The impact of forecasting errors on warehouse labor efficiency: A case study in consumer electronics

Thai Young Kim[a], Rommert Dekker[b] and Christiaan Heij[c]

Abstract

Efficiency of outbound warehouse operations depends on the management of demand forecasts and associated labor planning. A case study in consumer electronics shows that warehouse management systematically over-forecasts actual orders, by 3% on average and by 6-12% in busy periods (at the end of each month and also in the months September, October, and November). A time series model that corrects order forecasts for the biases in preceding weeks reduces the bias to less than 2%, both on average and also in busy periods. The arrangements with the labor provider imply potential benefits of intentional over-forecasting and the associated ample labor supply for the warehouse. As compared to under-forecasted days, labor productivity on over-forecasted days is higher by 12% for loading activities and by 4% for picking and total outbound activities. Similar productivity gains are found if unbiased forecasts are compared with the optimal bias obtained from non-linear models estimated from daily data on bias and labor efficiency. The positive effects of intentional over-forecasting on productivity are confirmed in a structural equations model. By following similar methodologies as described in this paper, warehouse managers can determine the amount of intentional forecast bias that works best for their situation. The information required for this kind of evidence-based labor management consists of historical data on order sizes, forecasts, and labor productivity, and the outcomes depend on the available hiring strategies and cost structures.

Keywords

Decision support, warehouse planning, forecasting, labor efficiency, case study, time series

[a] Samsung Electronics Europe Logistics; email: thaiyoung@gmail.com
[b] Econometric Institute, Erasmus University Rotterdam; email: rdekker@ese.eur.nl
[c] Corresponding author, Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, Netherlands; phone: +31-10-4081264; fax: +31-10-4089162; email: heij@ese.eur.nl

1 Introduction

Warehousing serves as the primary link between producers and customers in a supply chain. It provides buffering for manufacturing operations to manage varying customer demand (Bowersox et al., 2002). Warehouse operations are highly dependent on labor management, and many companies run third-party logistics (3PL) warehouses. Because warehouse labor planning is based on order forecasts with associated uncertainties, demand fluctuations result in gaps between planned and required labor. Therefore, operators of 3PL warehouses prefer flexible labor pools to fixed ones to reduce inefficiencies.

Warehouse operations are initiated after receiving an order from the sales department. Labor resources have to be arranged for processing orders quickly, thereby minimizing the lead time to customers. The sales and warehouse functions should meet periodically to align expected labor needs to the order forecasts to reduce the risks of high sunk costs of labor and labor shortages. From a sales perspective, order forecasts contain unavoidable errors because of changes in requested delivery dates from customers. Therefore, it may not be easy to translate order forecasts into accurate labor resource planning. The work of Ali et al. (2012) and Trapero et al. (2012) clarified the benefits of sharing information among the various parties involved. If a warehouse succeeds in obtaining a better understanding of the nature of the forecast errors, it may help improve the efficiency of hiring strategies.

We will investigate warehouse labor efficiency by a case study conducted at Samsung Electronics. We consider the following two questions. First, what is the quantitative nature of the errors in delivery order forecasting? Second, in which way does forecast bias, defined as the ratio of the forecast error over the actual order size, influence labor efficiency (the ratio of the actually required labor over hired labor)? The first question is analyzed using statistical methods, employing weekly time series data on managers' forecasts and actual orders. For the second question, we construct a structural equation model from daily labor productivity data and order forecast biases. Productivity is measured at the three stages of the outbound operations of warehouses: picking, packing, and loading. We aim to assist warehouse managers in labor resource planning by showing how forecasting biases permeate the labor efficiency of sequential procedures.

Although the specific details of our findings may be case-dependent, the analysis and general conclusions are relevant broadly for warehouse labor management, forecasting, and human factor operations management. By following similar methodologies as described in this paper, warehouse managers can determine the amount of forecast bias that works best for their situation. The information required for this kind of evidence-based labor management consists of historical data on order sizes, forecasts, and labor productivity, and the outcomes depend on the available hiring strategies and cost structures. We were able to perform our case study analysis because the company under investigation has a consistent and active attitude in collecting and storing the required operational performance data.

The remainder of this paper is structured as follows. Section 2 provides a brief literature review and states our research hypotheses. Section 3 describes the case study environment. Section 4 presents the statistical analysis of forecast errors, and Section 5 investigates the relationship between forecast bias and labor efficiency at various stages of the warehouse process. Section 6 summarizes the operational implications of our analysis.

2 Literature review and research hypotheses

This paper is concerned with biases in demand forecasting within a warehouse environment. Sanders & Graman (2009) provide simulation evidence that properly managed forecast biases may reduce costs, provided the bias is limited to prevent penalties and excessive stock holding costs. Benefits can be obtained if forecasts are biased in the least costly direction. Systematic under-forecasting is beneficial if labor costs dominate, and over-forecasting is beneficial if the main costs are customer-imposed delay penalties for discontinuity of service. Ritzman & King (1993) underscore the relevance of forecast bias for inventories in multi-stage manufacturing. Buffers are helpful in creating flexibility in labor and capacity utilization. They warn for undesirable biases that originate from optimistic sales projections and misguided attempts at inventory reduction. Pirttila & Hautaniemi (1995) analyze labor efficiency in terms of activity-based costing (ABC) for warehouse operations. They conclude that ABC information can be fragmentary and not as accurately recorded as production capacity. Studies by Schefczyk (1993) and Hamdan & Rogers (2008) suggest that the efficiency of 3PL warehouse operations is inversely related to warehouse size, contradicting expected economies of scale. Trapero et al. (2012) find that sharing of upstream sales information between retailers and manufacturers reduces costs. The benefit initially accrues to the manufacturers because of improved inventory policies, and retailers can subsequently profit from lower purchasing prices.

In our study, we focus on various stages of the outbound warehouse process, and we do not consider the inbound process and associated inventory strategies. We investigate whether forecast bias can increase labor efficiency if slack in capacity improves the transition of pallets among the various warehousing processes: picking, packing, and loading. Our main interest is on downstream effects of managers' forecast bias on warehouse operations, and we evaluate operational benefits in terms of labor efficiency. Our statistical analysis involves time series regressions and structural equation models. Box & Jenkins (1976) introduced a forecasting-oriented time series methodology, including historical data only of the variable of interest, supplemented by seasonal factors for yearly, monthly, weekly, and other cycles. Because of its simplicity, this forecast methodology has found widespread application in business and other fields. An early example of its use in labor productivity forecasting was demonstrated in airplane manufacturing by Masud (1985). Because warehouse orders are unstable, we extend the forecast model by including relevant factors besides past order sizes; in particular, we add the information on managers' forecasts and forecast bias.

The above brief review of related literature and our focus on the relation between forecast bias and labor productivity leads us to formulate the following four research hypotheses. First, expert forecasts of managers display systematic bias related to cost considerations. Second, statistical models assist in reducing forecast bias. Third, labor efficiency is affected asymmetrically by over- and under-forecasting of demand. And fourth, labor efficiencies at various stages of the outbound warehouse process (picking, packing, and loading) are interrelated.

3 Case study environment

We examine the Western European warehouse of Samsung Electronics. Distribution is a labor-intensive operation containing short-term orders for delivery. Labor constitutes more than 40 percent of the total warehouse costs. The outbound warehouse process comprises the consecutive stages of picking, packing, and loading. The packing process is the most labor intensive one, with an average handling of two-to-three pallets per man-hour. Picking and loading show larger productivity fluctuations than packing, and overall efficiency depends, to a large extent, on managing these two procedures.

The handled products comprise finished goods in consumer electronics. These are fast-moving items with a total inventory volume of less than two weeks of demand. The warehouse contains 250,000 pallet storage places, including racking and bulk storage, with a total space of 50,000 square meters.

The sales department provides weekly forecasts of order size, and the logistics department translates these forecasts twice per day and in a mechanical way into labor hiring decisions. Flexible labor pools of about 60 full time workers for each of two shifts (from 06:00 to 15:00 and from 15:00 to 23:00) are provided on a daily basis to the warehouse by a 3PL provider. The warehouse is charged a tariff based on standard activities needed to process orders rather than a tariff of man-hours consumed. This is profitable for the warehouse, because it carries less risk than managing a fixed labor force. However, accurate demand forecasts are required for limiting risks for the labor provider. Otherwise, unused labor capacity leads to low labor efficiency in the short term, with the risk of higher activity tariffs in the long term. Every Monday morning, a labor plan for the week is prepared, based on the weekly volume forecast. As actual labor needs may differ from plan, the labor provider and warehouse management meet daily for fine-tuning the man-hours required for the next half day. The labor provider permits the warehouse to furlough some of the workers without payment in case they have worked for more than three hours and the remaining workload does not justify hiring them for the remainder of the shift. On the other hand, if labor is insufficient in the previous shift, an impromptu demand for extra workers may be created up to four hours before a warehouse shift starts. However, such a last-minute demand for workers may be satisfied by hiring novices who are less productive. Furthermore, inaccurate planning may frustrate experienced workers.

Warehouse handling volume is measured by the number of boxes handled. Figure 1 shows weekly order sizes for week 38 of 2009 until 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 spread of demand across months and weeks within the month is shown in Figure 2. The order size tends to be the largest at the end of the month and at the end of the year. The September–November peak is typical for consumer electronics. The peak at the end of the month is related to the behavior of retailers, for example, to meet sales targets. Figure 2 also shows the managers' average forecast per month and per week. These forecasts systematically overestimate actual orders, particularly, at the end of the month and at the end of the year. This upward forecast bias can be explained by the responsibility of sales department for timely delivery to customers, whereas the risk of excess labor costs rests with the logistics department.

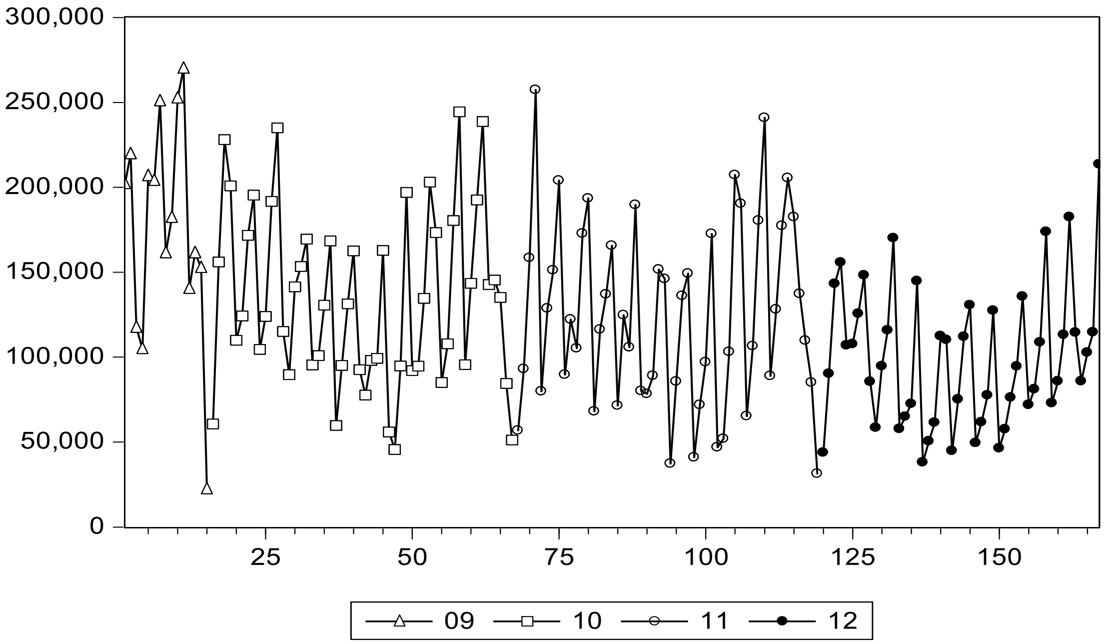

Figure 1.a: Time series of weekly order size, from week 38 of 2009 until week 48 of 2012.

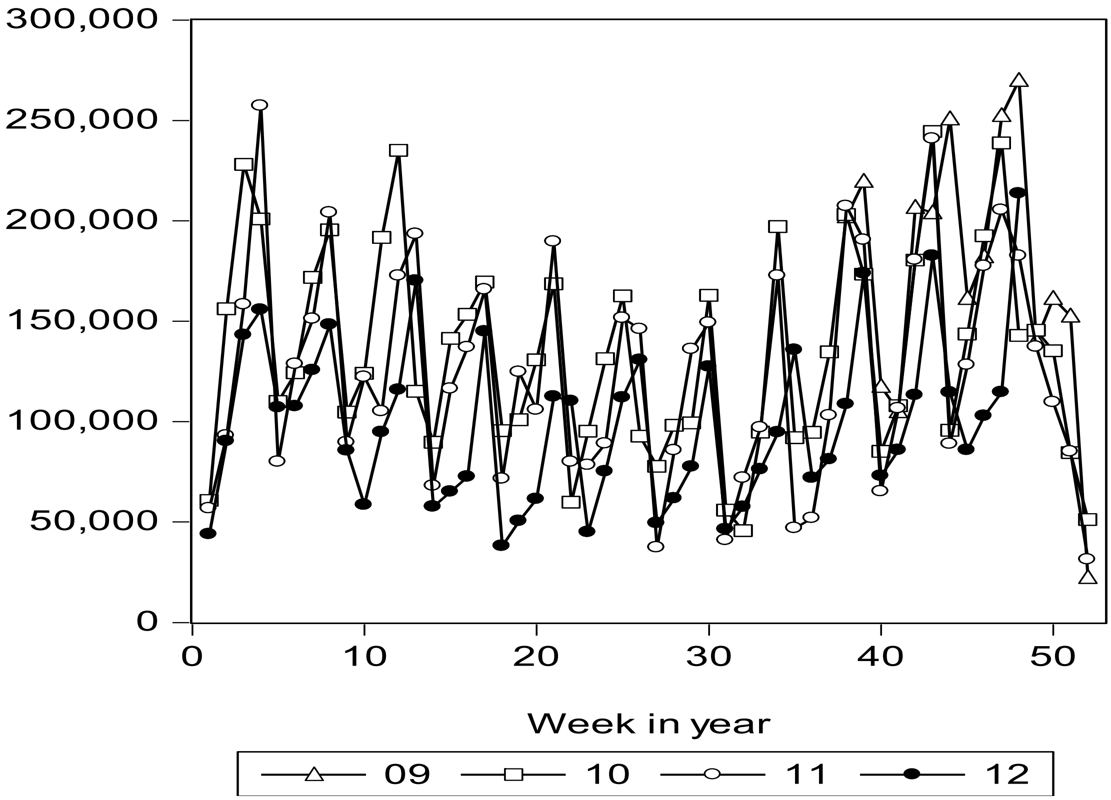

Figure 1.b: Four time series of weekly order size per calendar year, for weeks 38-52 of 2009, weeks 1-52 of 2010 and 2011, and weeks 1-48 of 2012.

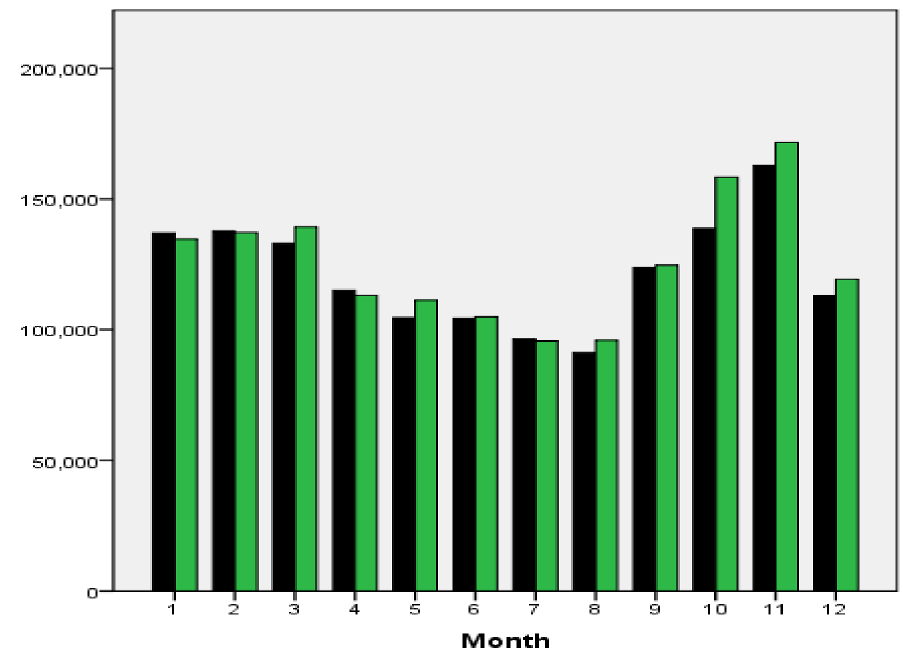

Figure 2.a: Mean value (per month of the year) of actual weekly order size (left bars) and weekly manager forecast (right bars); sample period runs from week 38 of 2009 until week 48 of 2012, and the sample size (number of observed weeks per month of the year) ranges from 12 to 17.

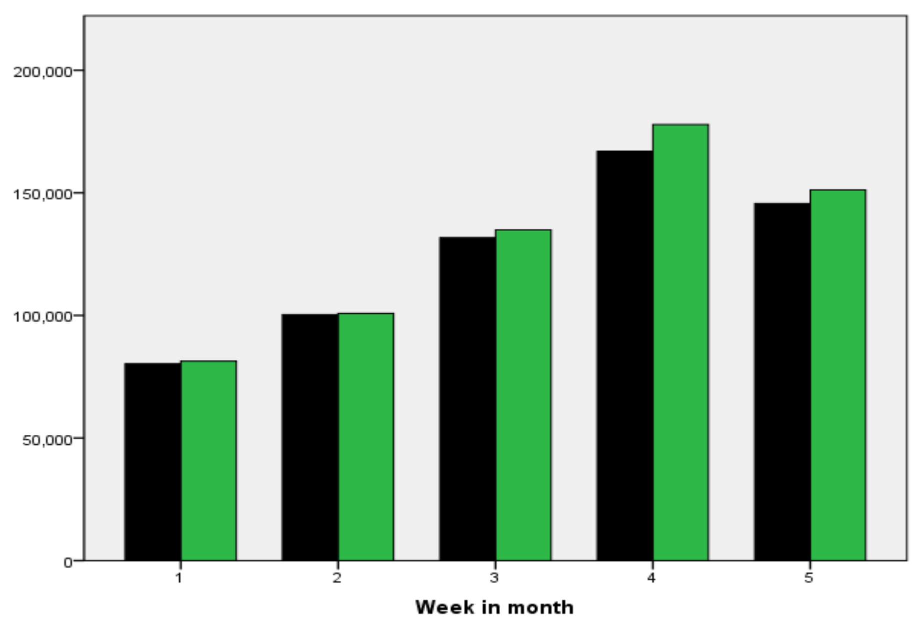

Figure 2.b: Mean value (per week of the month) of actual weekly order size (left bars) and weekly manager forecast (right bars); sample period runs from week 38 of 2009 until week 48 of 2012, and the sample size (number of observations per week of the month) is 38 or 39 for weeks 1 to 4 and 13 for week 5.

4 Forecasting order size

4.1 Statistical model

Reliability of warehouse planning crucially depends on the accuracy of order forecasts. Goodwin (2002) reports that forecasts are mostly based on expert judgment, which often contains systematic bias, possibly due to asymmetric costs of under- and over-forecasting (see also Sanders & Graman (2009)). In such a situation, planning strategies may benefit from combining judgment-based forecasts with statistical ones.

We analyze time series of weekly managerial forecasts provided by the sales department and actual orders processed by the logistics department. The uncertainty in realized order sizes causes potential unbalances between planned and required labor input. If the order sizes can be forecasted more accurately, management would be better able to reduce sunk labor costs. The costs of inaccurate labor planning are asymmetric because of a risk-taking arrangement with the labor provider.

Weekly data are available from week 38 of 2009 until week 48 of 2012 (see Figures 1 and 2), with a total of 167 observations on manager forecasts (denoted by F) and actual orders (denoted by O), both of which are measured in terms of number of boxes. The weekly forecasts are retrieved from the sales forecasting database and the actual orders from the warehouse processing IT system. For week t, the forecast F(t) of the management system is confirmed on Monday morning of week t, and the actual delivery order O(t) is confirmed at the next Thursday's cut-off (as orders placed subsequently are carried over to the next week). We use these data to build a statistical model to enable warehouse management to improve forecasts of order sizes and associated labor requirements by combining expert judgement with historical data.

Some seasonal factors can be useful for reducing uncertainty in order size. Indeed, as seen in Figures 1 and 2, the order size tends to be relatively large at the end of the month and end of the year. We include an end-of-the-month indicator M(t) (with a value of 1 for the entire last week of the month and 0 otherwise) and an end-of-the-year indicator Y(t) (with a value of 1 for weeks in September, October, and November, and 0 otherwise) as possible calendar effects. Also of interest is the forecast accuracy of previous weeks. If, for example, the managers' forecast for the previous week underestimates the actual order size, then a similar bias may apply for the current week. Therefore, the forecast error $E(t) = O(t) - F(t)$ may have predictive power for future orders.

To obtain a statistical forecast model, we applied a "general-to-specific" specification procedure (Hendry, 1995). The starting point is a relatively rich model including information up to four weeks ago, that is, with values of the actual order sizes for the previous four weeks and with values of the manager's forecasts for the upcoming and the previous four weeks. Next, we simplified it by testing for various parameter restrictions and we applied diagnostic tests (on the absence of serial correlation and on normality of the model residuals) for the resulting simplified model. Finally, we considered inclusion of additional calendar effects for end-of-the-month and end-of-the-year effects.[1] The forecast model that is obtained by this procedure is

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

Here, r(t) denotes the residuals of the model, which contain no significant serial correlation (the Breusch-Godfrey LM-test for serial correlation of one-to-four lags has p-value 0.43). The distribution of residuals is not far from normal (the Jarque-Bera test for normality has p-value 0.06), although the residuals are somewhat positively skewed (with value 0.44). This means that large, positive residuals (up to a maximum of approximately 40% of the average order size) occur somewhat more often than large, negative residuals (down to a minimum of approximately 30% of the average order size). Put differently, the most sizeable errors of the statistical model correspond more often to under-prediction than to over-prediction of the

---

[1] Table A-1 shows the results for the initial model (Model 1, with four lags of manager's forecasts and delivery orders) and for two restricted versions. The results for Model 1 in Table A-1 show that four lags suffice, whereas using less lags is rejected. Further, the simplest model, Model 3, is not rejected against the alternative of Model 1, and therefore Model 3 is preferred.

actual order size. Additional calendar effects in the above model were insignificant (p-values 0.72 for end-of-the-year effects, 0.42 for end-of-the-month effects, and 0.72 for both effects jointly; allowing for distributed end-of-the-month effects does not provide significant results either, with p-values 0.24 for lag 1, 0.08 for lag -1, and 0.29 for lags -1, 0, and 1 jointly). These results may be explained by the fact that the manager's forecast, which is included as an explanatory factor, already includes these calendar effects.

The obtained model can be interpreted in terms of bias correction (see Goodwin, 2002). The coefficient 0.861 of F(t) means that approximately 86% of the manager's forecast for the coming week is taken as the expected order, adding approximately 20% of the statistical forecast bias from the previous week and from four weeks ago (which in most cases has the same position within the month as the upcoming week). If previous statistical forecasts were too low (with error $E = O - F > 0$), then the current forecast is corrected upward, and if they were too high, then it is corrected downward.

Before discussing the quality of the statistical forecasts, we highlight that the above model is obtained ex-post, using all available data. Therefore, these forecasts are not made in real time, as they employ future data that were used to obtain the numerical values of the coefficients. Real-time statistical forecasts are obtained if, for each week t, the model is estimated using only the data that would be available at the beginning of week t (i.e., forecasts F for times up to and including t, and order sizes O for times 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), obtaining different coefficients for every week t.[2] The comparison of the manager's forecasts with the ex-ante forecasts is a fair one, as both methods use compatible information sets of past historical data at each forecast origin, that is, each week a new forecast is made.

[2] This approach could be refined further by re-doing the model specification search at each forecast origin. For simplicity, we keep the model structure fixed over time and re-estimate the coefficients at each forecast origin. A disadvantage of our procedure may be that this model is over-parameterized in initial periods when few data are available, so that the quality of the presented ex ante forecasts could possibly be improved somewhat further by allowing for time-varying model specifications.

One might expect that ex-post forecasts are of a better quality than ex-ante forecasts because the latter employ less information.

4.2 Forecast comparison

The empirical results are summarized in Table 1. These results are for four periods: all weeks (159, as eight initial weeks are lost when using the ex-ante forecasts), and three busy periods, end-of-the-month (37 weeks), September through November (42 weeks), and end-of-the-month weeks in these three months (10 weeks). The actual order size is relatively large during busy periods. The manager's forecasts are consistently biased upwards, and the relative bias increases with order size. [3] The manager's forecast is larger than the actual order size in 56% of all weeks, 58% of end-of-the-year weeks, and even 72% in end-of-the-month weeks. For the ex-ante model, these percentages are 52%, 50%, and 53%, respectively, showing a better balance between over- and under-forecasting. The ex-ante forecasts have a much smaller bias and standard deviation, and they perform only slightly worse than ex-post forecasts. The average manager forecast bias is 3% on average and 6-12% in busy periods (8.5% for end of the month, 6.4% for the end of the year, and 11.9% for end of the month weeks in September, October, and November). The ex-ante forecast bias is less than 2% on average, also in busy periods (1.7% for end of the month, 1.0% for the end of the year, and 0.7% for end of the month weeks in September, October, and November). As measured by the root mean squared prediction error, which combines bias and variance, the error reduces from 16.4% to 13.5% on average (from 16% to 10% for end-of-the-month, from 19% to 13% for end-of-the-year, and from 22% to 12% for end-of-the-month weeks in September, October, and November). The ex-post forecasts are only slightly better.

[3] Measured as percentage or order size, the average manager forecast bias is 100×4.05/122.03 = 3.3 for all weeks, 100×9.17/142.98 = 6.4 at the end of the year, 100×15.12/176.97 = 8.5 at the end of the month, and 100×25.48/213.32 = 11.9 at the end of the month weeks in September, October, and November. The corresponding percentage average ex-ante forecast biases are respectively 1.9, 1.0, 1.7, and 0.7.

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

* Considered are 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).

The lower part of Table 1 shows the outcomes of various forecast comparison tests. The ex-ante forecasts have a significantly smaller bias (at the 5% level) than the manager's 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 encompassing test evaluates whether the statistical forecasts provide significant additional information as compared to the manager's judgment (Hendry, 1995; see also Trapero et al., 2012). This test is based on the following regression equation:

$$O(t) = \alpha + \beta\times F_A(t) + (1 - \beta)\times F_B(t) + \varepsilon(t) \qquad (2)$$

Here, O is the order size, F_A and F_B are the forecasts of methods A and B, and ε is the 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 for $0 < \beta < 1$. Table 1 shows that the ex-ante forecasts encompass the manager's forecasts in all four cases (all weeks and the three busy sub-periods), and the manager's forecast never encompasses the ex-ante forecasts. This means that the ex-ante forecasts are more reliable than the manager's forecasts and that the ex-ante forecasts cannot be further improved by taking a weighted combination average with the manager's forecast. 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 the forecasts in terms of absolute prediction errors, as an alternative to bias where large positive and negative errors cancel out. The ex-ante statistical forecasts have a smaller mean absolute prediction error than the manager's forecasts (reduction from 12% to 11% for all weeks, from 11% to 9% for end-of-the-month, from 14% to 10% for end-of-the-year, and from 17% to 10% for end-of-the-month weeks in end-of-the-year months). This difference is significant for all weeks and for end-of-the-year weeks.

4.3 Summary of results

The manager's forecasts of order sizes are systematically biased upward. As it is more costly for the warehouse under consideration to solve labor shortages than to dismiss excess workers before the end of their shift, the systematic over-forecasting is therefore in line with this asymmetric cost structure for the warehouse. This finding confirms our first research hypothesis that expert forecasts display systematic bias related to cost considerations. Further, the ex-ante statistical forecasts provide significant improvements by reducing bias, improving forecast quality, and reducing absolute prediction errors. This finding confirms our second research hypothesis that statistical models assist in reducing forecast bias. The real-time ex-ante forecasts are only slightly inferior to ex-post forecasts, which provide a benchmark that is unachievable in real-time.

5 Labor efficiency and forecast bias

5.1 Effects of positive and negative forecast bias

Even though accurate demand forecasts are useful for improving efficiency, some amount of bias may be profitable for operational reasons. For example, if the supply process comprises a sequence of inter-related tasks, then overall efficiency may be improved by introducing some slack in each stage of the process. We provide an empirical analysis of the effects of forecasting bias on labor efficiency for each of the three stages of the supply process. We also consider the combined efficiency of these stages, that is, the full outbound process. We define labor efficiency as the ratio of actually required labor over hired labor, so that an efficiency above (below) 1 corresponds to labor productivity being higher (lower) than standard. Further, a positive (negative) forecast bias corresponds to over-forecasting (under-forecasting) by the manager. We investigate the research hypothesis that labor efficiency is affected differently on days with a positive forecast bias than on days with a negative forecast bias. Further, we attempt to identify the optimal amount of bias, on average, for the considered period.

Table 2: Two hypothetical scenarios for over-forecasting and under-forecasting

	Forecasting situation (Forecast is 100)		
	Accurate	Over	Under
Actual order size	100	80	120
Scenario A: Fixed labor			
Actual labor units spent	100	100	100
Labor efficiency (%)	100	80	120
Scenario B: Flexible labor			
Actual labor units spent	100	70	140
Labor efficiency (%)	100	114	86

Table notes

* Labor and order size are standardized so that one unit of order size is processed by one unit of labor during normal operation.
* Labor efficiency is defined as the ratio of actual order size over actual labor units spent.
* In scenario A with fixed labor, workers adjust their working rate to actual needs.
* In scenario B with flexible labor, over-forecasting leads to 30% furlough, and under-forecasting leads to 40% extra labor units spent due to expensive or inefficient extra labor time.

Table 2 shows two hypothetical scenarios for workers' reactions to over- and under-forecasting. In scenario A, workers have a good overview of the total workload. They are dedicated to finish the extra work needed in case of under-forecasting the actual workload, whereas they slow down their working speed in case of over-forecasting to fill up their work shift time. Therefore, in scenario A, labor efficiency benefits from under-forecasting. In scenario B, work is done in segmented stages, and workers in flexible labor pools do not have an overview of the total workload. In case of under-forecasting, they will not adjust their working speed so that extra, last-minute labor is hired, which is relatively expensive due to overtime compensation and lower productivity. In case of over-forecasting, workers are not able to reduce their working speed strategically, their work speed is even higher than normal due to smooth transition of work through the various processing stages, and work will finish early so that workers can be dismissed before the end of their shift. Therefore, in scenario B, labor efficiency benefits from over-forecasting. The opposite results for scenarios A and B show that the effect of forecast bias on labor productivity depends on how labor is organized.

We will investigate for the warehouse under study whether labor efficiency is improved by under- or over-forecasting. Every week, the warehouse receives an order forecast for that week from the sales department. This weekly forecast is split into daily forecasts, based on historical spreads over the days and on operational information like order cancellation notifications and postponed orders of previous days. At the start of each day, the resulting order forecast for that day is translated into the labor plan for the first half and second half of the day, and the latter plan can still be updated during the day. These half-daily plans involve a trade-off between delivery service reliability and labor efficiency, Labor efficiency is measured continually and stored on a daily basis in the warehouse IT system. Actual task durations are measured (in seconds) by time clocking systems for each of the activities of picking, packing, and loading. For each of these activities, the warehouse employs standard durations based on about fifty sub-tasks. The labor efficiency of each activity is automatically registered in the IT system on a daily basis, by comparing the clocking system data with the standard durations.

Data on labor efficiency are available on a daily basis for the first 40 weeks of 2012. The total number of observations is 195 (40 weeks of 5 working days, excluding 5 bank holidays). Actual order sizes are also available on a daily basis for this period, as well as the daily order forecasts that the warehouse managers derived from the weekly order forecasts of the sales managers. Table 3 shows the distribution of actual order sizes and the manager's forecasts over the five working days of the week. The forecasts are considerably biased downward for Mondays and upward for the end of the week. One possible cause of these biases is a shifting demand pattern over the week as compared to previous years. The table shows approximate daily shares 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 where the ratios of Forecast over Order and that of Order over Forecast are both at most 1.5.

Table 3: Daily shares 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

* The daily shares for 2008-2011 are a rule-of-thumb obtained by interviewing operational experts.
* 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.

In other words, the sample is restricted to days where the forecast bias lies between -1/3 and +1/2. This eliminates 63 days (of which 18 are Mondays and 27 are Fridays), leaving 132 days for our analysis.[4]

Table 4 shows summary statistics for the comparison of labor efficiency on days with negative and positive forecast bias. The average efficiency is the largest 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, which is close to 1). The mean daily efficiency of the picking, loading, and overall outbound procedures is significantly higher for days with a positive forecast bias than for days with a negative forecast bias.

[4]Allowing for somewhat wider bias intervals leads to qualitatively similar results.

Table 4: Effect of forecasting bias on labor efficiency (daily data from week 1 to week 40, 2012)

				Mean labor 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

* The forecasting bias is defined as Bias = (Forecast - Order) / Order, where Forecast is the manager forecast and Order is the 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).
* Labor efficiency is defined as the ratio of actually required labor over hired labor (all measured per day), so that the efficiency is positive (negative) if productivity lies 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.

As compared to days with a 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, eliminating aberrant observations, are very similar. The table also reports the outcomes of rank comparison tests for days with negative and positive bias. These tests are not sensitive to outliers in efficiency, and the results confirm those of the conventional mean comparison t-tests described earlier. Negative forecast bias, that is, under-estimating actual labor needs, results in loss of efficiency. This is because too much overload will lead to costly overtime and long waiting times between the various stages of the warehouse process. Moreover, if workers foresee that the workload is too large to be finished in time, it may tempt them to reduce their working speed to profit from higher overtime wages.

5.2 Structural equations model

More detailed insight regarding the relationship between forecast bias and labor efficiency of warehouse activities can be obtained by constructing a structural equations model. The variables of interest consist of four labor efficiency variables ('Pick' for picking efficiency, 'Pack" for packing efficiency, 'Load' for loading efficiency, and 'Out' for the efficiency of combined outbound handling), forecast bias (defined as the ratio (Forecast - Order)/Order), and seasonal factors (day of the week, week of the month, and the holiday season, July and August, to account for possible variations in the experience of employed labor). Daily data on these variables are available from week 1 to 40 of 2012, and aberrant days with relatively large forecast errors are excluded (with a bias larger than 1/2 or smaller than -1/3), resulting in a sample size of 132 days. We follow a "general-to-specific" procedure, starting with a general model, including all variables that are available, followed by model simplification by omitting insignificant links among the variables. We refer to Bollen (1989) and Raftery (1993) for technical details on the estimation and selection of structural equation models. The resulting

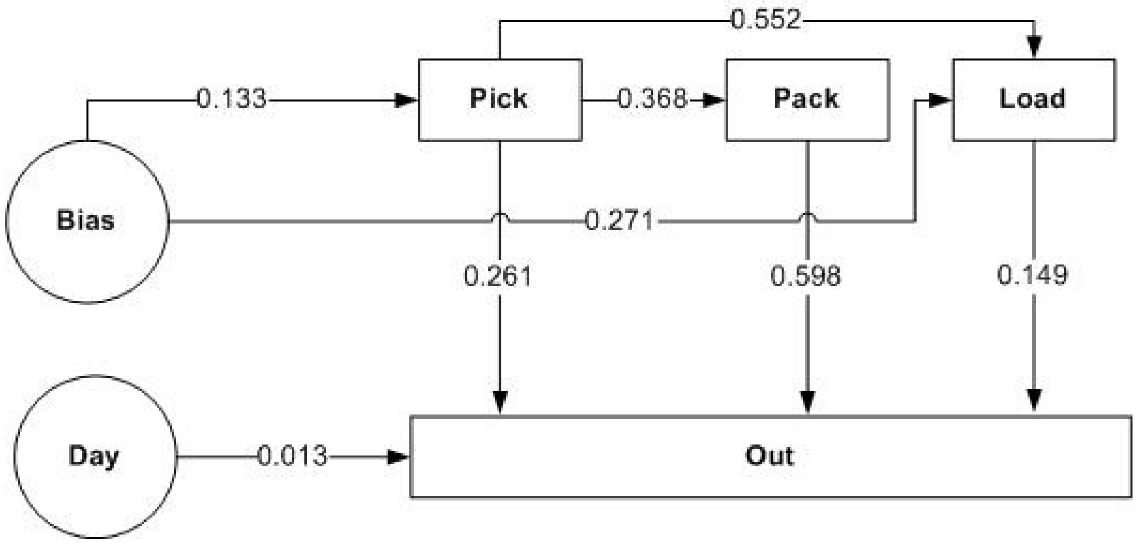

Figure 3: Structural equations model for the relations between forecast bias and labor efficiency. The input factors are the day of the week ('Day') and the forecast bias ('Bias'), and the output factors are the efficiency of the warehouse activities picking ('Pick'), packing ('Pack'), loading ('Load'), and the overall outbound process ('Out') comprised of these three activities. Definitions of bias and labor efficiency are given in Table 4. The graph depicts the relations among the variables that are statistically significant (at the 5% level); for example, Load = 0.552 × Pick + 0.271 × Bias shows that the efficiency of loading increases for higher efficiency of picking and for more positive forecast bias.

model is shown in Figure 3.[5] Of the seasonal factors, only the day of the week has a significant effect on total outbound activities, with higher efficiency for the latter part of the week. Forecast bias has positive effects on the efficiency of the activities of picking (p-value 0.008) and loading (p-value 0.008). The labor efficiency in the first stage of the outbound process, picking, has important effects on the subsequent stages of packing and loading, and also, both directly and indirectly, on total outbound labor efficiency. The efficiency of packing and loading also influences total outbound efficiency; however, there is no direct relationship between packing

[5] We used the AMOS module of SPSS for model selection. Depending on the selection of links between the variables, there are in total 65,536 candidate models that are ranked by means of the Bayes information criterion. This ranking delivers four models of comparable quality, in the sense that their BIC values are the smallest and differ by less than 2 from the top ranked model. We choose this top ranked model (with minimal BIC). The other three models differ only marginally from this model and provide qualitatively very similar results. Table A-2 provides more details of the model shown in Figure 3.

and loading. If the direct and indirect effects are combined, the total effect of picking efficiency on total outbound efficiency is 0.563.[6]

Of special interest are the significant positive direct effects of forecast bias on the efficiency of picking and loading activities, with associated positive indirect effects on the efficiency of packing and total outbound activities. These results support our earlier findings presented in Table 4. Over-forecasting the required labor pool will guarantee a large share of experienced labor and prevent the risk of having to hire less-experienced workers.

5.3 Optimal bias estimation

Because of the limited number of available observations, the relationships of the models considered before are all linear, which is an evident simplification. Although a reasonable amount of positive bias, that is, over-forecasting labor needs, may improve efficiency, there will be limits beyond which excessive bias will obstruct efficiency. Therefore, a relevant question for warehouse management is which amount of bias leads to optimal efficiency. We analyze this by investigating possible non-linear relationships between forecast bias and the efficiency of each activity separately. We use 180 instead of 132 daily observations by allowing for a somewhat wider range of bias. We exclude only 15 observations that have very large positive bias (larger than 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 can arise if a big customer cancels an order or if the warehouse suffers from 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 e denotes the residual:

$$\text{EPick} = 1.109 + 0.168 \times B + 0.182 \times B^2 - 0.564 \times B^3 + e \qquad (3)$$

The coefficient of the cubic term is significant (p-value: 0.004), whereas higher-order terms

[6] This value is obtained from 0.261 + 0.368×0.598 + 0.552×0.149 = 0.563. Table A-2 contains all direct, indirect, and total effects between the variables of Figure 3.

are not (the p-value for jointly omitting the fourth- and fifth-order terms, B^4 and B^5, has a p-value of 0.178).[7] 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 as compared to unbiased forecasts is approximately 5% (the maximal efficiency is 1.17 for bias 0.45 as 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. An approximate 95% confidence interval for the optimal bias runs from 0.3 to 0.6.[8]

With the same approach, we obtained similar results for loading and total outbound activities. 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 (the maximal efficiency is 1.33 for bias 0.48 as compared to 1.20 for bias 0) and approximately 5% for total outbound activities (maximum 1.08 for bias 0.49 as compared to 1.02 for bias 0). These outcomes provide additional support to our earlier conclusion that some amount of over-forecasting is beneficial for labor efficiency of the activities of picking, loading, and total outbound procedures. Bias has no significant direct (linear or non-linear) effect on packing efficiency. Packing is the most labor-intensive stage and is affected primarily by the preceding stage of picking. The efficiency of packing workers is determined mainly by that of the picking workers, as they depend on a smooth supply of pallets offered from the picking stage. For the warehouse under consideration,

[7] Details of this regression are in Table A-3, together with the results for the efficiency of loading and total outbound activities. Bias has no significant (linear or non-linear) effect on packing efficiency.

[8] More precisely, the interval runs from 0.34 to 0.57. The local maximum of the cubic relation is a non-linear function of the regression coefficients, which under standard regression assumptions are approximately normally distributed with mean and variance-covariance matrix computed from the data. The 95% confidence interval of the bias where the relation takes its local maximum is obtained by simulating the estimated normal distribution of the regression coefficients. We performed one million simulation draws of this normal distribution and computed the associated location of the local maximum. For some configurations of the coefficients, the curve has no local maximum, and we skipped such draws. This occurred only in 0.06% of the simulations.

positive bias in required labor leads to fast processing with associated gains of improved labor efficiency and reduced costs.

5.4 Summary of results

For the warehouse under consideration, over-forecasting of required labor leads to higher labor efficiency of picking, loading, and overall outbound procedures. This can be explained by the importance of smooth transitions between various labor process flows in the warehouse. This finding confirms our third research hypothesis that labor efficiency is affected asymmetrically by over- and under-forecasting of labor demand. Our analysis suggests that optimal efficiency of picking, loading, and outbound labor is obtained for this warehouse by using a positive forecast bias of roughly 30-70 percent. Further, the efficiency of picking activities has considerable effect on the efficiency of subsequent activities of packing and loading, and all these three activities affect the overall outbound efficiency. This finding confirms our fourth research hypothesis that labor efficiencies at various stages of the outbound warehouse process are interrelated.

6 Conclusions

We formulated the following four research hypotheses. First, expert forecasts of managers display systematic biases related to cost considerations. Second, statistical models assist in reducing forecast bias. Third, labor efficiency is affected asymmetrically by over- and under-forecasting of demand. And fourth, labor efficiencies at various stages of the outbound warehouse process are interrelated.

The conclusions for the warehouse of our case study are as follows. First, managers' forecasts are systematically biased upwards, particularly in busy periods. This finding can be explained by the asymmetric cost structure for the warehouse, because it is more practical dismissing workers before the end of their shift rather than experiencing labor shortages. Second, the bias can be reduced considerably by means of statistical forecasts that correct for recently observed biases in the manager's forecasts. The root mean squared prediction error is reduced by 18 percent on average and by 35 percent for busy periods. Third, over-forecasting the need for labor leads to higher labor efficiency of outbound warehouse activities. This finding can be explained by the importance of smooth transitions between various labor process flows in the warehouse. As compared to unbiased forecasts, estimated optimal biases lead to efficiency gains of approximately 10% for loading and of approximately 5% for picking and for the total outbound process. Fourth, some amount of excess labor, especially at early stages of the outbound process, improves the overall efficiency because of reductions in waiting time due to a smoother transition of tasks to next stages. Using the historical data of our study, we calculate that the optimal amount of daily bias lies roughly between 0.3 and 0.7. As labor is hired per stage of the warehouse process, it is best to control the amount of bias per stage.

Our main finding is that some amount of controlled bias improves the 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 amount of forecast bias that works best for their situation. The information required for this evidence-based labor management consists of available hiring strategies and cost structures as well as historical data on order sizes, forecasts, and labor productivity. Supply chain management may profit from further empirical case studies on the use of systematically collected warehouse data to support evidence-based management strategies.

7 References

Ali, M.M., Boylan, J.E., Syntetos, A.A., 2012. Forecast errors and inventory performance under forecast information sharing. International Journal of Forecasting, 28, 830-841.

Bollen, K.A., 1989. Structural equations with latent variables. New York: Wiley.

Bowersox, D.J., Closs, D.J., Cooper, M.B., 2002. Supply chain logistics management. New York: McGraw Hill.

Box, G.E.P., Jenkins, G.M., 1976. Time series analysis: forecasting and control. San Francisco: Holden Day.

Goodwin, P., 2002. Integrating management judgement and statistical methods to improve short-term forecasts. Omega, 30, 127-135.

Hamdan, A., Rogers, K.J., 2008. Evaluating the efficiency of 3PL logistics operations. International Journal of Production Economics, 113, 235-244.

Hendry, D.F., 1995. Dynamic econometrics. Oxford: Oxford University Press.

Masud, A.S.M., 1985. Forecasting plant labor productivity with a time series model. Computers and industrial Engineering, 9, 73-81.

Pirttila, T., Hautaniemi, P., 1995. Activity-based costing and distribution logistics management. International Journal of Production Economics, 41: 327-333.

Raftery, A.E., 1993. Bayesian model selection in structural equation models. In: Bollen, K.A., Long, J.S., (eds.), Testing structural equation models. Beverly Hills: Sage, 163-180.

Ritzman, L.P., King, B.E., 1993. The relative significance of forecasting errors in multistage manufacturing. Journal of Operations Management, 11, 51-65.

Sanders, N.R., Graman, G.A., 2009. Quantifying costs of forecast errors: A case study of the warehouse environment. Omega, 37, 116-125.

Schefczyk, M., 1993. Industrial benchmarking: A case study of performance analysis techniques. International Journal of Production Economics, 32, 1-11.

Trapero, J.R., Kourentzes, N., Fildes, R., 2012. Impact of information exchange on supplier forecasting performance. Omega, 40, 738-747.

Appendix: Detailed estimation outcomes

Table A1: Forecast models (ex post, 2009 week 42 - 2012 week 48)

Model	1			2			3		
Variable	Coeff	Std. Error	t-Stat	Coeff	Std. Error	t-Stat	Coeff	Std. Error	t-Stat
Constant	12.026	5.206	2.310	11.741	4.328	2.713	15.512	3.017	5.141
Order(-1)	0.202	0.081	2.479	0.184	0.077	2.380	0.195	0.067	2.920
Order(-2)	-0.060	0.084	-0.715	--			--		
Order(-3)	0.070	0.084	0.838	--			--		
Order(-4)	0.252	0.081	3.118	0.257	0.076	3.400	0.190	0.068	2.792
Forecast	0.836	0.028	30.415	0.839	0.027	31.631	0.861	0.022	39.859
Forecast(-1)	-0.200	0.072	-2.787	-0.186	0.068	-2.747	-0.195	--	
Forecast(-2)	0.059	0.074	0.799	--			--		
Forecast(-3)	-0.082	0.074	-1.103	--			--		
Forecast(-4)	-0.187	0.072	-2.611	-0.202	0.068	-2.979	-0.190	--	
R-squared	0.914			0.913			0.911		
Log-likelihood	-1808.8			-1809.7			-1811.9		
F and LR tests	0.157	0.135		0.786	0.773		0.443	0.401	

Table notes

* Considered are the 163 weeks for which up to 4 lagged forecasts are available (4 initial weeks are lost).
* The explained variable is order size, expressed in 1000 boxes per week (as are the forecasts).
* "Order" denotes the weekly actual order size, and "Forecast" is the weekly manager forecast.
* Model 2 is a restricted version of Model 1 (removing explanatory factors at lags 2 and 3), and Model 3 is a restricted version of Model 2 (imposing opposite coefficients for Order and Forecast at equal lags).
* F and LR tests show p-values for respectively the F-test and LR-test for comparing the following models: for model 1 against the alternative of the model with 5 lags of Order and Forecast; for model 2 against the alternative of model 1; and for model 3 also against the alternative of model 1.

Table A2: Structural equation model for forecast bias and efficiency (132 daily data, week 1-40, 2012)

Activity	Pick		Pack		Load		Out	
Effect	Coeff	Std. Error	Coeff	Std. Error	Coeff	Std. Error	Coeff	Std. Error
Direct								
Bias	0.133	0.050	--		0.271	0.102	--	
Pick	x		0.368	0.113	0.552	0.173	0.261	0.013
Pack	--		x		--		0.598	0.009
Load	--		--		x		0.149	0.006
Day	--		--		--		0.013	0.005
Indirect								
Bias	--		0.049		0.074		0.116	
Pick	x		--		--		0.302	
Total								
Bias	0.133		0.049		0.345		0.116	
Pick	x		0.368		0.552		0.563	
Pack	--		x		--		0.598	
Load	--		--		x		0.149	
Day	--		--		--		0.013	

Table notes

* The model is estimated using 132 daily data of manager forecasts and labor efficiency (sample period week 1-40 of 2012, 195 observations, of which 63 are lost by excluding aberrant forecasts and restricting to 132 days with Bias between -1/3 and +1/2).
* Each row shows the coefficients corresponding to the effects of the row variable on the column variable, that is, on the efficiency (actually needed labor divided by hired labor) of each activity.
* The direct effects shown in the table are significant at 5% level; all p-values are actually below 0.0005, except for the effects of Bias on Pick (p-value 0.004) and of Bias on Load (p-value 0.020).
* Coefficient cells "--" denote that these links are insignificant and removed in the specified model.
* Coefficient cells "x" denote superfluous cells (as a variable does not affect itself).
* The indirect effects are obtained by multilpying through the relevant direct effects.
* The total effects are obtained as the sum of the corresponding direct and indirect effects.

Table A3: Cubic relations between forecast bias and labor efficiency (180 daily data, week 1-40, 2012)

Activity	Pick			Load			Out		
Variable	Coeff	Std. Error	P-val	Coeff	Std. Error	P-val	Coeff	Std. Error	P-val
Constant	1.109	0.013	0.000	1.200	0.028	0.000	1.019	0.015	0.000
Bias	0.168	0.055	0.003	0.370	0.116	0.002	0.155	0.061	0.012
Bias^2	0.182	0.106	0.088	0.215	0.221	0.333	0.196	0.117	0.097
Bias^3	-0.546	0.187	0.004	-0.846	0.390	0.032	-0.489	0.207	0.019
R-squared	0.053			0.058			0.037		
Bias^4, Bias^5	0.178			0.062			0.633		
Local optimum	0.340	0.451	0.568	0.354	0.476	0.669	0.355	0.485	0.706

Table notes

* The models are estimated using 180 daily data (sample period week 1-40 of 2012, 195 observations, of which 15 are dropped because of excessively large bias: dropped if Bias > 1, mean bias of dropped is 2.3); the dependent variable is labor efficiency (the ratio of actually required labor over hired labor), and the explanatory factors are defined in terms of the forecasting bias, that is, (Forecast - Order) / Order, where Forecast is the manager forecast and Order is the actual order size.
* The row "Bias^4, Bias^5" shows the p-value of the F-test for adding the two terms Bias^4 and Bias^5.
* The row "Local optimum" shows three values; the middle value is the local maximum of the estimated cubic relation, and the first (last) value is the left (right) endpoint of the 95% confidence interval for this local maximum.
* For Pack, the cubic relation is not significant (the p-value for the F-test that the coefficients of the three terms Bias, Bias^2, and Bias^3 are all zero is equal to 0.755, so that Bias has no significant effect on Pack).