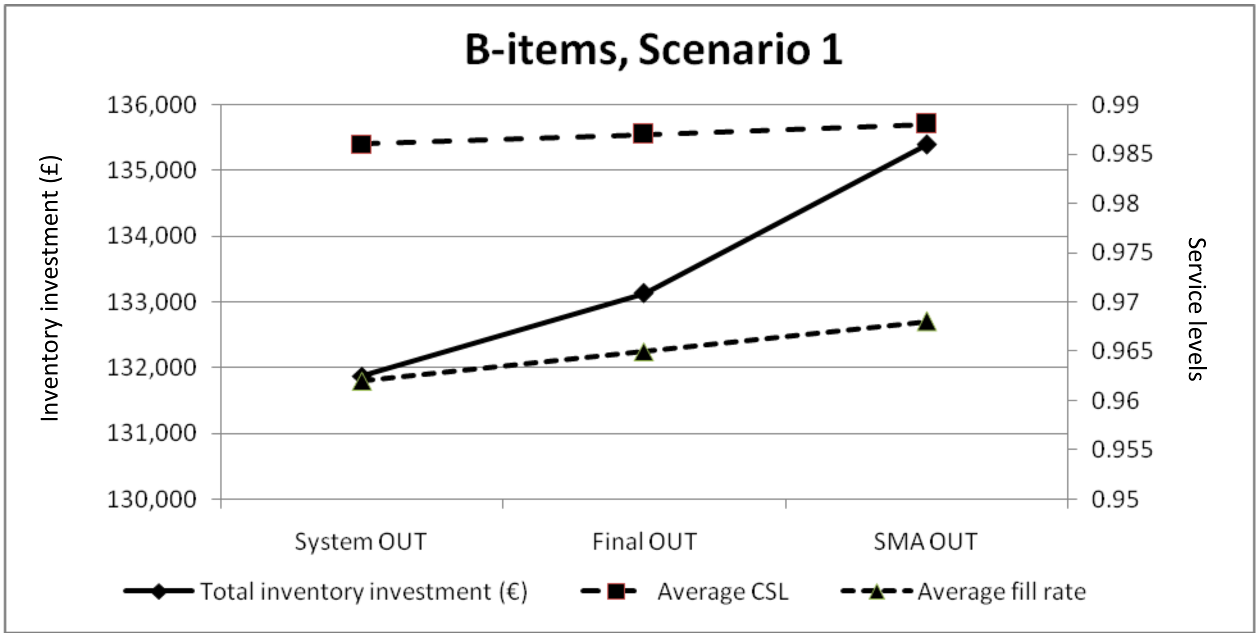

Figure 4. Analysis of compromise inventory outcomes

6.3. The effects of the sign of adjustments[2]

The purpose of this part of the analysis is to evaluate the effects of the sign of adjustments (positive/increasing adjustments and negative/decreasing adjustments) on inventory performance. To conduct such analysis, we consider the average adjustment per SKU across time (as compared to the average System OUT level) as it would simply be impractical to evaluate the effects of each adjustment separately (which has been the case in some forecasting related studies). The average adjustment is used to classify SKUs into two categories: positive and negative average adjustments. In the next step we analyse the inventory performance for each of these two groups in terms of inventory investment, cycle service level and fill rate across all the SKUs in each group.

For both A and B items, the negative adjustment category for the Final OUT replenishment levels produces a lower inventory investment as compared to the System OUT levels. On the other hand, the inventory investment associated with positive adjustments is higher than that related to unadjusted replenishment levels. Regarding the CSL and fill rate, the simulation

---

[2] The detailed analysis pertaining to sub-sections 6.3 – 6.5 can be found in *Appendix D* in the Electronic Companion of this paper.

produces similar values between the replenishment order methods being tested. Thus, we may conclude that negative adjustments deliver more benefit than the positive adjustments towards improving the inventory performance.

This result is in line with the outcomes of judgemental forecasting research which has found that negative adjustments perform well in increasing forecast accuracy for products that are subject to intermittent demand in the pharmaceuticals sector (Syntetos et al., 2009) and also for fast moving items in the pharmaceuticals, food, household products and retailing sectors (Fildes et al., 2009). The relatively poor performance of positive adjustments is a result of an optimism bias on the part of the managers. Forecasters tend to over-weight the statistical system's forecasts when contextual information is available (but in the absence of reliable evidence). Alternatively, and as previously discussed in this paper, excessive upward adjustments may be motivated by political factors such as pressure from senior management to obtain high service levels. Insights from the judgemental forecasting literature helps to explain why negative adjustments of replenishment related decisions perform better than positive ones.

6.4. The effects of the relative size of adjustments

The analysis of the effects of the size of adjustments is conducted by using the average of the absolute sizes of the adjustments (the sign of adjustments is not considered in calculating the average), followed by its expression as a percentage of the average System Out level for every SKU across time and then the classification of the SKUs into three categories: small, medium, and large adjustments.

To enable such a categorisation we define the cut-off points as follows: i) small adjustments if: 0 < average adjustment/average System OUT level ≤ 10%, ii) medium adjustments if: 10% < average adjustment/average System OUT level ≤ 20%, iii) large adjustments if: average adjustment/average System OUT level > 20%. For B items, as the smallest value of average adjustment/average System OUT level is 11.18%, we change the grouping into: less than or equal to 20% for small adjustments, between 20% and 40% for medium adjustments, and above 40% for large adjustments. Subsequently, the inventory performance is reported for each category and scenario.

For A items, our analysis reveals that in scenario 1 the medium adjustment category results to a best performance in term of inventory costs, while in scenario 2 the best performance is produced by the small adjustment category (≤10%). For B items, the small adjustment category (≤20%) results in the best performance under both experimental scenarios considered. The differences are not very big but one important finding is that, in contrast with the forecasting literature, large adjustments account for the least benefit. This suggests that managers are able to successfully 'fine-tune' the system OUT replenishment levels when such delicate refinements are required.

From the discussion conducted in sub-sections 6.3 and 6.4 we provide an answer to the second **research question, Q2,** about the potential effect of the sign and size of OUT level adjustments on the performance of inventory systems; we find that, in agreement with the forecasting literature, negative adjustments offer an advantage in stock control decision making. However, in contrast with the forecasting literature, large adjustments are found not to perform well.

6.5. The effects of adjustments' justification

This part of the analysis examines the effects of the justifications accompanying the adjustments on inventory performance. It is achieved by calculating the number of adjustments that are accompanied by a justification for each SKU and expressing that as a percentage of the total number of judgemental adjustments for the same SKU. Then, four categories are introduced based on that percentage: i) Justifications <25%; ii) 25%≤Justifications<50%; iii) 50%≤Justifications<75%; iv) Justifications≥75%. Subsequently, the inventory performance is separately summarised for each category. The findings indicate that the SKUs associated with justifications that are in excess of 75% of the cases (judgemental adjustments) are associated with the best trade-off between inventory investments and service levels achieved. That is, the higher the number of justifications the better the performance is. From the analysis we have conducted we provide an answer to the third **research question, Q3,** about the potential performance improvement resulting from adjustments that are accompanied by a justification; we find that the provision of a justification does lead to an improved inventory performance.

The case organisation has documented some of the justifications for the adjustments made by managers; however, managers often change the OUT level frequently without offering any reason for doing so. Providing the rationale behind an adjustment is useful in terms of the learning process of practitioners and towards an understanding of why decisions are erroneous (see also Section 3). In the area of forecasting, it has been suggested to document what the decision makers do when they adjust statistical forecasts since this should facilitate learning and hence lead to an improvement of the forecast performance (Franses and Legerstee, 2009). The same is suggested in the case of inventory control. Finally, it is also important to choose the best way in terms of presenting the justifications of the adjustments so transparency may be facilitated and communication of such information to other stake holders is eased. Provision, for example, of a drop-down menu to select the appropriate reason for intervening into the system, along with an 'Other' option that can be accompanied by the insertion of some relevant text, should lead to a uniform presentation of the justifications thereby allowing their easy descriptive analysis and their internal communication in an unambiguous manner.

7. CONCLUSION AND EXTENSIONS

Our work aimed at the investigation of the effects of incorporating human judgement into inventory-related decisions (in particular OUT levels). To our knowledge, this is the first study to consider such effects using a comprehensive company specific data set and to deliberate on the justifications behind such judgemental adjustments. A company representing the European logistics headquarters of a major international electronics manufacturer was considered for the purposes of our research. Hence, our findings are influenced by the specific context, where generally there is a high degree of innovation and rapid introduction of new products, meaning relatively short life-cycles with a high degree of substitution from one generation of product to the next and a high degree of obsolescence for either finished goods or spare parts. A number of research questions were developed based on the literature and a simulation experiment was constructed to evaluate the effects of judgemental adjustments in an empirical context.

By means of considering the trade-off between inventory costs and service levels achieved, the results revealed that judgemental adjustments account for an improvement in inventory investment. However, the effects are not prevailing in increasing the CSL or fill rate. Overall, the results indicate that human intervention offers a rather modest advantage in stock control decision making. This result contradicts previous empirical research in the area of demand forecasting conducted by Syntetos et al. (2009, 2010) which shows that the inventory implications of adjusting demand forecasts are considerable. Nevertheless, it is in line with previous simulation-based work (Syntetos et al., 2011) which shows that the effect of adjusting inventory decisions is less prominent than that associated with adjusting forecasts. Our results in conjunction with the already existing empirical knowledge base in the area of adjusting demand forecasts, confirm the findings of Syntetos et al. (2011) about the comparatively larger inventory implications of adjusting at the forecasting rather than at the inventory control level.

Our research has also shown that performance improvements are conditional to: a) the nature of the adjustments (size and direction); b) the demand class of the SKUs. With regards to the characteristics of the adjustments, we have found that negative adjustments result in better performance than positive ones, and that adjustments of a small/medium size perform better than larger ones. The former finding validates previous outcomes in the area of forecasting on the comparatively better performance of negative adjustments that reflect the asymmetric loss functions of managers that lead them to overly (unnecessarily) adjust upwards. The latter though is not in agreement with forecasting results which show that small adjustments are reaction to noise. In contrast, small adjustments in stock control account for a successful fine-tuning of the process and result in comparatively better performance than large adjustments.

With regards to the class of items, the results favour the better behaved time series, volatility-wise, and this is in agreement with previous findings in the area of judgemental adjustments for intermittent demand forecasts. They contradict however previous results for fast moving items where the positive effects of judgement increases with the underlying variability of the series. More research is needed to link the performance of adjustments to time series characteristics and to the differences between fast and slow moving SKUs.

Since the justification of the adjustments was recorded, we attempted to assess whether offering a justification is associated with a better performance. Justifications very often related to a perceived change in the underlying demand pattern, (recorded as 'demand increasing' or 'decreasing') something that our analysis showed though that very often was not the case. The results indicate that managers sometimes make significant errors in adjusting the OUT levels. However, it was also found that the very process of justifying one's own decision does lead to a better inventory performance – confirming previous speculations of that nature in the area of demand forecasting (Goodwin, 2000b).

Knowing the distribution that resembles the range and shape of the relevant decisions made by managers is very important in terms of the design of support systems. Through goodness-of-fit testing, some theoretical distributions were found that fit the distribution of judgemental adjustments. Based on such information, managers define specific percentiles above or below which adjustments are 'permitted'. For example, as we found that large adjustments are not performing well, managers might decide that no adjustments are allowed beyond, say, the 90^{th}percentile or they ask for recording information on the rationale behind large adjustments (exceeding such a percentile). Such functionality would complement well current attempts to 'control' the 'sign' of adjustments, and is something that could be considered by software developers for inclusion in existing decision support packages.

For example, and with reference to another sector in order to show the possible generalisation of our findings, Cash Management Solutions (CMS, http://www.cashmanagement.co.uk/),is a company that provides financial consulting and administers the cash management practices of a great number of clients including banks and building societies, bureau de change branches and ATMs. In such a context of application, forecasts and stock replenishment decisions refer to actual cash banknotes and smaller denominations. CMS produces a recommended order for every customer / location combination to cover requirements over the lead time plus review period. The order is then subject to evaluation at the local level, where the recommendations are viewed via an internet based cash ordering system. The outlet may then choose one of the following options: i) accept the recommended order; ii) suggest back a new 'requested order', where requested order > recommended order; iii) adjust directly the recommended order downwards without that being subject to approval from CMS. Therefore, only the upward adjustments need to be authorised (Donafee, 2012) which confirms the findings of this study (and those of other studies in the forecasting field) that downward adjustments perform well. Complementing such a system with ‘size’ related control mechanisms should potentially further enhance performance. However, it should be noted here that previous laboratory research (Goodwin et al., 2011) on the effects of restrictiveness in forecasting with regards to the size of adjustments (small adjustments were prohibited) did not provide encouraging results. The researchers found that although restrictiveness did reduce unnecessary adjustments, it also deterred desirable adjustments and also encouraged over-large adjustments so that accuracy was overall damaged.

From a theoretical perspective there is scope for further contributions and advancement in the area of Behavioural OR by exploiting an empirical data set as studied here. Given the prevalence of judgemental interventions into inventory management systems, further research in the following areas would appear to be merited:

- Although the dataset employed for the purposes of our research covers a wide range of scenarios both in terms of demand / cost combinations (that determine performance across an entire stock base) and judgemental adjustments (in terms of their descriptive statistics) and their justifications, replication of the analysis conducted here in other datasets/organizations would help to further understand and clarify ‘how’ managers perform adjustments in stock control and the conditions under which such adjustments add value to the system.

- Complementing such an analysis with interviews would be particularly important for linking ‘how’ to ‘why’ managers intervene into the system. Given the sensitivity of information related to judgemental adjustments conducting such interviews does not appear to be an easy task. However, analysis of datasets such as the one used in this study, coupled with interviewing the managers that conducted the adjustments would certainly introduce new perspectives into this area. Further, linking the background of the individuals to the effects of judgemental adjustments (along the lines, for example, considered by Eroglu, 2006, Eroglu and Croxton, 2010) is an exciting opportunity for further research and something the authors are currently trying to pursue with another company.

- In addition, in this study we have worked with a number of research questions, which admittedly may be reformulated into testable hypotheses. Our approach is in line with other empirical studies in the area of judgemental adjustments on demand forecasts (see, e.g., Syntetos et al., 2009), and in the field of forecasting in general, where many researchers have argued against testing for statistical significance (see, e.g., Armstrong, 2007). However, we recognize that hypothesis testing offers an opportunity for further research and although such an approach is conditioned to rather large datasets (that would allow statistically significant differences to emerge) it would offer an interesting methodologically-alternative addition to the literature.
- There are many decisions involved in inventory control (a wide spectrum of 'variables' that could be potentially judgementally adjusted), but we have only looked at one of those decisions, i.e. setting Order-Up-To (OUT) levels. Investigation of other organisational contexts, operating under different inventory policies, would help assess the validity of our results when looking at other relevant decisions (such as order quantities, re-order points, quantity discounts etc.) Similarly, and in terms of the effects of those decisions, we have looked only at the trade-offs between service levels and inventory investment. Although such trade-offs reflect to a great extent practitioners' concerns, there are other effects that one could have looked at, should appropriate data were available, such as, for example, order variability, batch sizes, lead times (as a result of potentially placing larger orders with the suppliers), etc.
- Previous studies have found that there is no learning effect over time in the forecasting function, neither from an organisational nor from an individual perspective (Lim and O'Connor, 1996; Klassen and Flores, 2001; Nikolopoulos et al., 2006; Syntetos et al., 2009). 'Learning' here implies how best to intervene into a forecasting system in order to produce more accurate forecasts. Further, Kolassa et al. (2008) discussed another type of learning, that associated with feeling more comfortable with the system in place and trusting its results – that is, learning not to interfere with the system as opposed to learning how best to interfere with the system. Investigation of learning effects into an inventory context would complement existing results in the forecasting literature and enable the design of

appropriate feedback mechanisms in real world applications. Further, and given the study by Kolassa et al. (2008) discussed above, elaboration on what actually constitutes 'learning' in a demand forecasting and inventory control context would appear to be merited.

- Many studies have investigated the issue of bias in the process of judgementally adjusting statistical forecasts (for example, Mathews and Diamantopoulos, 1990; Goodwin and Wright, 1994). As judgemental forecasting introduces bias, and the effects of bias impact on the performance of forecasting, it suggests that bias can also be found in the process of inventory decision making. By analysing whether a judgementally adjusted stock control decision is biased or not, further analysis could be conducted in order to investigate how and why managers are making adjustments.
- Expand the analysis to systems where adjustments are performed at both the forecasting and inventory control stage. A laboratory study of the effects of incorporating adjustments on already-adjusted forecasts has resulted in some very interesting insights (Onkal et al, 2008) and extending such work to subsequent empirical forecast-inventory interventions should be of great value for improving real world inventory management systems.

ACKNOWLEDGEMENTS

We would like to thank the anonymous referees for their comments that greatly helped to improve the content of our paper and its presentation.

REFERENCES

Armstrong, S. (2007). Significance tests harm progress in forecasting. *International Journal of Forecasting*, 23, 321-327.

Berglund, M., & Karltun, J. (2007).Human, technological and organizational aspects influencing the production scheduling process. *International Journal of Production Economics*, 110, 160-174.

Disney, S. M., Naim, M. M., & Towill, D. R. (1997). Dynamic simulation modelling for lean logistics. *International Journal of Physical Distribution and Logistics Management*, 27, 174-196.

Disney, S. M., Naim, M. M., & Potter, A. (2004). Assessing the impact of e-business on supply chain dynamics. *International Journal of Production Economics*, 89, 109-118.

Donafee, A. (2012). Personal communication to the corresponding author.

Eroglu, C. (2006). *An investigation of accuracy, learning and biases in judgemental adjustments of statistical forecasts*. Unpublished PhD thesis, The Ohio State University, USA.

Eroglu, C., & Croxton, K. (2010). Biases in judgmental adjustments of statistical forecasts: The role of individual differences. *International Journal of Forecasting*, 26, 116-133.

Fildes, R., Goodwin, P., Lawrence, M., &Nikolopoulos, K. (2009). Effective forecasting and judgemental adjustment: An empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting,* 25, 3-23.

Franses, P. H., &Legerstee, R. (2009). Properties of expert adjustments on model-based SKUs-level forecasts. *International Journal of Forecasting*, 25, 35-47.

Franses, P. H., & Legerstee, R. (2011). Combining SKU-level sales forecasts from models and experts. Expert Systems with Applications, 38, 2365-2370.

Franses, P. H., & Legerstee, R. (2013). Do statistical forecasting models for SKU-level data benefit from including past expert knowledge? *International Journal of Forecasting*, 29, 80-87.

Gonçalves, P., Hines, J., & Sterman, J. (2005).The Impact of endogenous demand on push-pull production systems. *System Dynamics Review,* 21, 187-216.

Goodwin, P. (1996). Statistical correction of judgmental point forecast and decisions. *Omega: The International Journal of Management Science,* 24, 551-229.

Goodwin, P. (2000a). Correct or combine? Mechanically integrating judgemental forecasts with statistical methods. *International Journal of Forecasting*, 16, 261-275.

Goodwin, P. (2000b).Improving the voluntary integration of statistical forecasts and judgment. *International Journal of Forecasting*, 16, 85-99.

Goodwin, P. (2002). Integrating management judgment and statistical methods to improve short-term forecast. *Omega: The International Journal of Management Science,* 20, 127-135.

Goodwin, P. (2005). How to integrate management judgmental with statistical forecasts. *The International Journal of Applied Forecasting*, 1, 8-12.

Goodwin P., Fildes, R., Lawrence, M., & Stephens, G, (2011) Restrictiveness and guidance in support systems. *Omega: The International Journal of Management Science* 39, 242-253.

Goodwin, P., & Wright, G. (1994). Heuristics, biases and improvement strategies in judgmental time series forecasting. *Omega: The International Journal of Management Science*, 22, 553-568.

Hamalainen, R. P., Luoma, J., & Saarinen, E. (2013). On the importance of behavioral operational research: The case of understanding and communicating about dynamic systems. *European Journal of Operational Research*, 228, 623-634.

Harvey, N. (1995). Why are adjustments less consistent in less predictable task situation? *Organizational Behavior and Human Decision Processes*, 63, 247-263.

Klassen, R. D., & Flores, B. E. (2001). Forecasting practices of Canadian Firms: Survey results and comparison. *International Journal of Production Economics,* 70, 163-174.

Kolassa S., Schütz, W., Syntetos, A. A., &Boylan, J. E. (2008). Judgemental changes to retail sales forecasts and automatic orders. Paper presented at the 28thInternational Symposium on Forecasting; 22-25 June 2008; Nice, France.

Lawrence, M., Edmundson, R. H., & O'Connor, M. (1986).The accuracy of combining judgemental and statistical forecasts. *Management Science*, 32, 1521-1532.

Leitner, J., Leopold-Wildburger, U. (2011). Experiments on forecasting behaviour with several sources of information - A review of the literature. *European Journal of Operational Research*, 213, 459-469.

Lim, J. S., & O'Connor, M. (1996). Judgemental forecasting with time series and causal information. *International Journal of Forecasting*, 12, 139-153.

Lundberg, C. G. (1988). On the structuration of multiactivity task-environments. *Environment and Planning A*, *20*, 1603-1621.

Mathews, B. P., & Diamantopoulos, A. (1986). Managerial intervention in forecasting: An empirical investigation of forecast manipulation. *International Journal of Research in Marketing*, 3, 3-10.

Mathews, B. P., & Diamantopoulos, A. 1990. Judgemental revision of sales forecasts: Effectiveness of forecast selection. *International Journal of Forecasting*, 9, 407-415.

Mathews, B. P., & Diamantopoulos, A. (1992). Judgmental revision of sales forecasts-the relative performance of judgementally revised versus non revised forecasts. *International Journal of Forecasting*, 11, 569-576.

McKay, K. N., &Wiers, V. (1999). Unifying the theory and practice of production scheduling. *Journal of Manufacturing Systems*, 18, 241-255.

Nikolopoulos, K., Fildes, R., Goodwin, P., & Lawrence, M. (2005). On the accuracy of judgemental interventions on forecasting support systems. *Lancaster University Management School Working paper 2005/022*.

Nikolopoulos, K., Stafylarakis, M., Goodwin, P., & Fildes, R. (2006). Why do companies not produce better forecasts over time? An organizational learning approach. In: Proceedings of the 12th IFAC Symposium on Information Control Problems in Manufacturing, Saint Etienne, France, 3, 167–172.

Olsmats, C.M.G., Edghill, J.S., & Towill, D.T. (1988). Industrial dynamics model of a close-coupled production-distribution system. *Engineering Costs and Production Economics*, 13, 295-310.

Onkal, D., & Gonul, S. (2005). Judgemental adjustment: A challenge for providers and users of forecasts. *Foresight*, 1, 13-17.

Onkal, D., & Gonul, S., &Lawrence, M. (2008). Judgmental adjustments of previously adjusted forecasts. *Decision Sciences*, 39, 213-238.

Onkal, D., Goodwin, P., Thomson, M., Gonul, S., &Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making*, 22, 390-409.

Sanders, N. R. (1992) Accuracy of judgmental forecasts: a comparison. *Omega: The International Journal of Management Science*, 20, 353-364.

Sanders, N. R., & Manrodt, K. B. (1994). Forecasting practices in US corporations: survey results. *Interfaces*, 24, 92-100.

Sanders, N. R., & Manrodt, K.B. (2003).The efficacy of using judgmental versus quantitative forecasting methods in practice. *Omega: The International Journal of Management Science*, 31, 511-522.

Sanders, N. R., & Ritzman, L. P. (1995). Bringing judgment into combination forecasts. *Journal of Operations Management,* 13, 311-321.

Sanders, N. R., &Ritzman, L. P. (2001).Judgmental adjustment of statistical forecasts. In J. S. Armstrong (Ed.), *Principles of Forecasting* (pp. 405–416). Boston: Kluwer Academic Publishers.

Syntetos, A. A., Georgantzas, N. C., Boylan, J. E., & Dangerfield, J. E. (2011). Judgement and supply chain dynamics. *Journal of the Operational Research Society*, 62, 1138-1158.

Syntetos, A. A., Nikolopoulos, K., Boylan, J. E., Fildes, R., & Goodwin, P. (2009). The effects of integrating management judgement into intermittent demand forecasts. *International Journal of Production Economics*, 118, 72-81.

Syntetos, A. A., Nikolopoulos, K., & Boylan, J. E. (2010). Judging the judges through accuracy-implication metrics: The case of inventory forecasting. *International Journal of Forecasting,* 26, 134-143.

Teunter, R. H., Syntetos, A. A., & Babai, M. Z., (2011). Intermittent demand: Linking forecasting to inventory obsolescence. *European Journal of Operational Research*, 214, 606-615.

Wolfe, C., & Flores, B. (1990). Judgemental adjustment of earning Forecasts. *International Journal of Forecasting*, 9, 389-405.