Improving warehouse labour efficiency by intentional forecast bias

Abstract

Purpose – This paper shows that intentional demand forecast bias can improve warehouse capacity planning and labour efficiency. It presents an empirical methodology to detect and implement forecast bias.

Design/methodology/approach – A forecast model integrates historical demand information and expert forecasts to support active bias management. A non-linear relationship between labour productivity and forecast bias is employed to optimise efficiency. The business analytic methods are illustrated by a case study in a consumer electronics warehouse, supplemented by a survey among thirty warehouses.

Findings – Results indicate that warehouse management systematically over-forecasts order sizes. The case study shows that optimal bias for picking and loading is 30-70 percent with efficiency gains of 5-10 percent, whereas the labour-intensive packing stage does not benefit from bias. The survey results confirm productivity effects of forecast bias.

Research implications – Warehouse managers can apply the methodology in their own situation if they systematically register demand forecasts, actual order sizes and labour productivity per warehouse stage. Application is illustrated for a single warehouse, and studies for alternative product categories and labour processes are of interest.

Practical implications – Intentional forecast bias can lead to smoother workflows in warehouses and thus result in higher labour efficiency. Required data includes historical data on demand forecasts, order sizes and labour productivity. Implementation depends on labour hiring strategies and cost structures.

Originality/value – Operational data support evidence-based warehouse labour management. The case study validates earlier conceptual studies based on artificial data.

Keywords Warehouse planning, Demand forecasting, Labour management, Labour efficiency, Forecast bias

Paper type Research paper

Introduction

Warehousing serves as the primary link between producers and customers in the supply chain. It provides buffering for manufacturing operations to manage varying customer demand (Bowersox *et al*., 2002). Labour constitutes about half of all (non-automated) warehouse operation costs (Bartholdi and Hackman, 2016). For retail warehouses it is often difficult to determine the exact workforce, as the workload tends to be variable and activities, especially outbound work, have tight deadlines. Many warehouse managers therefore prefer flexible labour pools (De Leeuw and Wiers, 2015). Even with flexible pools, labour planning may be inaccurate with negative effects on labour productivity. Forecasting the workload and hence the required capacity is therefore an essential step in warehouse manpower planning (Bond, 2012). As managers usually have a good view of upcoming orders, quantitative forecasting methods using historical data can be combined with expert judgement, although this may introduce bias, i.e. systematic differences between forecasts and actual order sizes (Goodwin, 1996, 2002). Important questions are how to detect such biases, how to control them, and how they affect labour efficiency, defined as the ratio of required labour over actually hired labour.

This paper presents an empirical methodology to detect forecast bias, defined as the ratio of forecast error over actual order size. It shows how to implement a controlled level of bias to optimise labour efficiency in warehousing, in particular for Original Equipment Manufacturer (OEM) warehouses serving retail distribution centres. The main three research questions are the following. What is the quantitative nature of errors in demand forecasting? How does forecast bias affect labour efficiency? What is the optimal level of forecast bias to optimise labour efficiency? Two statistical models are used to investigate these research questions. One relates historical demand patterns to expert forecasts and forecast errors, and the other relates historical labour efficiency measurements to forecast bias.

The business analytic methods are illustrated in a case study conducted at Samsung Electronics. The analysis provides an empirical test of the theoretical claim in Sanders and Graman (2009) that forecast bias can improve labour efficiency and extends recent work of Van Gils *et al*. (2017) on forecasting methods for personnel planning. The data for the first research question consist of weekly time series on management forecasts and on actual orders. The second and third question are analysed by means of daily labour productivity data and order forecast biases, where productivity is measured at the three consecutive stages of the outbound warehouse operations: picking, packing and loading. To supplement the case study results, information from a survey among thirty warehouses is used to further investigate the productivity effects of forecast bias.

Although the specific details of our empirical findings are case-dependent, the methodology and general conclusions are relevant for warehouse labour management and forecasting. By following this methodology, warehouse managers can determine the level of forecast bias that works best for their situation. Figure 1 depicts the methodology as a flow diagram incorporating data, models and activities in data-driven adaptive labour management. We were able to implement this methodology in our case study because the company under investigation continuously collects and stores the required analytics data, including a refined barcode-based labour management system that registers and stores the activities of each individual worker. The company has incorporated the results presented here in the evaluation and redesign of their interrelated management strategies for demand forecasting and labour planning.

This paper is structured as follows. We first provide a literature review and formulate our research hypotheses. Next, we describe the case study environment and summarise data and methods. Then we present our results, starting with the statistical analysis of forecast errors and the demand forecast model in Figure 1, followed by the empirical investigation of the relationship between labour efficiency and forecast bias using the labour efficiency model in Figure 1. We also discuss the results of a small-scale survey among thirty warehouses. Finally, we summarise operational implications of our analysis and discuss topics for future research.

Insert Figure 1 about here

Literature review and research hypotheses

This literature review examines the three main aspects of our methodology: warehousing, labour management and demand forecasting.

Warehousing

Warehouses receive products in large quantities, reorganise and repackage them and send them out in smaller quantities. Warehouse operations thus consist of inbound processes (receiving and storing goods) and outbound processes (order-picking, packing and shipping). As warehouses can usually regulate the inflow of products well because of tight links with their suppliers, most research studies focus on stocking and outbound processes (Bartholdi and Hackman, 2016). Several types of warehouses can be distinguished, e.g. warehouses for retail distribution, for spare parts and for e-commerce. This paper considers OEM warehouses serving retail distribution centres. Such warehouses often play a subordinate role between sales departments and buyer purchasing departments. Their outbound processes tend to be under time pressure as their customers usually operate in just-in-time modes and require deliveries at short notice. Their inbound processes are generally under less time pressure as incoming goods are typically stored as safety stock without small unit handling (Bartholdi and Hackman, 2016), except for cross-docked goods. Such warehouses must adapt their workforce to fast demand fluctuations, which makes accurate short-term workload forecasting and efficient labour management essential for smooth operations (Van Gils *et al*., 2017). Although automation is steadily spreading, De Koster *et al*. (2007) report that 80% of all warehouses apply manual picking. Order picking still accounts for about 60% of total labour costs and 50% of overall operational costs (Bartholdi and Hackman, 2016). Reviews of the warehousing literature (De Koster *et al.,* 2007; Gu *et al*., 2010; Gong and De Koster, 2011; Bartholdi and Hackman, 2016) show that most research has focused on warehouse design and improved order picking, such as optimising picking routes and selecting good storage locations for fast picking.

Our paper examines interlinked outbound warehouse processes and focuses on labour efficiency and demand forecasting. It does not consider inbound processes and associated inventory strategies, or optimisation of used warehouse space.

Labour management

Customer demand of warehouses for retail distribution is usually characterised by short-term fluctuations. A recent survey by De Leeuw and Wiers (2015) indicates that many warehouses employ both permanent staff and temporary labour to accommodate workload fluctuations. Although temporary work agencies can provide workers at short notice, worker quality may be lower. Some warehouses deliberately operate with excess permanent staff capacity if they cannot rely on temporary staff being available at the right time (Van den Berg, 2007). Some studies (Brusco and Johns, 1995; Riley and Lockwood, 1997) have investigated the balance between permanent and temporary labour under the restrictive assumption that fluctuations are known in advance. Managing the number of warehouse workers remains a key issue in tackling daily demand fluctuations (Ruben and Jacobs, 1999; De Koster *et al*., 2007).

Our paper analyses the effect of an intentional level of forecast bias on labour efficiency and uses this relationship to determine the optimal level of bias to maximise this efficiency.

Demand forecasting

Demand and workload forecasting are crucial steps in manufacturing and warehouse management. Forecasts can be based on expert judgement or on quantitative methods. The first are easy to make and include information on future orders, but are often systematically biased (Goodwin, 1996, 2002). Quantitative methods using historical data are more complex, but often provide more accurate forecasts (Sanders and Manrodt, 2003). Even though unbiased forecasts are valuable for management purposes, some bias may be preferable if the costs of over-forecasting differ substantially from those of under-forecasting. Sanders and Graman (2009) provide simulation evidence that properly managed forecast biases reduce costs if the bias is related to labour and inventory costs. Under-forecasting is attractive if labour costs dominate, whereas over-forecasting is more appealing if the main costs are delay penalties for stock-outs. Ritzman and King (1993) stress the relevance of forecast bias for inventories in multistage manufacturing, as buffers between stages like picking and packing are helpful in creating flexibility in labour and capacity utilisation. They warn for undesirable biases that originate from optimistic sales projections and misguided attempts at inventory reduction.

Demand forecasts can often be improved by combining expert judgements and historical sales statistics. Combination forecasts outperform individual forecasts especially when the latter employ diverse sources of information (Aiolfi *et al.*, 2011). Managers incorporate qualitative information in their forecasts, but cannot extrapolate recent demand trends as accurately as statistical models. Wacker and Lummus (2002) emphasise the importance of the managerial side of sales forecasting. Managers need to understand the value and the limitations of forecasting strategies to be able to adopt them successfully. Van Gils *et al.* (2017) investigate several statistical forecasting methods for determining order picking work in a case study of a Belgian warehouse. A method combining exponential smoothing with (SARIMA) time series models outperforms current expert forecasts, but the authors do not consider bias or the integration of expert forecasts and statistical forecasts.

Our paper investigates how expert forecasts and historical sales information can be integrated to improve demand forecasts and how a certain level of intentional forecast bias can be implemented to optimise labour resource planning.

Research hypotheses and contributions

The brief review above and our focus on the relation between forecast bias and labour productivity leads us to formulate the following three research hypotheses:

Hypothesis 1

Expert forecasts of managers display systematic bias related to cost considerations.

Hypothesis 2

Integrating expert forecasts in statistical models supports intentional management of forecast bias.

Hypothesis 3

A controlled amount of intentional forecast bias derived from operational warehouse data improves labour efficiency.

Support for these three hypotheses in our study can provide the basis for the following three-step, business analytics strategy to optimise warehouse labour efficiency:

- Maintain a periodically (daily or weekly) updated database with (*ex ante*) management demand forecasts and (*ex post*) received order sizes.
- Implement a detailed labour productivity measurement system at the disaggregated level of individual activities and workers.
- Optimise labour capacity per activity by means of the relationship between demand forecast bias and productivity.

In terms of our case study, Hypothesis 1 provides an empirical test of the assertion in Goodwin (1996, 2002) that judgemental forecasts are often systematically biased because of asymmetric loss considerations. We relate management forecast bias of the warehouse in our case study to its labour cost structure. To test Hypothesis 2, we implement and empirically validate the recommendations in Sanders and Ritzman (1991, 2004) and in Goodwin (2002), and use composite methodologies that integrate judgemental and statistical forecasts. The evaluation of the quantitative forecast gains of this integration in the warehouse of our case study supplements the cross-firm survey results of Sanders and Manrodt (2003) on the benefits of quantitative methods compared to judgemental methods. Finally, Hypothesis 3 is closely connected to bias management proposed in Ritzman and King (1993) and to forecast bias exploitation proposed in Sanders and Graman (2009). By using real-world data, our results provide an empirical validation of these two studies, which were based on artificially generated data to support their proposals.

Case study environment

Warehouse characteristics

We illustrate our methodology in a case study for a Samsung Electronics warehouse in Western Europe. The warehouse has 250,000 pallet storage places, including racking and bulk storage, with a total space of 50,000 square meters. The products comprise finished goods in consumer electronics. These are fast-moving items with a total inventory volume of less than two weeks of demand. Distribution is a labour-intensive operation, and labour constitutes more than 40% of total warehouse costs. The outbound process of this warehouse has a multi-server queuing structure with consecutive stages of picking, packing and loading. As the warehouse delivers goods to customer warehouses instead of to end customers, delivery sizes are massive and on average comprise more than ten pallets. To satisfy delivery size constraints at destination, 80% of the pallets need re-palletising at the packing stage. On average this stage requires 85% of total labour, and overall outbound labour efficiency depends crucially on a smooth packing workflow. It is therefore essential that the pallets are retrieved from the picking stage and transported to the loading area as quickly as possible to prevent workflow disruption at the packing stage.

Demand characteristics

Warehouse operation volume is measured by the number of boxes handled. Figure 2 shows weekly order sizes from week 38 of 2009 to week 48 of 2012 (167 observations). The long-term average is rather stable, with considerable seasonal and short-term fluctuations. The annual cycle shows the same pattern for all four years. The September-November peak is typical for consumer electronics, and the end-of-the-month peak is related to the retailer behaviour, for example, to meet sales targets. Such historical sales patterns help sales managers with their weekly forecasts.

<u>Insert Figure 2 about here</u>

Labour characteristics

The warehouse translates the weekly order size forecasts into daily labour hiring decisions based on expected order sizes and past experiences. As order sizes fluctuate considerably, the warehouse employs flexible labour pools of about 60 full-time workers for each of two eight-hour shifts per day from a third-party logistics (3PL) provider of temporary labour. The provider permits the warehouse to furlough workers without payment if they have worked for more than three hours, and if the remaining workload does not justify hiring them for the remainder of the shift. This arrangement limits the costs of over-forecasting. On the other hand, if labour is insufficient, then impromptu demand for extra workers can often only be satisfied by hiring novices who are less productive, so that under-forecasting is costly.

As discussed in the literature review, forecast bias may reduce costs, but it depends on the labour situation which kind of bias is profitable (Sanders and Graman, 2009). As over-forecasting is less expensive than under-forecasting for the warehouse of the case study, we specialize our Hypothesis 3 as follows:

Hypothesis 3*

In the case study warehouse, where over-forecasting is less expensive than under-forecasting, some amount of over-forecasting is beneficial for labour productivity.

Case study data and statistical methods

The warehouse has an active database management strategy with weekly recorded order sizes and management demand forecasts, daily recorded handled orders and labour hours, and continuously measured and daily recorded labour productivity data per stage of the warehouse process.

In our empirical analysis, we compare weekly management forecasts with weekly order sizes. A regression-type forecast model with lagged effects integrates these forecasts with historical sales data. We follow the Box-Jenkins methodology (Box et al., 1994), which consists of the stages of model identification, estimation and diagnostic checking. Because of its simplicity, this forecast methodology has found widespread application in business and in other fields. Because warehouse orders are unstable, we extend the forecast model by including seasonal effects and management forecast information. For the model identification stage, we use the "general-to-specific" procedure (Hendry, 1995). This procedure has the advantage of working with correctly specified models as it starts with a general model where no factors have been omitted, followed by model simplification by removing insignificant factors. The models are estimated by (ordinary or recursive) least squares and simplified by standard tests. Standard diagnostics check for correct specification of the dynamic structure of the demand process (Breusch, 1978; Godfrey, 1978) and for normality of the model residuals (Jarque and Bera, 1987). The significance of forecast gains is evaluated by forecast comparison tests, including standard paired t-tests and encompassing tests (Hendry, 1995). An encompassing test for two forecast methods, A and B, is based on the following regression equation:

$$O(t) = \alpha + \beta \times F_A(t) + (1 - \beta) \times F_B(t) + \varepsilon(t) \qquad (1)$$

Here O is the variable of interest, order size in our case, F_A and F_B are the forecasts of methods A and B and ε is an error term. Method B is said to encompass method A if $\beta = 0$, that is, if the forecast of method A does not add to the forecast power of B. Similarly, method A encompasses B if $\beta = 1$, and the two complement each other if $0 < \beta < 1$.

In our further empirical analysis, we use daily data to study the relationship between forecast bias and labour efficiency. The weekly sales forecast is split into daily forecasts, based on historical spreads over the days and on operational information such as order cancellation notifications and postponed orders of previous days. Labour efficiency is measured continuously. Task durations are measured in seconds by time clock systems for picking, packing and loading activities. For each of these activities, the warehouse employs standard durations based on about 50 sub-tasks. The labour efficiency of each activity is automatically registered in the IT system daily, by comparing clock system data with standard durations. We compare efficiency on days with positive and negative bias, and we estimate the optimal amount of bias for each stage of the outbound warehouse process. In addition to the case study results on the relationship between forecast bias and labour efficiency, we also present the outcome of a small-scale survey among thirty warehouses and results of a simple simulation study.

Forecasting order size

Statistical model

To illustrate our forecast methodology, we analyse sales department forecasts and actual orders processed by the logistics department. The weekly data consist of 167 observations of management forecasts (denoted by F) and actual orders (denoted by O), both of which are measured in terms of number of boxes. For week t, the management forecast F(t) is confirmed on Monday morning of week t, and the actual delivery order O(t) is confirmed at the following Thursday's cut-off (as later orders are carried over to the next week). As the order size tends to be relatively large at the end of the month and at the end of the year (see Figure 2), we include end-of-the-month effects (for the last week of the month) and end-of-the-year effects (for weeks in September, October and November) as possible calendar effects. We also consider the forecast accuracy of previous weeks. For example, if management forecasts for the previous week underestimated the actual order size, a similar bias may apply for the current week. In other words, forecast error $E(t) = F(t) – O(t)$ may have predictive power for future orders.

The statistical forecast model is obtained from the "general-to-specific" specification procedure. The starting point is a relatively rich model including information on order sizes and management forecasts of up to the last four weeks. We simplify this model by testing for various parameter restrictions and we apply diagnostic tests (on absence of serial correlation and on normality of model residuals) for the simplified models. We finally consider inclusion of calendar effects. The forecast model obtained by this procedure is

$$O(t) = 15{,}512 + 0.861\times F(t) - 0.195\times E(t-1) - 0.190\times E(t-4) + r(t) \quad (2)$$

Here r(t) denotes residuals of the model, which contain no significant serial correlation (p-value for lags 1-4 is 0.43) and which are reasonably normal (p-value 0.06). Additional calendar effects are insignificant (p-values 0.72 for end-of-the-year, 0.42 for end-of-the-month and 0.72 for both effects jointly), which can be explained by the fact that management forecasts, which is included as an explanatory factor, already includes these calendar effects.

Model interpretation

The model in equation (2) integrates judgemental forecasts and statistical sales data and can be interpreted in terms of bias correction (Goodwin, 2002). The coefficient 0.861 of F(t) means that approximately 86% of the management forecast for the coming week is taken as the expected order, with corrections of about 20% of the management forecast error from the previous week and from four weeks earlier (which in most cases has the same position within the month as the upcoming week). If previous forecasts were too high (with error $E = F – O > 0$), the current forecast is corrected downward, and if they were too low, it is corrected upward.

As the model is obtained ex post and uses all available data, the forecasts from model (2) are not made in real time as they employ future data that were used to obtain numerical values of the coefficients. Real-time statistical forecasts are obtained if, for each week t, the model is estimated using only data that were available at the beginning of week t (i.e. management forecasts F for weeks up to and including t, and order sizes O for weeks up to and including t - 1). We construct such ex ante forecasts by re-estimating the above model with factors F(t), E(t - 1), and E(t - 4) by means of recursive least squares with different coefficients for every week t. The comparison of management forecasts and ex ante model forecasts is a fair one, as both methods use compatible information sets of past historical data at each forecast week. One might expect that ex post forecasts are qualitatively better than ex ante forecasts because the latter employ less information.

Forecast comparison

The empirical results are summarised in Table 1 for all weeks and for three busy periods: end-of-the-month (EM), end-of-the-year (EY, September through November), and end-of-the-month weeks in these three months (EMY). Management forecasts are consistently upward biased, and relative bias increases with order size. They are larger than actual order size in 56% of all weeks, 58% in EY weeks, and 72% in EM weeks. For the ex ante model, these percentages are 52%, 50% and 53%, respectively, showing a better balance between over-forecasting and under-forecasting. The ex ante forecasts have a much smaller bias and standard deviation, and perform only slightly worse than the ex post forecasts. The average management forecast bias is 3% and 6%-12% in busy periods (8.5% in EM, 6.4% in EY, and 11.9% in EMY). The ex ante forecast bias is less than 2% on average, also in busy periods (1.7% in EM, 1.0% in EY, and 0.7% in EMY). Measured by the root mean squared prediction error, which combines bias and variance, the error decreases from 16.4% to 13.5% on average (from 16% to 10% in EM, from 19% to 13% in EY and from 22% to 12% in EMY). The ex post forecasts are only slightly better. All these findings support Hypothesis 2 that integrating expert forecasts and historical sales data reduces forecast bias.

The lower part of Table 1 shows outcomes of various forecast comparison tests. The ex-ante forecasts have a significantly smaller bias (at 5% level) than management forecasts, and the ex-post bias is significantly smaller than the ex-ante bias only when evaluated over all forecasts (but not for the busy sub-periods). The ex-ante forecasts encompass management forecasts in all four cases (all weeks and the three busy sub-periods), but management forecasts never encompass the ex ante forecasts. This means that the ex-ante forecasts are more reliable than management forecasts, and that the ex ante forecasts cannot be further improved by taking a weighted combination average with management forecasts. The ex-post forecasts and the ex-ante forecasts are of equal quality, except if all weeks are considered, in which case the ex-post forecasts encompass the ex-ante forecasts.

It is also useful to compare forecasts in terms of absolute prediction errors, as an alternative to bias where large positive and negative errors cancel out. Ex ante statistical forecasts have smaller mean absolute prediction error than management forecasts: the error decreases from12% to 11% in all weeks, from 11% to 9% in EM, from 14% to 10% in EY and from 17% to 10% in EMY. This difference is significant for all weeks and for end-of-the-year weeks.

We further mention that management forecasts contain crucial information. If this information is excluded and order forecasts are based only on past orders and calendar effects, the general-to-specific procedure produces a model that contains order sizes of one and four weeks earlier as well as significant end-of-the-year and end-of-the-month effects. This model has about 80% higher root mean squared forecast error than model in equation (2), which provides further support of Hypothesis 2, and shows that the combination of judgement and statistical information provides better sales forecasts than can be obtained from either separately.

Insert Table 1 about here

Labour efficiency and forecast bias

In this section, we first investigate Hypothesis 3* by comparing labour efficiency on days with positive and negative forecast bias. Next, we analyse the relationship between forecast bias and labour efficiency to determine the amount of bias that optimises this efficiency, and we perform a small simulation study that confirms the empirical case study findings. Finally, we present outcomes of a small-scale survey among thirty warehouses that confirm labour efficiency benefits of forecast bias.

Comparison of labour efficiency for days with positive and negative forecast bias

We investigate the relationship between forecast bias and labour efficiency for the warehouse in our case study. Labour efficiency is defined as the ratio of required labour over hired labour, so that an efficiency above (below) 1 corresponds to labour productivity being higher (lower) than standard. Forecast bias is defined as the ratio (F – O)/O where F and O are respectively the forecast and the actual order size, so that a positive (negative) forecast bias corresponds to management over-forecasting (under-forecasting).

Daily labour efficiency data are available for the first 40 weeks of 2012. The total number of observations is 195 (40 weeks of five working days, excluding five bank holidays).

Daily information on actual order sizes are available for this period, as well as daily order forecasts derived from sales managers' weekly order forecasts. Table 2 shows the distribution of actual order sizes and management forecasts over the five working days of the week. The forecasts are considerably downward biased for Mondays and upward biased for the end of the week. One possible cause of these biases is a shifting demand pattern over the week compared to previous years. The table shows approximate daily distributions for 2008–2011 reported in interviews with warehouse managers, and the bias for Mondays may have been caused by these past expectations. Consequently, the forecast bias varies considerably and contains some aberrant values. In our analysis, we sometimes exclude aberrant observations by restricting the sample to days when the ratios of Forecast over Order (F/O) and of Order over Forecast (O/F) are both at most 1.5, so that the forecast bias (F-O)/O lies between -1/3 and +1/2, which eliminates 63 days (including 18 Mondays and 27 Fridays).

Insert Table 2 about here

Table 3 shows the effect of forecasting bias on labour efficiency. Average efficiency is the highest for loading, followed by picking, and the lowest for packing. The relatively high efficiency in picking and loading does not lead to appreciable efficiency gains in the overall outbound activities (efficiency of 1.034, close to 1). The mean daily efficiency of picking, loading and overall outbound procedures is significantly higher for days with a positive forecast bias than for days with a negative forecast bias. Compared to days with negative bias, the efficiency gain on days with positive bias is approximately 12% for loading, 3% for picking and overall outbound handling and 0% for packing. The results for the restricted data set, after eliminating aberrant observations, are very similar. The table also reports outcomes of rank comparison tests for all days with negative and positive bias. These tests are not sensitive to outliers in efficiency, and the results confirm those of conventional mean comparison t-tests described earlier. All these findings support Hypothesis 3*: for the given labour situation of this warehouse, over-forecasting is beneficial for labour productivity. More specifically, because packing comprises 85% of all outbound labour, efficiency is improved by introducing extra labour in the preceding picking stage and in the subsequent loading stage to guarantee a smooth workflow in the intermediate packing stage.

Insert Table 3 about here

Estimation of optimal bias

Although some bias may improve efficiency, excessive bias can obstruct efficiency (Sanders and Graman, 2009). Therefore, a relevant question for warehouse management is what level of bias leads to optimal efficiency. We analyse this by investigating the (non-linear) relationship between forecast bias and efficiency of each activity. We use 180 instead of 132 daily observations by allowing for a somewhat wider range of bias. We exclude only 15 observations with positive bias above 1, meaning that the forecast is more than twice the actual order size (the mean bias of these 15 observations is 2.3). Such forecast errors arise if big customers cancel orders or if the warehouse has an ICT system collapse. The efficiency of picking (denoted by EPick) is related as follows to the forecast bias (denoted by B), where the coefficients are obtained by regression and r denotes the residual:

$$\text{EPick} = 1.109 + 0.168 \times B + 0.182 \times B^2 - 0.564 \times B^3 + r \tag{3}$$

The coefficient of the cubic term is significant (p-value 0.004), whereas higher-order terms are not (the p-value for jointly omitting B^4 and B^5 is 0.178). Within the bias range from -0.5 to +1.0, the above relationship has a local maximum for a bias of approximately 0.5. The associated gain in efficiency compared to unbiased forecasts is approximately 5% (maximum efficiency is 1.17 for bias 0.45 compared to 1.11 for bias 0). A rather wide bias range leads to similar efficiencies (the estimated efficiency is at least 1.16 for biases between 0.26 and 0.59). Since the data information is rather limited, the precise optimal value is uncertain, and an approximate 95% confidence interval for the optimal bias runs from 0.3 to 0.6.

We obtained comparable results for loading and total outbound activities with the same approach. The 95% confidence interval for optimal bias runs from 0.4 to 0.7 in both cases. The efficiency gain is approximately 10% for loading (maximum efficiency is 1.33 for bias 0.48 compared to 1.20 for bias 0) and 5% for total outbound activities (maximum 1.08 for bias 0.49 compared to 1.02 for bias 0). These outcomes confirm the previously obtained support for Hypothesis 3* in Table 3 that a certain level of over-forecasting is beneficial for labour efficiency of picking, loading and total outbound procedures. Bias has no significant direct (linear or non-linear) effect on packing efficiency. As was discussed before, packing is the most labour-intensive stage and is affected primarily by the preceding stage of picking and the subsequent stage of loading.

Simulation of optimal bias

We extend the above empirical analysis with a small-scale simulation study. The simulation model consists of three consecutive multi-server queuing blocks for the outbound process. The waiting space for each block is unlimited, and a server comprises a team of four workers for packing and of individual workers for picking and loading. Table 4 shows empirical warehouse data on daily order size, hourly peak order size and labour productivity. We assume fixed service rates at each queuing block, corresponding to stationary working speeds. The random part of the process consists of the arrival of pallets at the first queuing block for picking. These arrivals are assumed to follow exponential distributions with hourly varying mean based on historical data of the warehouse. After arrival at the picking stage, pallets go through the consecutive stages of picking, packing and loading. For each stage, throughput consists of the number of handled pallets and depends on the number of workers.

If arrivals were non-random and evenly spread over all hours, the results for this theoretical operation mode would show an overall outbound labour efficiency of 98.0% (see Table 4). In practice, arrivals are random, and the warehouse data show an hourly peak order of 13% on average per day at the packing stage (125.5 instead of 110.7 pallets). The standard labour plan in Table 4 accommodates for this hourly peak and allocates labour based on unbiased forecasts. Based on 1,000 simulation runs, overall labour efficiency is 77.7% on average. We use the same set of simulation runs to compute labour efficiency for a range of alternative labour plans. Table 4 shows the labour allocation that maximises average overall outbound labour efficiency. By allocating one extra worker both to the picking and to the loading stage, the resulting higher productivity (82.7% - 76.6% = 6.1 percentage points) of the intermediate (labour-intensive) packing stage compensates for lower productivity at the picking and loading stages, as the average overall outbound labour efficiency increases by 4.1 percentage points to 81.8%. This optimal plan corresponds to forecast biases of 11.1% for picking and 50.0% for loading. This simulation result is roughly in line with the previously discussed empirical study results indicating optimal biases of 30%-70%.

The simulation illustrates the beneficial effect of positive forecast bias on labour productivity as postulated in Hypothesis 3*. The quantitative results differ somewhat from those of the empirical study, which is due to several simplifying assumptions. In practice, arrival rates are not exponentially distributed and labour service rates are not constant over time. The simulation model ignores details of the (around fifty) sub-activities of the outbound process and the effects of furlough and overtime policies. Because of the limited number of available data, the simulation model is kept simple and is meant only as illustration.

Insert Table 4 about here

Survey outcomes on forecast bias and labour efficiency

We conduct a survey to supplement the case study results of a single warehouse. As we were primarily interested in sensitive information on forecast bias and labour productivity, we selected warehouses for which we knew how to find the right person in charge of manpower planning who has this exclusive information. Another selection criterion was that the warehouses should be comparable with respect to other relevant aspects, such as handled products, floor space and labour situation. We approached 34 warehouses, thirty of which participated in the survey. The thirty warehouses are located in ten European countries: France, Germany, Great Britain, Greece, Lithuania, the Netherlands, Poland, Portugal, Slovakia and Sweden. The warehouses deal with consumer electronics products of various manufacturers and deliver to retailers (22), to end users (3), or to retail or repair shops (5). We supported participation by providing Likert-scale answer options to overcome possible reluctance in providing sensitive information. The two main questions were related to the level of bias they usually applied in labour planning (11 options, ranging from 'below -40%' to 'above +40%') and to their average productivity measured as the ratio of required over actually available labour (six options, from 'less than 80%' to 'over 120%'). Other survey questions were related to warehouse conditions.

The warehouses are divided into three groups according to their bias strategy: 22 employ a positive bias strategy (use more workers than required), four have negative bias strategy (use fewer workers than required), and four do not employ bias. These three groups of warehouses are similar with respect to all considered warehouse conditions: handled products, floor space, labour contract flexibility, consignee type, planning flexibility, order fluctuation level, forecast frequency, shift size, shift cost, job complexity (number of stages), labour takt time and number of shippers. ANOVA tests for equal means in the three bias groups are all insignificant at the 5% level, except for shift size that is similar for warehouses with positive and negative bias, but is somewhat larger for the group without bias (p-value 0.04). Average labour productivity differs significantly among the three groups (p-value 0.02), with the highest productivity for warehouses with positive bias, followed by those without bias, and with the lowest productivity for warehouses with negative bias. The four warehouses of the latter type reported productivities of 80%-90% (2), 91%-100% (1), and 101%-110% (1), whereas the 22 warehouses with a positive bias strategy reported productivities of 80%-90% (1), 91%-100%

(4), 101%-110% (10), 111%-120% (5) and over 120% (2).

These survey outcomes support the case study results. As the survey warehouses are similar to the one of the case study, the predominance of positive bias in labour planning supports Hypothesis 1, and the beneficial effect for labour productivity supports Hypothesis 3*. The survey also indicates that many European warehouses already implement bias strategies to improve labour efficiency. However, as detailed information on sensitive aspects such as bias strategies and labour productivity at the disaggregated level of individual operations is unavailable from the survey, the case study provided a unique opportunity to study these mechanisms at actual floor level.

Implications

Forecast bias methodology for labour planning

The efficiency of warehouse operations largely depends on labour costs. Overall efficiency is high if sequential stages of the warehouse process are synchronised so that each stage receives a smooth stream of tasks from previous stages. This requires flexible and accurate labour planning. We investigated three research hypotheses related to forecast bias and labour productivity. First, management forecasts display systematic bias related to cost considerations. Second, integrating management forecasts in statistical models supports intentional management forecast bias. And third, intentional forecast bias derived from operational warehouse data improves labour efficiency.

We proposed a predictive analytic methodology to integrate management forecasts and statistical forecasts (Hypothesis 2) and to obtain the optimal level of forecast bias (Hypothesis 3). The required operational information consists of management forecasts, actual order sizes and labour productivity at various stages of the warehouse process. Our proposed strategy to optimise warehouse labour efficiency consists of three steps. First, maintain a database containing management demand forecasts and actual order sizes. Second, measure labour productivity at the level of individual warehouse activities and workers. Third, determine the predictive analytic relationship between demand forecast bias and productivity and optimise labour capacity planning accordingly for the sequential stages of the warehouse process.

This methodology was illustrated with a case study, and we now summarise the main results. Management forecasts of order sizes are systematically upward biased, particularly in busy periods. As it is more expensive for our case study warehouse to solve labour shortages than to dismiss excess workers before the end of their shift, this systematic over-forecasting is in line with the asymmetric cost structure for this warehouse. This finding supports our first hypothesis. The real-time ex ante statistical forecasts that integrate expert forecasts provide significant improvements by reducing bias, improving forecast quality and reducing absolute prediction errors. This finding supports our second hypothesis. The bias can be managed by correcting for recently observed biases in management forecasts. Compared to these forecasts, the root mean squared prediction error is reduced by 18% on average and by 35% for busy periods. Real-time ex ante forecasts are only slightly inferior to ex post forecasts, which provide a benchmark that is unachievable in real-time. The combination of expert forecasts and statistical information provides better sales forecasts than can be obtained from either separately. Our findings further show that over-forecasting of required labour leads to higher labour efficiency of picking, loading and overall outbound procedures. Allocating more labour during the preliminary picking stage and during the final loading stage reduces waiting times and guarantees a smooth workflow for the labour-intensive intermediate packing stage. These findings support our third research hypothesis. Optimal efficiency of picking, loading and outbound labour is obtained by a positive forecast bias of roughly 30%-70%, including systemic bias from warehouse managers. Compared to unbiased forecasts, these biases lead to efficiency gains of approximately 10% for loading and 5% for picking and for the total outbound process. A small-scale survey among thirty warehouses confirms that over-forecasting generally improves labour efficiency, and this result is also confirmed in a simple simulation study.

The case study company has incorporated these results in their evaluation and redesign of their interrelated management strategies for demand forecasting and labour planning. It acknowledges the importance of investing more labour during picking and loading to support the packing stage. The company rewards its workers periodically by individual or team bonuses to sustain higher efficiency and flexibility among workers.

Implementation aspects

Implementing our methodology for warehouse labour planning involves two predictive analytic relations, that is, a demand forecast model and a labour productivity model, as summarised in Figure 1. The forecast model integrates expert judgement and historical demand data, and managers can decide what type of expert judgements are relevant for their situation and how to incorporate them. The respective weights of the various forecast sources can be determined empirically, for example, by means of forecast combination methods (Aiolfi *et al.*, 2011). In the case study example, we apply such methods to determine the weights of management forecasts and historical demand. Various other strategies can be employed, e.g. using historical demand data to produce a benchmark forecast and adjusting the outcome by expert judgement. The productivity model relates labour efficiency to forecast bias based on historical labour productivity data for each stage of the warehouse process. This relationship depends on warehouse characteristics, including prevailing cost structures and labour hiring options. Managers can develop forecast bias strategies depending on their situation, and our advice is to analyse historical patterns of productivity and forecast bias.

The advantage of the above two predictive analytic steps is that their implementation is flexible and can be tuned directly to the warehouse situation. It should be noted, however, that the resulting demand forecast and labour planning strategies will be case dependent, as the relative weight of expert judgement and the amount of bias are determined empirically. Such an empirical approach provides only approximations of reality and may not represent the real nature of the process in its full extent, which is a common shortcoming of empirical research.

Future research and study limitations

Our main finding is that some controlled amount of bias improves overall efficiency of warehouse procedures. The specific results on optimal bias and associated efficiency gains obtained for our case will be different for other periods and other warehouses. By following similar methodologies as described in this paper, warehouse managers can determine the level of forecast bias that works best for their situation. The business analytic information required for this evidence-based labour management consists of available hiring strategies and cost structures as well as historical data on order sizes, forecasts and labour productivity. Such an implementation requires integrating information flows from various warehouse departments and provides an example of the potential benefits of the rapidly increasing interest for big data and business analytics (Waller and Fawcett, 2013; Wang *et al*., 2016). The case study illustrates the methodology, and the results are confirmed in a small-scale survey among thirty warehouses and in a simple simulation study.

Because our methodology follows an empirical approach, the investigation of its benefits for other warehouse situations is an important topic for future research. More in general, supply chain management may benefit from further empirical case studies on the use of systematically collected warehouse data to support evidence-based management strategies.

References

Aiolfi, M., Capistran, C. and Timmerman, A. (2011), “Forecast combinations”. Chapter 12 in Clements, M.P. and Hendry, D.F. (Eds.), *The Oxford Handbook of Economic Forecasting*, Oxford University Press, Oxford, pp. 355-390.

Bartholdi, J. J. and Hackman, S. T. (2016). *Warehouse & Distribution Science: Release 0.97,* The Supply Chain and Logistics Institute, Georgia Institute of Technology, Atlanta, Georgia, www.warehouse-science.com.

Bond, J. (2012), "Labor management systems: The (very near) future of LMS", available at: http://www.mmh.com/article/labor_management_systems_the_very_near_future_of_lms (accessed May 19, 2017).

Bowersox, D.J., Closs, D.J. and Cooper, M.B. (2002), *Supply Chain Logistics Management*, McGraw Hill, New York.

Box, G.E.P., Jenkins, G.H. and Reinsel, G.C. (1994). *Time Series Analysis: Forecasting and Control* (3rd ed), Prentice Hall, Englewood Cliffs, New Jersey.

Breusch, T.S. (1978), “Testing for autocorrelation in dynamic linear models”, *Australian Economic Papers*, Vol. 17 No. 31, pp. 334-355.

Brusco, M.J. and Johns, T.R. (1995), “The effect of demand characteristics on labour scheduling methods”, *International Journal of Operations & Production Management*, Vol. 15 No. 1, pp. 74-88.

De Koster, R., Le-Duc, T. and Roodbergen, K.J. (2007), “Design and control of warehouse order picking: a literature review”, *European Journal of Operational Research*, Vol. 182 No. 2, pp. 481-501.

De Leeuw, S. and Wiers, V.C.S. (2015), “Warehouse manpower planning strategies in times of financial crisis: evidence from logistics service providers and retailers in the Netherlands”, *Production Planning & Control*, Vol. 26 No. 4, pp. 328-337.

Godfrey, L.G. (1978), “Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables”, *Econometrica*, Vol. 46 No. 6, pp. 1293-1301.

Gong, Y. and De Koster, R.B.M. (2011), "A review on stochastic models and analysis of warehouse operations", *Logistics Research,* Vol. 3 No. 4, pp. 191-205.

Goodwin, P. (1996), "Statistical correction of judgmental point forecasts and decisions", *Omega,* Vol. 24 No. 5, pp. 551-559.

Goodwin, P. (2002), "Integrating management judgement and statistical methods to improve short-term forecasts", *Omega,* Vol. 30 No. 2, pp. 127-135.

Gu, J., Goetschalckx, M. and McGinnis, L.F. (2010), "Research on warehouse design and performance evaluation: A comprehensive review", *European Journal of Operational Research,* Vol. 203 No. 3, pp. 539-549.

Hendry, D.F. (1995), *Dynamic Econometrics*, Oxford University Press, Oxford.

Jarque, C.M. and Bera, A.K. (1987), "A test for normality of observations and regression residuals", *International Statistical Review*, Vol. 55 No. 2, pp. 163-172.

Riley, M. and Lockwood, A. (1997), "Strategies and measurement for workforce flexibility: an application of functional flexibility in a service setting", *International Journal of Operations & Production Management*, Vol. 17 No. 4, pp. 413-419.

Ritzman, L.P. and King, B.E. (1993), "The relative significance of forecasting errors in multistage manufacturing", *Journal of Operations Management*, Vol. 11 No. 1, pp. 51-65.

Ruben, R.A. and Jacobs, F.R. (1999), "Batch construction heuristics and storage assignment strategies for walk/ride and pick systems", *Management Science*, Vol. 45 No. 4, pp. 575-596.

Sanders, N.R. and Graman, G.A. (2009), "Quantifying costs of forecast errors: A case study of the warehouse environment", *Omega*, Vol. 37 No. 1, pp. 116-125.

Sanders, N.R. and Manrodt, K.B. (2003), "The efficacy of using judgmental versus quantitative forecasting methods in practice", *Omega*, Vol. 31 No. 6, pp. 511-522.

Sanders, N.R. and Ritzman, L.P. (1991), "On knowing when to switch from quantitative to judgemental forecasts", *International Journal of Operations & Production Management*, Vol. 11 No. 6, pp. 27-37.

Sanders, N.R. and Ritzman, L.P. (2004), “Integrating judgmental and quantitative forecasts: methodologies for pooling marketing and operations information”, *International Journal of Operations & Production Management*, Vol. 24 No. 5, pp. 514-529.

Van den Berg, J.P. (2007). *Integral Warehouse Management,* Management Outlook Publications, Utrecht, Netherlands.

Van Gils, T., Ramaekers, K., Caris, A. and Cools, M. (2017), “The use of time series forecasting in zone order picking systems to predict order pickers’ workload”, *International Journal of Production Research*, Vol. 55 No. 21, pp. 6380-6393.

Wacker, J.G. and Lummus, R.R. (2002), “Sales forecasting for strategic resource planning”, *International Journal of Operations & Production Management*, Vol. 22 No. 9, pp. 1014-1031.

Waller, M. A. and Fawcett, S. E. (2013), “Click here for a data scientist: Big data, predictive analytics, and theory development in the era of a maker movement supply chain”, *Journal of Business Logistics*, Vol. 34 No. 4, pp. 249-252.

Wang, G, Gunasekaran, A., Ngia, E.W.T. and Papadopoulos, T. (2016), “Big data analytics in logistics and supply chain management: Certain investigations for research and applications”, *International Journal Production Economics*, Vol. 176 No. C, pp. 98-110.

Table 1: Comparison of manager forecasts with model-based forecasts (2009 week 46 - 2012 week 48)

	Week			
	All	End month	End year	End month year
Sample size	159	37	42	10
Mean value				
Actual order size	122.03	176.97	142.98	213.32
Forecast management	126.08	192.09	152.15	238.80
Forecast ex ante models	124.36	179.91	144.38	214.86
Forecast ex post model	122.40	178.50	143.62	215.68
Prediction bias (forecast minus actual)				
Forecast management	4.05	15.12	9.17	25.48
Forecast ex ante models	2.33	2.94	1.41	1.54
Forecast ex post model	0.37	1.53	0.64	2.36
Mean absolute prediction error				
Forecast management	14.66	19.64	20.49	35.60
Forecast ex ante models	13.15	15.06	14.67	21.99
Forecast ex post model	12.74	14.07	15.46	23.60
Standard deviation				
Actual order size	53.49	43.75	61.39	33.80
Forecast error management	19.65	23.75	26.24	38.45
Forecast error ex ante models	16.35	18.14	18.50	26.07
Forecast error ex post model	16.17	17.16	19.37	27.06
Root mean squared prediction error				
Forecast management	20.06	28.15	27.80	46.12
Forecast ex ante models	16.52	18.38	18.55	26.12
Forecast ex post model	16.17	17.23	19.38	27.16
Prediction error comparison tests				
F-test variance EMAN vs EMOD_a	0.17	0.29	0.06	0.26
t-test EMAN vs EMOD_a	0.05	0.00	0.00	0.00
t-test EMOD_a vs EMOD_p	0.00	0.07	0.21	0.36
E-test MAN vs MOD_a	0.09 / 0.00	0.95 / 0.00	0.51 / 0.00	0.38 / 0.01
E-test MOD_a vs MOD_p	0.08 / 0.01	0.94 / 0.05	0.06 / 0.93	0.45 / 0.98
Absolute prediction error comparison tests				
F-test variance EMAN vs EMOD_a	0.06	0.01	0.11	0.09
t-test EMAN vs EMOD_a	0.05	0.06	0.02	0.09
t-test EMOD_a vs EMOD_p	0.16	0.15	0.15	0.24
W-test EMAN vs EMOD_a	0.26	0.15	0.04	0.14
W-test EMOD_a vs EMOD_p	0.10	0.09	0.31	0.36

Table notes

* We consider 159 weeks for which ex ante model forecasts are available (8 initial weeks are lost).
* The order size and all forecasts and forecast errors are expressed in terms of 1000 boxes per week.
* The forecast errors are denoted by EMAN for the manager, EMOD_a for the ex ante (real-time) models that vary per week, and EMOD_p for the ex post model that is estimated using data for all weeks.
* The tests show p-values (underlined if at most 0.05) for the following tests: Levene's F-test for equal variance (2-sided), paired samples t-test for mean (1-sided), Wilcoxon signed rank W-test (1-sided), and two encompassing E-tests (2-sided) for forecasts A vs B, with test equation O = c + dA + (1-d)B, with O actual order size; first test is for B encompasses A (d=0), second for A encompasses B (d=1).

Table 2: Daily distribution of weekly orders and manager forecasts

	Year	Sample	Day				
			Monday	Tuesday	Wednesday	Thursday	Friday
Orders	2008-2011	--	18.0	18.0	19.0	22.0	23.0
Orders	2012	195	25.4	20.8	19.0	23.7	11.0
Forecast	2012	195	18.5	20.2	20.4	27.2	13.7
Forecast error	2012	195	-27.2	-2.9	7.4	14.8	24.5
Bias > 1/2	2012	34	2	1	4	7	20
Bias < -1/3	2012	29	16	2	3	1	7

Table notes

* Daily distribution for 2008-2011 is obtained through interviews with warehouse managers.
* The 195 daily observations are for week 1 to 40 of 2012 (200 days, excluding 5 bank holidays).
* The first three rows of the table show daily shares (in percentages).
* The row "Forecast error" shows the percentage relative mean forecast error, that is, 100*(Forecast - Orders) / Orders.
* The rows "Bias > 1/2" and "Bias < -1/3" show the number of days with such large bias.

Table 3: Effect of forecasting bias on labour efficiency (daily data from week 1 to week 40, 2012)

				Mean labour efficiency per activity			
Bias situation	Sample	Bias interval	Bias	Pick	Pack	Load	Out
Comparison of means							
All	195	All	0.190	1.123	0.954	1.220	1.034
Negative	95	Below 0	-0.254	1.103	0.957	1.147	1.016
Positive	100	Above 0	0.611	1.142	0.953	1.292	1.049
Difference (%)	95+100			3.5	-0.4	12.6	3.2
Equal means	95+100			0.023	0.946	0.000	0.050
Non-aberrant	132	-1/3 to +1/2	0.006	1.110	0.950	1.209	1.023
Negative	66	-1/3 to 0	-0.180	1.087	0.949	1.140	1.004
Positive	66	0 to 1/2	0.192	1.133	0.951	1.277	1.041
Difference (%)	66			4.2	0.2	12.0	3.7
Equal means	66+66			0.023	0.946	0.003	0.061
Comparison of ranks							
Negative	95	Below 0	-0.254	90.1	99.2	81.7	91.0
Positive	100	Above 0	0.611	105.5	96.8	113.5	104.7
Equal ranks	95+100			0.028	0.883	0.000	0.044

Table notes

* Forecast bias is defined as Bias = (Forecast - Order)/Order, where Forecast = manager forecast and Order = actual order size.
* To exclude aberrant forecasts, the data are limited to 132 days where the ratios Forecast / Order and Order / Forecast are both at most 1.5, that is, with Bias between -1/3 and +1/2; in this way, 63 of the 196 observations are lost (34 with Bias > 1/2, mean 1.43, and 29 with Bias < -1/3, mean -0.42).
* Labour efficiency is defined as the ratio of actually required labour over hired labour (all measured per day), so that the efficiency is positive (negative) if productivity is above (below) the value 1.
* The rows "Difference (%) " show the percentage difference: 100*(Positive - Negative)/Negative.
* The rows "Equal means" show the (one-sided) p-value for the t-test that the mean efficiency is larger in the Positive than in the Negative bias group (equal variances not assumed).
* The row "Equal ranks" shows the (one-sided) p-value of the Wilcoxon rank sum test.
* The column "Bias" shows the mean bias.
* The last four columns show the mean (rank) efficiency for four activities and (tests for) the difference in efficiency between the negative and positive bias groups.

Table 4: Simulation results for the effect of excess labour on efficiency

		Unit	Formula	Pick	Pack	Load	Out
Theoretical operation (non-random arrivals, no peaks)							
(1)	Order size per day (15 labour hours)	Pallet	Empirical	1,661	1,661	1,661	1,661
(2)	Average order size per labour hour	Pallet	(1)/15	110.7	110.7	110.7	
(3)	Maximum pallets per hour per worker	Pallet	Empirical	13.0	2.0	60.5	
(4)	Required labour hours	Hour	(1)/(3)	128	829	27	985
(5)	Required workers per shift (7.5 hours)	Worker	(2)/(3)	9	56	2	67
(6)	Labour efficiency	%	100*(4)/(15*(5))	94.8	98.7	91.4	98.0
Standard operation (random arrivals, peaks, no bias)							
(7)	Peak order size per hour	Pallet	Empirical	110.7	125.5	110.7	
(8)	Allocated labour size per shift	Worker	(7)/(3)	9	64	2	75
(9)	Actual labour hours per day	Hour	15*(8)	135	960	30	1,125
(10)	Simulated average throughput per day	Pallet	Simulation	1,488	1,473	1,470	
(11)	Labour efficiency	%	100*((10)/(3))/(9)	85.0	76.6	80.9	77.7
Optimal operation (random arrivals, peaks, optimal bias)							
(12)	Peak order size per hour	Pallet	Empirical	110.7	125.5	110.7	
(13)	Allocated labour size per shift	Worker	Optimised	10	64	3	77
(14)	Actual labour hours per day	Hour	15*(13)	150	960	45	1,155
(15)	Simulated average throughput per day	Pallet	Simulation	1,613	1,590	1,588	
(16)	Labour efficiency	%	100*((15)/(3))/(14)	82.9	82.7	58.3	81.8
(17)	Labour bias	%	100*((13)-(8))/(8)	11.1	0.0	50.0	2.7

Table notes

* Labour days consist of two shifts, each with 7.5 working hours.
* Peak order size per hour in row (7) is derived from empirical data as the average hourly peak load per day.
* Labour efficiency in rows (11) and (16) is defined as the ratio of needed labour hours over actual labour hours.
* Labour efficiency of 'Out' is the weighted average of productivity of the three tasks, with labour shares as weights.
* Labour bias is defined as the extra allocated labour compared to standard operation without bias.
* Each packing lane requires four workers, so that planned labour size per shift for packing is a multiple of four.
* The column 'Formula' shows how rows are computed from previous ones; the (rounded) data in rows (1), (3), (7), and (12) are based on empirical data, and the averages in rows (10) and (15) are obtained from 1000 simulations.

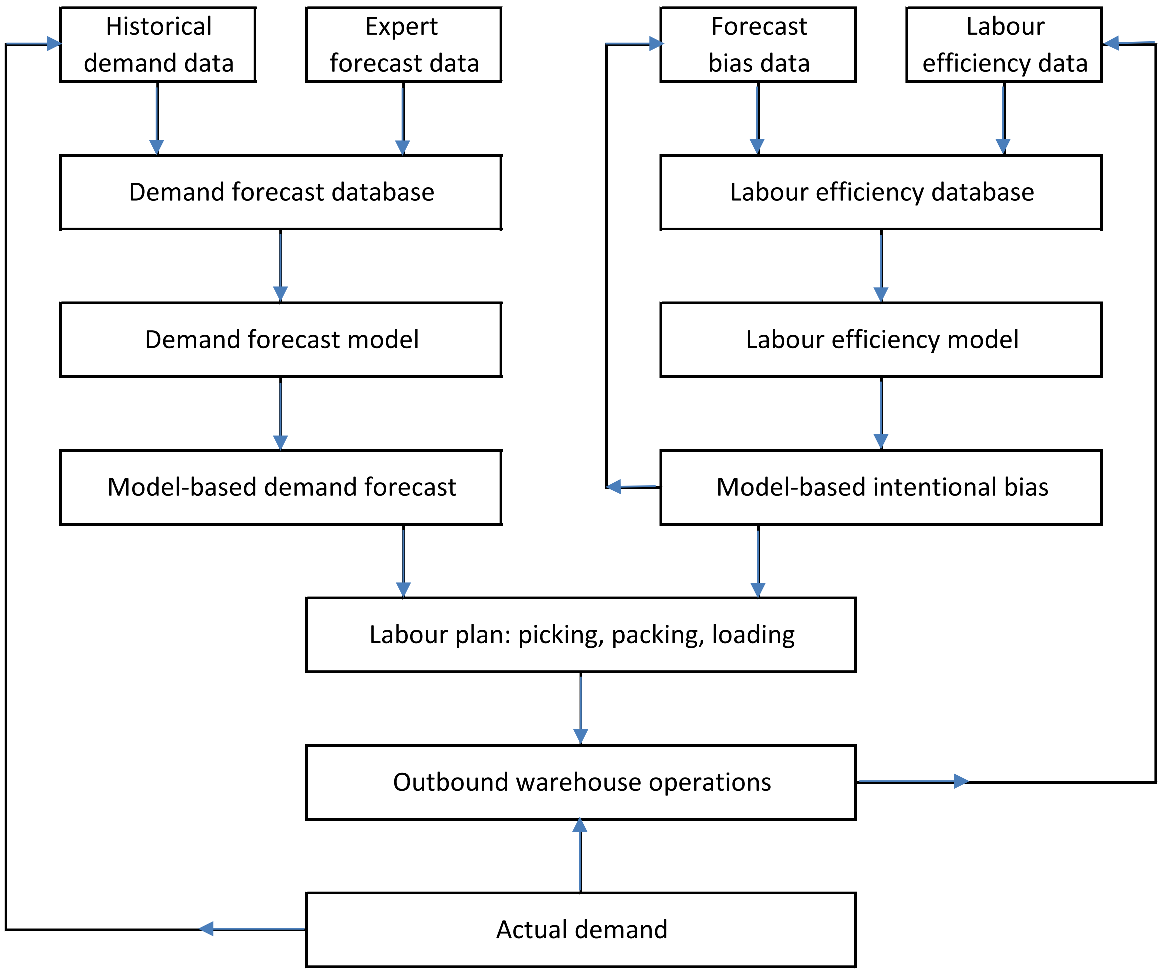
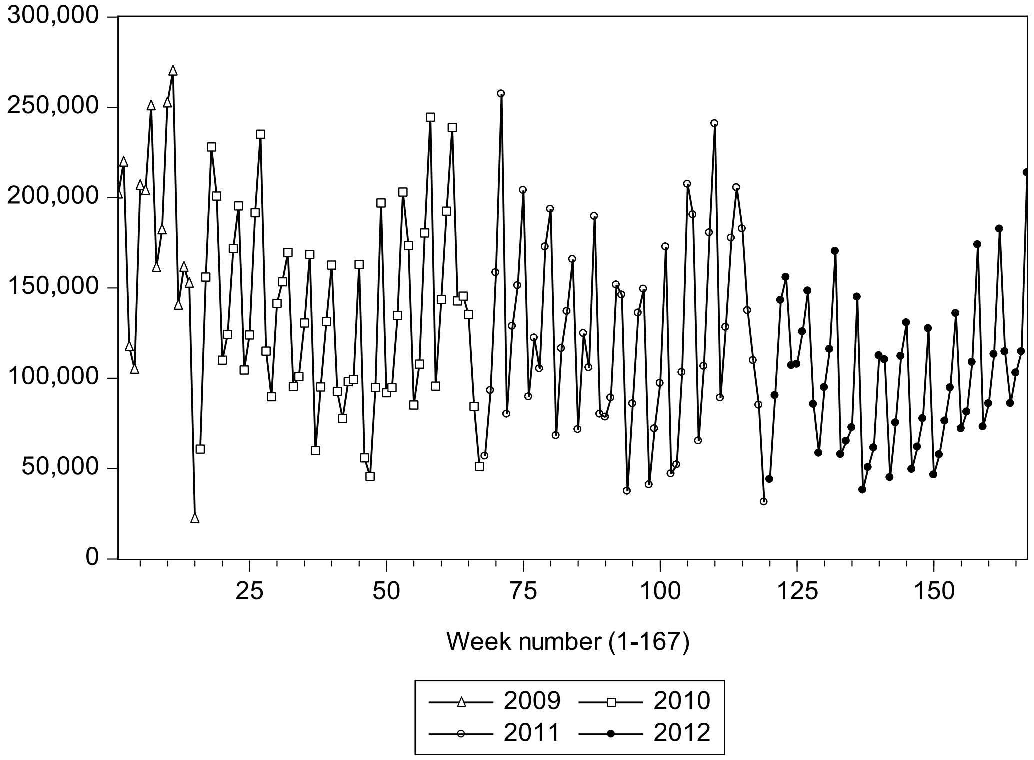


Figure 1: Flow diagram of warehouse data, models, activities, and management.



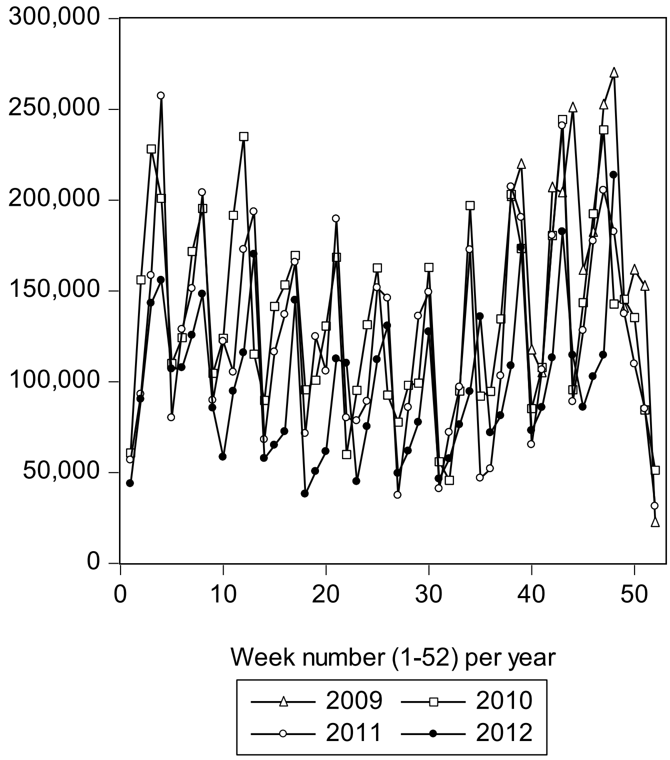


Figure 2: Time series of weekly order size, measured on the vertical axis as the number of boxes, from week 38 of 2009 to week 48 of 2012 (top) and after split-up per calendar year (bottom).