MANAGEMENT SCIENCE
Vol. 56, No. 9, September 2010, pp. 1519–1533
ISSN 0025-1909 | EISSN 1526-5501 | 10 | 5609 | 1519

informs

DOI 10.1287/mnsc.1100.1209
© 2010 INFORMS

Do Inventory and Gross Margin Data Improve Sales Forecasts for U.S. Public Retailers?

Saravanan Kesavan
Kenan-Flagler Business School, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, skesavan@unc.edu

Vishal Gaur
Johnson Graduate School of Management, Cornell University, Ithaca, New York 14853, vg77@cornell.edu

Ananth Raman
Harvard Business School, Harvard University, Boston, Massachusetts 02163, araman@hbs.edu

Firm-level sales forecasts for retailers can be improved if we incorporate cost of goods sold, inventory, and gross margin (defined by us as the ratio of sales to cost of goods sold) as three endogenous variables. We construct a simultaneous equations model, estimated using public financial and nonfinancial data, to provide joint forecasts of annual cost of goods sold, inventory, and gross margin for retailers using historical data. We show that sales forecasts from this model are more accurate than consensus forecasts from equity analysts. Further, the residuals from this model for one fiscal year are used to predict retailers for whom the relative advantage of model forecasts over consensus forecasts would be large in the next fiscal year. Our results show that historical inventory and gross margin contain information useful to forecast sales, and that equity analysts do not fully utilize this information in their sales forecasts.

Key words: sales forecasting; retail; inventory; empirical
History: Received October 15, 2007; accepted May 17, 2010, by Aleda Roth, operations and supply chain management. Published online in *Articles in Advance* August 3, 2010.

1. Introduction

The forecast of sales for a firm is a necessary input for its valuation. It is used to project both future earnings and growth rates, which are then used to determine the value of the firm in standard valuation models. Thus, investors and equity analysts place enormous importance in forecasting sales accurately.

In the U.S. retailing industry, which is the focus of our paper, equity analysts conduct a number of activities related to sales forecasting. They visit stores, seek management guidance, watch for leading indicators of sales, and issue their own sales forecasts, which they keep revising as new information becomes available. Many retailers realize the importance placed by investors on forecasting sales and voluntarily release monthly reports—containing information such as same store sales growth rate, year-to-date sales, and plans to open or close stores—to help investors forecast sales. Many third parties, such as comScore Inc., MasterCard SpendingPulse, Retail Metrics Inc., and Thomson Reuters, also cater to the investment community by releasing sales metrics to aid in the forecasting process.

Once the sales forecasts are developed, the stock market uses them in valuation and to track actual sales that get reported in financial statements. It rewards retailers who outperform expectations and penalizes those who fall short of expectations. For example, when Walmart reported its results for the fourth quarter of 2009, it announced a 4.6% increase in sales. Yet, several analysts expressed concerns over Walmart's sales growth being lower than the expected growth rate of 6%, and Walmart's stock price declined (Maestri 2010). The decline in stock price occurred even though Walmart exceeded the analysts' expectations regarding earnings for that quarter. There are many such examples in the business press. Although not conclusive, they suggest that sales forecasts matter, and even small deviations of actual sales from forecasts are associated with large impacts on stock prices. Therefore, it is important to improve the accuracy of sales forecasting.[1]

The goal of our paper is to determine if historical inventory and gross margin data can be applied to improve forecasting of annual sales at the firm

[1] According to some researchers, the stock market reacts differently to revenue and expense surprises. Ertimur et al. (2003) find that investors value a dollar of revenue surprise more highly than a dollar of expense surprise. Swaminathan and Weintrop (1991) document incremental information content of revenues beyond earnings for a sample of companies.

INFORMS holds copyright to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

level in the context of U.S. public retailers. The paper is motivated by the fact that sales, inventory, and gross margin for retailers are related to each other. Retailers commonly use inventory and gross margin to increase sales. Conversely, sales provide input to retailers' decisions on inventory and gross margins. Inventory and gross margin also influence each other because procuring more inventory increases the probability of markdowns, whereas higher gross margin increases the incentive for retailers to carry more inventory. Therefore, the historical values of sales, inventory, and gross margin will be the joint outcome of all these relationships occurring simultaneously. This motivates our research questions: How do we incorporate these simultaneous relationships in a time-series model for sales forecasting? Does the resulting model provide higher forecast accuracy than a benchmark such as the consensus of sales forecasts given by equity analysts?

We use the consensus of equity analysts' forecasts as the benchmark to evaluate our model for several reasons. First, equity analysts have strong incentives to produce accurate forecasts because their compensation and reputation are tied in part to their forecast accuracy. Investors pay millions of dollars every year to purchase analysts' forecasts and stock recommendations (Ivkovic and Jegadeesh 2004). Trade journals such as the *Wall Street Journal* and *Institutional Investor* rank equity analysts based on their track record on metrics including the accuracy of their forecasts. Second, our research parallels a vast number of papers in accounting that study the forecasts of earnings by equity analysts. Past work suggests that equity analysts' consensus earnings forecasts outperform earnings forecasts from time-series models (see Conroy and Harris 1987). Similarly, we compared our model and analysts' consensus sales forecasts against several time-series forecasting models. We found that the model and the consensus were both more accurate than the time-series models. Thus, equity analysts' consensus forecasts should serve as a good benchmark to evaluate forecasts from our model.

Our methodology is based on a simultaneous equations model (SEM) relating cost of sales,[2] inventory, and gross margin for retailers at the firm-year level. We use annual and quarterly data for a large cross-section of U.S. retailers listed on NYSE, AMEX, or NASDAQ for the 1993–2007 fiscal years for our analysis. Besides cost of sales, inventory, and gross margin, we use data for several other variables that serve as curve shifters in the SEM. These include selling, general and administrative expenses, store growth, capital investment per store, index of consumer sentiment, accounts-payable-to-inventory ratio, and lagged values of cost of sales, inventory, and margin. We obtain data for these variables from Standard and Poor's Compustat database, the 10-K filings of firms, and the University of Michigan's Survey Research Center. We generate forecasts on a rolling-horizon basis using lagged data and use observations for 2004–2007 as the test sample. To evaluate forecast accuracy of our model on the test sample, we obtain individual analysts' forecasts of annual sales from the Institutional Brokers Estimate System (I/B/E/S) for the years 2004–2007.

Our paper has two main results. First, we show that the forecasts from our model have lower mean absolute percentage error (MAPE) and median of absolute percentage error than consensus forecasts from equity analysts. We find that the MAPE of model forecasts for our test sample of 230 firm-year data points for 2004–2007 is 4.13%. This is significantly lower ($p < 0.1$) than the MAPE of the consensus forecasts obtained five days after the earnings announcement date (EAD), which is 4.49%. Furthermore, the forecasts of our model have a lower absolute percentage error (APE) than the consensus forecasts for 52.2% of the observations. Here, we note that our model is based only on data from prior years' financial statements, whereas analysts have access to considerably more information. In particular, the EAD of a firm can occur up to 90 days after the fiscal year end (FYE) in our data set. Because the majority of the firms in our data set release monthly sales figures and we only consider analysts' forecasts issued after the EAD, analysts would have information from the first one to two months of the fiscal year they are forecasting for.

The second result is that the residuals from our SEM for a given firm year are predictive of the model's forecast accuracy for that firm for the next year. Using the percentiles of the residuals from the inventory and margin simultaneous equations, we identify four categories of retailers: overinventoried (OI), underinventoried (UI), overpriced (OP), and underpriced (UP). We show that the performance difference between our model forecasts and analysts' forecasts is greater for these categories of retailers than in the general case with all retailers. For example, when we use the top and bottom 10th percentiles of inventory and margin residuals to construct the above categories, we find that one quarter of the observations lie in these categories. The MAPE of the model is 4.67% and the MAPE of the consensus forecasts is 5.59%. The model represents a 16% improvement over the consensus, with the difference in the MAPEs statistically significant at $p < 0.05$. Moreover, the model

[2] Cost of sales is also called cost of goods sold. It is reported in the income statement of a firm. We use it in our analysis as a proxy for sales measured at cost. Thus, we decompose sales revenue into cost of sales and gross margin.

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

yields more accurate forecasts than the consensus for 60.3% of the firms, which is significantly higher than 50% ($p < 0.1$). Because these categories are predicted based upon previous year's residuals, it shows that our model is diagnostic in nature and complements analysts' forecasts.

Our paper builds on the growing literature in empirical research in operations management that uses secondary data. A number of researchers in operations management have used event studies applied to accounting and stock market data to study the impact of operational issues on firm performance (e.g., Hendricks and Singhal 2008; Chen et al. 2005; Corbett et al. 2005; Hendricks and Singhal 2008, 2009). Other researchers have analyzed firm-level inventories through correlation studies between inventory turns and independent variables (e.g., Gaur et al. 2005, Gaur and Kesavan 2009, Rumyantsev and Netessine 2007). Kekre and Srinivasan (1990) use an SEM to study firm-level inventories, product variety, costs, etc. Our forecasting model builds on this prior research by setting up three equations to describe the relationships among variables. We expand the set of explanatory variables in the model to include nonfinancial data from 10-K statements, such as store growth and the number of stores.

The rest of this paper is organized as follows: in §2, we present our research setup, definitions of variables, and data set description; in §3, we present the forecasting model and estimation methodology; in §4, we discuss the methodology for comparing the analysts' forecasts with model forecasts; in §5, we show the results on the comparison of forecast accuracy; and in §6, we discuss the limitations of our study and directions for future work. Technical details and additional sensitivity analysis of our model are presented in the online appendix, which is provided in the e-companion.[3]

2. Research Setup

We define the variables in our model and explain how they are measured in §2.1. Then, we describe the data sets used for constructing and testing the forecasting model in §2.2.

2.1. Definition of Variables

Let i be the index for firms and t be the index for fiscal years. From the Compustat annual data for firm i in fiscal year t, let S_{it} be the total sales (Compustat field DATA12); $COGS_{it}$ be the cost of sales (DATA41); SGA_{it} be the selling, general, and administrative expenses (DATA189); $LIFO_{it}$ be the LIFO reserve (DATA240); and $RENT_{it,1}$, $RENT_{it,2}$, …, $RENT_{it,5}$ be the rental commitments for the next five years (DATA96, DATA164, DATA165, DATA166, and DATA167, respectively). From the Compustat quarterly data for firm i in fiscal year t quarter q, let PPE_{itq} be the net property, plant, and equipment (DATA42); AP_{itq} be the accounts payable (DATA46); and I_{itq} be the ending inventory (DATA38). Additionally, from the 10-K statement of firm i for year t, let N_{it} be the total number of stores open for firm i at the end of year t, and $\widehat{N}_{i,t+1}$ be the projected number of stores to be open at the end of year $t+1$. All these data items are available in the 10-K and 10-Q statements for year t. Our analysis is conducted at the firm-fiscal year unit. Thus, we use year and fiscal year interchangeably throughout the paper.

We make the following adjustments to the above data. The use of FIFO versus LIFO methods for valuing inventory produces an artificial difference in the reported ending inventory and cost of sales. Thus, we add back LIFO reserve to the ending inventory and subtract the annual change in LIFO reserve from the cost of sales to ensure compatibility across observations. The value of PPE could vary depending on the values of capitalized leases and operating leases held by a retailer. We compute the present value of rental commitments for the next five years using $RENT_{it,1}, \ldots, RENT_{it,5}$, and add it to PPE to adjust uniformly for operating leases. We use a discount rate $d = 8\%$ per year for computing the present value, and verify our results with $d = 10\%$ as well.

Using these data and adjustments, we define the following variables for each firm i in year t; q denotes fiscal quarters, $q = 1, \ldots, 4$.

Average cost of sales per store:

$$CS_{it} = [COGS_{it} - LIFO_{it} + LIFO_{i,t-1}]/N_{it}.$$

Average inventory per store:

$$IS_{it} = \left[\tfrac{1}{4}\sum_{q=1}^{4} I_{itq} + LIFO_{it}\right] \Big/ N_{it}.$$

Gross margin:

$$GM_{it} = SR_{it}/[COGS_{it} - LIFO_{it} + LIFO_{i,t-1}].$$

Average SGA per store:

$$SGAS_{it} = SGA_{it}/N_{it}.$$

Average capital investment per store:

$$CAPS_{it} = \left[\frac{1}{4}\sum_{q=1}^{4} PPE_{itq} + \sum_{\tau=1}^{5} \frac{RENT_{it\tau}}{(1+d)^{\tau}}\right] \Big/ N_{it}.$$

Store growth:

$$G_{it} = N_{it}/N_{i,t-1}.$$

[3] An electronic companion to this paper is available as part of the online version that can be found at http://mansci.journal.informs.org/.

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

Projected store growth:

$$\widehat{G}_{it} = \widehat{N}_{it}/N_{i,t-1}.$$

Accounts-payable-to-inventory ratio:

$$PI_{it} = \sum_{q=1}^{4} AP_{itq} \Big/ \left[\sum_{q=1}^{4} I_{itq} + 4LIFO_{it}\right].$$

Also, let ICS_t denote the value of the index of consumer sentiment for the month of December in fiscal year t. We normalize some of the above variables, as shown, by the number of retail stores in order to avoid correlations that could arise due to scale effects caused by increase or decrease in the size of a firm. We compute the logarithm of each variable in order to construct a multiplicative model. The variables obtained after taking logarithm are denoted by respective lowercase letters, i.e., cs_{it}, is_{it}, gm_{it}, $sgas_{it}$, $caps_{it}$, g_{it}, $\hat{g}_{it}$, pi_{it}, and ics_t. Hereafter, we omit the prefix "logged" in the names and notation for variables throughout the paper, unless necessary.

Average cost of sales per store, average inventory per store, and gross margin are the three *endogenous* variables in our study. The rest of the variables are used as *exogenous* variables in the SEM as explained in §3. The forecasts shall be based only on historical (or lagged) values of exogenous variables and coefficients estimated from historical data to avoid any look-ahead.

2.2. Data Description

The Compustat database provides Standard Industry Classification (SIC) codes for all firms, assigned by the U.S. Department of Commerce based on their type of business. The U.S. Department of Commerce includes eight categories, identified by two-digit SIC codes, under retail trade: *lumber and other building materials dealers* (SIC: 52); *general merchandise stores* (SIC: 53); *food stores* (SIC: 54); *eating and drinking places* (SIC: 55); *apparel and accessory stores* (SIC: 56); *home furnishing stores* (SIC: 57); *automotive dealers and service stations* (SIC: 58) and *miscellaneous retail* (SIC: 59). We do not include retailers in categories *eating and drinking places* and *automotive dealers and service stations* in our study because they contain significant service components to their businesses. We collect financial data for fiscal years 1993–2007 for the entire population of public firms listed with the remaining six SIC codes on the U.S. stock exchanges, NYSE, NASDAQ, and AMEX, from Standard and Poor's Compustat database using the Wharton Research Data Services (WRDS). There are 670 firms that reported at least one year of data to the U.S. Securities and Exchange Commission (SEC) for these years. Table 1 reports the number of retailers in each category. It also shows the distribution of retailers for the final data set and test sample.

We also collect data on the number of stores in each year and the projected number of stores for the following year from 10-K statements accessed through the Thomson Research database. Because generally accepted accounting principles (GAAP) does not mandate retailers to reveal store related information, many retailers did not report their number of stores in their 10-K statements. We find that 208 of the 670 firms did not report any store information. They include Internet, mail, and catalog retailers (SIC 5,961); credit card companies (SIC 5,900); and a few other retailers. From the remaining 462 retailers, we consider only those that have at least five consecutive years of data on number of stores to enable us to perform longitudinal analysis; 355 of the 462 firms meet this criterion.

From this data set, we remove foreign retailers that are listed as *American Depositary Receipt* (ADR) in the U.S. stock exchanges. We also eliminate jewelry firms from the miscellaneous retail sector because their inventory depends on commodity prices and macroeconomic conditions not captured by our model. We then combine SIC 52 with SIC 57 because SIC 52 has a small number of firms and its closest match is SIC 57. After applying these rules and removing firms

Table 1 Description of Initial, Final, and Test Data Sets by Retail Sectors, 1993–2007

Retail sector	Two-digit SIC code	Examples of firms	No. of firms	No. of firms that reported store information for at least five years	Final data set for 1993–2007		Test sample 2004–2007	
					No. of firms	No. of obs.	No. of firms	No. of obs.
Lumber and other building materials	52	Home Depot, Lowe's, National Home Centers	29	13	53	390	19	48
Home furnishing stores	57	Williams-Sonoma, Jennifer Convertibles, Circuit City	69	44				
General merchandise stores	53	Costco, Dollar General, Walmart	78	48	47	346	21	47
Food stores	54	Safeway, Dairy Mart Convenience stores, Shaws	92	57	51	328	5	11
Apparel and accessory stores	56	Mens Wearhouse, Harolds, Childrens Place	91	73	70	570	36	88
Miscellaneous retail	59	Toys R Us, Officemax, Walgreens	311	120	93	609	19	36
Total			670	355	314	2,243	100	230

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

with missing data, our final data set has 314 retailers classified into five segments and with 2,243 firm-year observations across 1993–2007. All further analysis is performed on this data set.

We obtain index of consumer sentiment (ICS) collected and compiled by the University of Michigan's Survey Research Center. The index of consumer sentiment represents consumers' confidence and is collected on a monthly basis. We obtain the monthly time series for 1993–2006.

2.2.1. Test Sample. Our test sample consists of 230 observations for 2004–2007. We generate forecasts for the test sample on a rolling-horizon basis. That is, data for 1993–2003 are used to fit the model and generate forecasts for 2004, then data for 1993–2004 are used to fit the model and generate forecasts for 2005, and so on. We describe the test sample and the setup for the comparison with analysts' forecasts in §4, after discussing the forecasting model.

3. Forecasting Model

3.1. System of Simultaneous Equations

We set up a system of three simultaneous equations to represent the relationships that our forecasting model is based on:

$$cs_{it} = F_i + \alpha_{11} is_{it} + \alpha_{12} gm_{it} + \alpha_{13} sgas_{it} + \alpha_{14} pi_{i,t-1} + \alpha_{15} g_{it} + \alpha_{16} ics_{t-1} + \varepsilon_{it}, \quad (1)$$

$$is_{it} = J_i + \alpha_{21} cs_{it} + \alpha_{22} gm_{it} + \alpha_{23} cs_{i,t-1} + \alpha_{24} pi_{i,t-1} + \alpha_{25} g_{it} + \alpha_{26} caps_{i,t-1} + \eta_{it}, \quad (2)$$

$$gm_{it} = H_i + \alpha_{31} cs_{it} + \alpha_{32} is_{it} + \alpha_{33} gm_{i,t-1} + \upsilon_{it}. \quad (3)$$

We refer to these equations as the (1) cost of sales equation, (2) inventory equation, and (3) gross margin equation. Each equation consists of firm fixed-effects (F_i, J_i, and H_i), coefficients of endogenous variables, coefficients of exogenous variables, and error terms (ε_{it}, η_{it}, and υ_{it}). The estimates of α_{11}, α_{12}, α_{21}, α_{22}, α_{31}, and α_{32} give the relationships among the endogenous variables. Figure 1 shows a block diagram of the SEM.

This model is based on the premise that cost of sales, inventory, and gross margin for retailers are jointly determined. First, consider the impact of inventory and gross margin on cost of sales. As the amount of inventory carried by a retailer increases, its number of units sold and cost of sales would increase due to various mechanisms: increasing service level, stimulating demand, or increasing choice for consumers. As shown in the literature, all three of these effects work in the same direction; see Dana and Petruzzi (2001), Hall and Porteus (2000), or Balakrishnan et al. (2004) for the demand stimulating effect of inventory, and see van Ryzin and Mahajan (1999) for the effect of variety on optimal inventory and sales. Next, as gross margin increases, the number of units sold and cost of sales would decline because demand is downward sloping in price. These relationships are represented in Equation (1).

Next, consider the impact of cost of sales and gross margin on inventory. A retailer's optimal stocking quantity is increasing in mean demand according to common inventory models, such as the EOQ model or the newsvendor model. Hence, we expect average inventory per store to be increasing in cost of sales per store. We also expect average inventory per store to be increasing in gross margin because higher margin induces retailers to carry more inventory, either through a higher optimal safety stock or higher optimal level of variety.

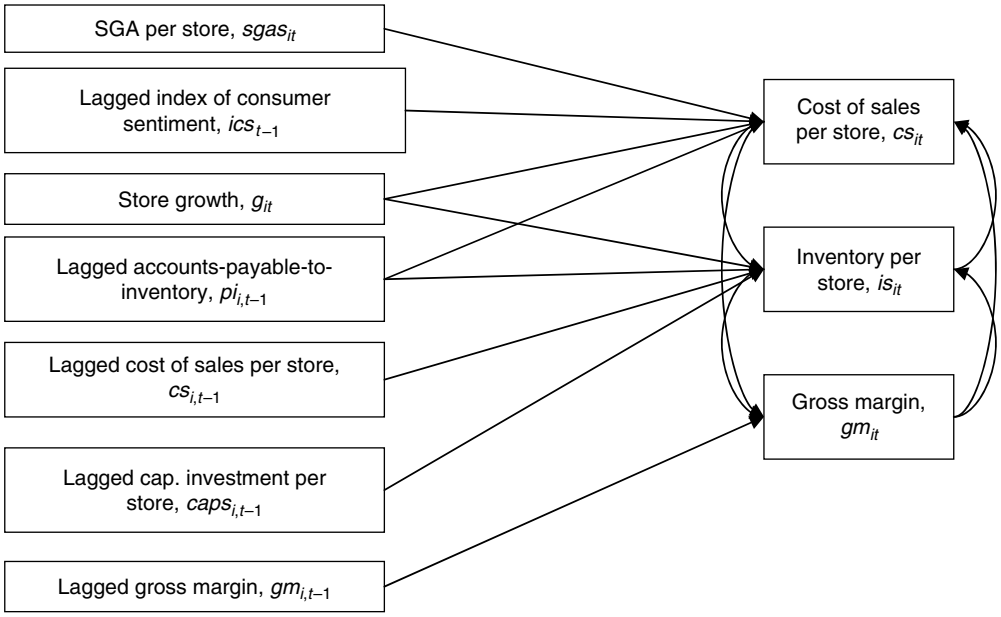
Finally, consider the impact of cost of ssales and inventory on gross margin for a retailer. For a given level of inventory, we expect that as demand increases, retailers would have fewer promotions and clearance sales. Cost of sales is a proxy for demand. Hence, retailers would have higher gross margin when cost of sales are higher. For given cost of sales, as inventory increases, the number of unsold units goes up, which would force the retailer to take larger markdowns on its merchandise or undertake larger liquidation of its merchandise through clearance sales. Hence, we expect gross margin to decrease with average inventory per store.

Apart from the variables discussed above, each equation has exogenous variables. SGA expenses for a retailer captures costs involved in building brand image, providing customer service, and conducting marketing activities that help to implement a retailer's competitive strategy which would lead to increase in sales (Palepu et al. 2004). We control for store growth because the composition of new and old stores would affect total sales and inventory differently. Sales in new stores can be less than that from old stores because new stores take time to reach maturity, or it can be more than that from old stores due to a fad effect; see Lundholm et al. (2010) for details. Similarly inventory level in new stores could be different from that of old stores. We use lagged accounts-payable-to-inventory ratio as it has been used in practice for sales forecasting (Raman et al. 2005).

We use the lagged index of consumer sentiment as a leading indicator of personal consumption expenditures as shown by Carroll et al. (1994). An increase in personal consumption expenditures in the economy would affect the demand faced by a retailer. We control for lagged capital investment per store as it captures retailer's investment in warehouses, information technology, supply chain infrastructure, etc. that could lead to increased efficiency and thus lower inventories. Finally, we control for lagged values of

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

Figure 1 Relationships Among Endogenous and Exogenous Variables in Our Model



cost of sales per store and gross margin as they are good predictors of future values of those variables. These exogenous variables enable us to identify the six directional relationships among the endogenous variables.

The equations are written in terms of logarithms of variables. We use a multiplicative model for several reasons: (a) multiplicative models of supply and demand are used extensively in the operations management, economics, and marketing literature; (b) multiplicative models deal with elasticities, which are more intuitive and translate more easily across firms of different sizes than the coefficients of a linear model; and (c) multiplicative models have been found to fit aggregate inventory levels in recent empirical research (Gaur et al. 2005, Rumyantsev and Netessine 2007).

We eliminate the fixed effects in (1)–(3) by first differencing. This gives us

$$\Delta cs_{it} = \alpha_{10} + \alpha_{11}\Delta is_{it} + \alpha_{12}\Delta gm_{it} + \alpha_{13}\Delta sgas_{it} + \alpha_{14}\Delta pi_{i,t-1} + \alpha_{15}\Delta g_{it} + \alpha_{16}\Delta ics_{t-1} + \Delta\varepsilon_{it}, \quad (4)$$

$$\Delta is_{it} = \alpha_{20} + \alpha_{21}\Delta cs_{it} + \alpha_{22}\Delta gm_{it} + \alpha_{23}\Delta cs_{i,t-1} + \alpha_{24}\Delta pi_{i,t-1} + \alpha_{25}\Delta g_{it} + \alpha_{26}\Delta caps_{i,t-1} + \Delta\eta_{it}, \quad (5)$$

$$\Delta gm_{it} = \alpha_{30} + \alpha_{31}\Delta cs_{it} + \alpha_{32}\Delta is_{it} + \alpha_{34}\Delta gm_{i,t-1} + \Delta\upsilon_{it}. \quad (6)$$

Here, Δ prefix for each variable denotes the first difference, e.g., $\Delta cs_{it} = cs_{it} - cs_{it-1}$.

Equations (4)–(6) serve two purposes in our analysis. They yield the forecasting equations after we eliminate the contemporaneous and endogenous variables from the right-hand side. Moreover, we shall show that the residuals from these equations in one year are predictors of the model's forecast accuracy for the next year.

3.2. Forecasting Equations and Estimation

We generate our forecasting equations from the simultaneous equations (4)–(6) as follows. First, we eliminate endogeneity by conducting a linear transformation of (4)–(6). This yields the *reduced form* of the equations where cost of sales per store, average inventory per store, and gross margin are written as linear functions of exogenous variables only. Next, we replace SGA per store by lagged SGA per store in both equations to remove look-ahead in forecasting. Then, we add lagged dependent variables in the SEM to reduce serial correlation; see Liu (1960) for a conceptual discussion of this type of augmentation. Since $\Delta cs_{i,t-1}$ and $\Delta gm_{i,t-1}$ are already present in the equations, we add $\Delta is_{i,t-1}$ and lagged gross margin $gm_{i,t-1}$ as two more variables to control for serial correlations.[4] This yields the following equations:

$$\Delta cs_{it} = \boldsymbol{\beta}_{1,s(i)}\mathbf{W}_{i,t-1} + \gamma_{1,s(i)}\Delta g_{it} + \varepsilon'_{it}, \quad (7)$$

$$\Delta is_{it} = \boldsymbol{\beta}_{2,s(i)}\mathbf{W}_{i,t-1} + \gamma_{2,s(i)}\Delta g_{it} + \eta'_{it}, \quad (8)$$

$$\Delta gm_{it} = \boldsymbol{\beta}_{3,s(i)}\mathbf{W}_{i,t-1} + \gamma_{3,s(i)}\Delta g_{it} + \upsilon'_{it}. \quad (9)$$

[4] Alternatively, one could estimate the autocorrelation coefficient and use it for forecasting. We find that the forecasts are more accurate when we use lagged variables to control for autocorrelation as opposed to estimating the autocorrelation coefficient and using it for forecasting.

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

Here, we use matrix notation for simplicity. Let $\mathbf{W}_{i,t-1}$ denote the column vector of all explanatory variables for firm i and year $t-1$; $\mathbf{W}_{i,t-1} = (\Delta cs_{i,t-1}, \Delta gm_{i,t-1}, \Delta is_{i,t-1}, \Delta sga_{i,t-1}, \Delta pi_{i,t-1}, \Delta caps_{i,t-1}, \Delta ics_{t-1}, gm_{i,t-1})'$. Let $\boldsymbol{\beta}_{1,s(i)}$, $\boldsymbol{\beta}_{2,s(i)}$, and $\boldsymbol{\beta}_{3,s(i)}$ be row vectors of the corresponding coefficients for the three equations, and let $s(i)$ be the segment of firm i. Thus, we use segment-wise estimates of coefficients. Store growth, Δg_{it}, is the only contemporaneous explanatory variable in these equations; all other variables are lagged. Finally, let $\gamma_{1,s(i)}$, $\gamma_{2,s(i)}$, and $\gamma_{3,s(i)}$ denote the coefficients of Δg_{it} in the three equations, and let ε'_{it}, υ'_{it}, and η'_{it} be the error terms.

To generate forecasts for year T, we estimate Equations (7)–(9) and obtain segment-wise coefficients using data up to year $T-2$. Our estimation methodology is based on Wooldridge (2002, Chaps. 8 and 9). We use the instrument variable generalized least squares (IVGLS) method to estimate the SEM because errors in our model can be both heteroscedastic and autocorrelated. We discuss the IVGLS procedure, tests of endogeneity and identification, and additional technical details in the online appendix.

We do not use data from year $T-1$ to estimate the coefficients because firms announce their financial results on different dates in the year. If we were to use data for year $T-1$ to estimate the coefficients, then we would have to wait until the *last* firm in the data set announced its financial results for year $T-1$.[5] We avoid this delay by estimating coefficients using data only up to year $T-2$. Then, in year $T-1$, we use these coefficients and generate forecasts for year T for each firm as soon as *that* firm's financial results for year $T-1$ are announced.

Let the estimated coefficients for segment s, based on data till year $T-2$, be denoted as $\hat{\boldsymbol{\beta}}_{1,s(i)}$, $\hat{\boldsymbol{\beta}}_{2,s(i)}$, $\hat{\boldsymbol{\beta}}_{3,s(i)}$ and $\hat{\gamma}_{1,s(i)}$, $\hat{\gamma}_{2,s(i)}$, $\hat{\gamma}_{3,s(i)}$. We then replace the contemporaneous variable store growth, Δg_{iT}, with projected store growth, $\Delta \hat{g}_{iT}$. Then, the forecast of sales for year T for firm i is generated using the following forecasting equations:

$$\widehat{cs}_{iT} = cs_{i,T-1} + \hat{\boldsymbol{\beta}}_{1,s(i)} \mathbf{W}_{i,T-1} + \hat{\gamma}_{1,s(i)} \Delta \hat{g}_{iT}, \quad (10)$$

$$\widehat{gm}_{iT} = gm_{i,T-1} + \hat{\boldsymbol{\beta}}_{3,s(i)} \mathbf{W}_{i,T-1} + \hat{\gamma}_{3,s(i)} \Delta \hat{g}_{iT}, \quad (11)$$

$$\hat{S}_{iT} = \mathrm{Exp}(\widehat{cs}_{iT} + \widehat{gm}_{iT}) * N_{i,T-1} * \mathrm{Exp}(\hat{g}_{iT}). \quad (12)$$

Here, $\widehat{cs}_{iT}$ and $\widehat{gm}_{iT}$ denote the forecasts of logged cost of sales per store and gross margin, and $\hat{S}_{iT}$ denotes the forecast of sales revenue for the firm obtained from $\widehat{cs}_{iT}$, $\widehat{gm}_{iT}$, the total number of stores in the previous year ($N_{i,T-1}$), and the projected store growth ($\hat{g}_{iT}$). Note that the forecasting equations consist of coefficients estimated from year $T-2$ and data for firm i for year $T-1$. The forecast $\hat{S}_{iT}$ can be calculated after the EAD for fiscal year $T-1$ for firm i, i.e., the date on which firm i reports its 10-K statement for fiscal year $T-1$ to its investors. Thus, we avoid any potential look-ahead in forecasting.

We use the above process to generate forecasts for our test sample by setting $T = 2004, \ldots, 2007$.

3.3. Remarks

We estimate the coefficients of the SEM equations (4)–(6) for the entire data set with data pooled across all segments as well as separately for each of the five segments. In the latter case, we have a total of 30 coefficient estimates made up of 6 coefficient estimates for each of the five segments. We find the coefficients to be significant in 27 of the 30 cases ($p < 0.05$) as shown in Table 2. Thus, our estimation results show that cost of sales, inventory, and gross margin for a retailer are jointly determined. Furthermore, the heterogeneity in magnitude of our coefficient estimates across different segments indicate that it is important to use segment-wise coefficients' estimates for forecasting purposes.

4. Comparison with Analysts' Forecasts: Method

The most important consideration in our methodological design is ensuring that the information environment for the model and analysts are similar, so that the comparison of forecast accuracy is fair. Although we cannot control the information available to analysts, we can ensure that all the information used in the model is publicly available at the time when the model's and analysts' forecasts are issued. This section explains the timeline that we set up to ensure this and the resulting test sample. Figure 2 depicts the timeline.

Our final data set, described in §2, has 595 observations across 2004–2007, divided as 163, 156, 142, and 134 retailers in the years 2004, 2005, 2006, and 2007, respectively. First, we combine these observations with data on equity analysts' forecasts. We obtain analysts' forecasts of annual sales from the I/B/E/S accessed using the WRDS for 2004–2007. The database has estimates from over 200 brokerage houses and 2,000 individual analysts (I/B/E/S Glossary 2001). I/B/E/S contains three main sections: *detail history, summary history*, and *recommendations*. We obtain individual analysts' estimates from the *detail history* database for our analysis. Each record in the *detail history* database contains information such

[5] For example, suppose that two firms, A and B, announce their results for fiscal year 2003 on February 28, 2004, and May 31, 2004, respectively. If we were to use the data for year 2003 for estimating the model, we would have to wait at least until May 31, 2004. Thus, we would not be able to issue forecasts of sales for firm A during the period from February 28, 2004, to May 31, 2004.

Table 2 Coefficients' Estimates for the Endogenous Variables in Simultaneous Equations (4)–(6) for All Retail Segments, 1993–2007

	Cost of sales equation (Δcs_{it})		Inventory equation (Δis_{it})		Gross margin equation (Δgm_t)	
Retail industry segment	Gross margin (Δgm_{it})	Inventory per store (Δis_{it})	Cost of sales per store (Δcs_{it})	Gross margin (Δgm_{it})	Cost of sales per store (Δcs_{it})	Inventory per store (Δis_{it})
General merchandise stores	−1.222**	0.375**	0.932**	−3.321**	0.089**	−0.087**
	(0.131)	(0.012)	(0.016)	(0.229)	(0.016)	(0.017)
Food stores	−1.938**	0.441**	0.737**	0.074	0.064**	−0.050**
	(0.343)	(0.019)	(0.016)	(0.323)	(0.012)	(0.014)
Apparel and accessory stores	−1.283**	0.475**	0.768**	0.521**	0.138**	−0.069**
	(0.115)	(0.026)	(0.013)	(0.112)	(0.026)	(0.030)
Home furnishing stores	−0.530	0.368**	0.909**	−0.879	0.173**	−0.172**
	(0.426)	(0.014)	(0.013)	(0.520)	(0.062)	(0.062)
Miscellaneous retail	−2.023**	0.332**	0.852**	1.711**	0.116**	−0.135**
	(0.087)	(0.017)	(0.006)	(0.116)	(0.016)	(0.019)

Notes. All variables have been first differenced. $n = 2{,}243$.
**Denotes statistically significant at 0.05. Standard errors are reported in parentheses below the parameters' estimates.

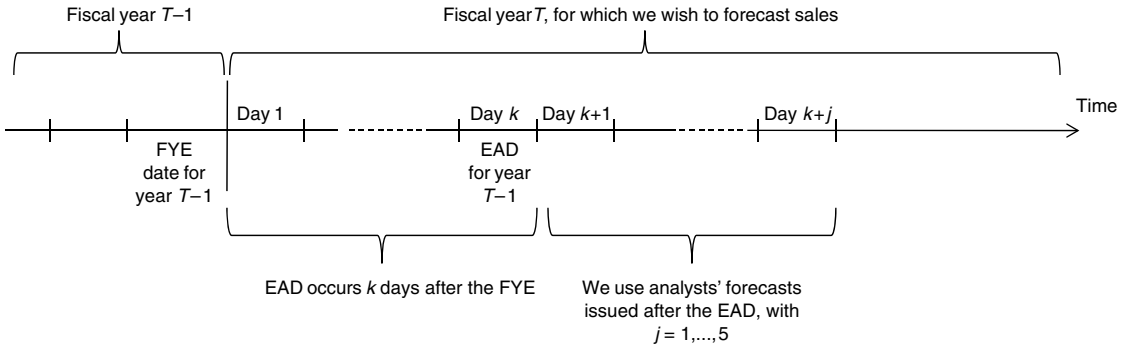
as the analyst's code, brokerage company, forecasted company, forecast year, forecast, estimate date, etc. Not all retailers have analyst coverage. We find 96, 93, 97, and 82 retailers that had sales forecasts from analysts during 2004, 2005, 2006, and 2007, respectively, which brings the number of test observations down from 595 to 368.

Then, we combine the 368 observations with projected store growth rates. Eighty percent of the retailers in our test data set reported in their 10-K statements the projected number of stores that will be closed or opened during the next fiscal year. For the remaining retailers, we set the projected store growth in year t as the actual store growth in year $t-1$, i.e., $\hat{g}_{it} = g_{i,t-1}$.

A firm announces its earnings and financial statements several days after the FYE date. This gap is illustrated in Figure 2, which shows the FYE and EAD for a retailer. Let k be the number of days between the FYE and EAD. In our test data set with analysts' coverage ($n = 368$), 24% of observations have financial results reported within 30 days after the FYE, 89% within 60 days, and 97% within 90 days. We conduct analysis using cutoff values of 60 and 90 days for k. Thus, we have 327 observations with $k \leq 60$ and 356 observations with $k \leq 90$.

Each firm is covered by many analysts. Each analyst issues sales forecasts for a covered firm many times, revising older forecasts as newer information becomes available. These forecasts are time-stamped with the dates on which they are issued. For each firm year, we pick analysts' forecasts that were issued *after* the EAD to ensure that the financial statements for the previous fiscal year are available to analysts when their forecasts are issued. Let j denote the number of days between the date when a retailer announces its earnings (EAD) and the date when an analyst issues a sales forecast for the next fiscal year, i.e., $j = 0, 1, \ldots$ corresponds to the dates EAD, EAD + 1, etc.

Figure 2 Timeline Used to Compare Model Forecasts with Analysts' Consensus Forecasts



Notes. In our test sample, we do two sets of analysis, one for all observations with $k \leq 60$ days and the other for all observations with $k \leq 90$ days. The model's sales forecast for any given firm is computed on its EAD using coefficients estimated from data from all firms up to year $T-2$ and data from this firm up to year $T-1$.

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

Table 3 Summary Statistics of Variables for the Test Sample of 230 Observations for 2004–2007

Variable name	Notation	Mean	Standard deviation	Min	Max
Sales ($ M)	S_{it}	14,239.07	44,763.95	158.655	375,376
Number of stores	N_{it}	1,228.51	1,545.15	52	7,929
Average cost of sales per store ($ M)	CS_{it}	8.11	16.32	0.335	115.267
Average inventory per store ($ M)	IS_{it}	1.523	2.183	0.054	15.118
Gross margin	GM_{it}	1.582	0.245	1.124	2.612
Average SGA per store ($ M)	$SGAS_{it}$	2.345	3.109	0.158	17.44
Average capital investment per store ($ M)	$CAPS_{it}$	3.207	4.404	0.129	20.484
Store growth	G_{it}	1.089	0.144	0.770	2.072
Accounts-payable-to-inventory ratio	PI_{it}	0.469	0.219	0.122	1.280
Index of consumer sentiment	ICS_t	90.288	5.649	74.2	103.8

All forecasts where $j \leq 0$ are dropped. We perform an additional check to ensure that the I/B/E/S field *fiscal period indicator* is set to 1 to ensure that we do not pick any analyst forecast that was issued before the EAD.[6]

The main analysis reported in this paper is based on analysts' forecasts with $j = 1, \ldots, 5$. We find that there are 220 and 230 firm-year observations in the cases with $k \leq 60$ days and $k \leq 90$ days that have at least one analyst forecast issued during $j = 1, \ldots, 5$. These form the final data sets for our comparison of forecast accuracy. Table 3 provides the summary statistics of the 230 firm-year observations. Sometimes, an analyst may issue multiple sales forecasts for a firm within this five-day period. In all such cases, we use the last of these forecasts from that analyst and discard all previous. We compute the median of all individual analysts' forecasts for a given firm-year combination and call it the *consensus forecast*. We prefer using consensus forecasts over individual analysts' forecasts to benchmark our model performance for the following reasons. First, comparison with consensus forecasts would be more conservative than comparison with individual forecasts. Second, comparison with individual forecasts is not straightforward because of cross-sectional dependencies among analysts' forecast errors, which could violate the independence assumption necessary for conducting *t*-tests. We also measure consensus forecasts alternatively as the average of the individual analysts' forecasts and obtain the same results.

Table 4 gives the descriptive statistics of analysts' forecasts issued during $j = 1, \ldots, 5$ days for retailers with $k \leq 90$ days. On average, there are 4.4 analysts' forecasts per firm. Any one analyst may have revised his or her forecast many times during this five-day window, but each analyst is counted only once in the average. The maximum number of analysts covering a firm is 14 and the minimum is 1.

[6] The fiscal period indicator variable is set to 2 by the I/B/E/S if a forecast is issued before the financial statement was released and to 1 otherwise.

Table 4 Summary Statistics of Analysts' Forecasts for 2004–2007 for the Time Window $j = 1, \ldots, 5$ Days for Retailers with the EAD Within 90 Days After the FYE

	2004	2005	2006	2007
Average number of analysts per firm	4.48	4.17	4.54	4.29
Min number of analysts per firm	1	1	1	1
Max number of analysts per firm	11	14	14	14
Dispersion among analysts' forecasts (%)	1.16	1.24	1.43	1.31

Notes. Dispersion for each firm-year observation is defined as the standard deviation of analysts' forecasts divided by average of analysts' forecasts for that firm year, with analysts' forecasts taken for $j = 1, \ldots, 5$. Dispersion is undefined when only one analyst forecast is available. Thus, we compute the dispersion for all firm-year observations and then report the average for each year. For example, the reported number of 1.16% is the average of dispersions for all firms in year 2004.

The above methodology avoids look-ahead and is conservative for several reasons. Perhaps the most important is that the majority of retailers in our data set announce realized sales at the end of each month. In such cases, by the time of earnings announcement, analysts would have not only information contained in financial statements, but also sales for the first one to two months of the current year for which they are forecasting. They could also have other valuable information such as the state of economy and sales of competing retailers for the months that have elapsed. Such information is not contained in our model and as such gives an advantage to analysts.

5. Comparison with Analysts' Forecasts: Results

We compare the forecast accuracy of our model with equity analysts using the APE, defined as APE = $|(Sales - Forecast)/Sales|$, where sales and forecasts are in dollars. The sales of retailers in our sample vary widely. For example, in 2007, the largest retailer, Walmart (Ticker: WMT), had sales of US$375 billion and the smallest retailer, Cache Inc. (Ticker: CACH),

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

had sales of US$274 million. Thus, it is important to choose a metric that normalizes for the size of the retailer. The APE is a commonly used metric that fills this requirement. As a validation test on this metric, we checked if analysts' sales forecasts are biased. We found that the sales forecasts in our data set have no statistically significant bias. This is consistent with past research on analysts' sales forecasts; see Mest and Plummer (2003).

We report the median of APE, its mean (MAPE), and the percentage of retailers for which forecasts from our model have lower APE than the benchmark forecasts. Through these metrics, our model can be compared with analysts' consensus forecasts using two different but equally important criteria. First, we determine if the model yields a lower MAPE than the consensus forecasts. We test the statistical significance of difference in the MAPE using one-tailed t-test and Johnson's skewness-adjusted t-test (Johnson 1978); the latter is used because the difference in the MAPE can have a skewed distribution. Second, we determine if the model performs better than the consensus forecasts for more than one half of the firms, i.e., if it yields a lower APE for more than 50% of the firms. This criterion requires a nonparametric binomial sign test. Note that the first criterion determines if the model produces smaller forecast errors on average, whereas the second criterion determines if the model beats the consensus forecast for a randomly picked firm.

Table 5 presents the forecast accuracies of our model and analysts' consensus forecasts for retailers who released earnings within 60 and 90 days after the FYE, i.e., $k \leq 60$ or 90 days in Figure 2. For each firm year, we consider analysts' forecasts made during $j = 1$ to 5 days after the EAD to generate consensus forecasts. For $k \leq 60$ days, our model gives a MAPE of 4.09% and the consensus forecast gives a higher MAPE of 4.40%, a difference of 0.31% of sales. For $k \leq 90$ days, the model gives MAPE = 4.13% and the consensus forecasts give MAPE = 4.49%, a difference of 0.36% of sales. These differences are statistically significant at $p < 0.1$ and $p < 0.05$, respectively. The skewness-adjusted t-statistic yields the same conclusions. Thus, the MAPE of model forecasts is significantly lower than that of consensus forecasts in each case.

The median of the APE of model forecasts is also lower than that for analysts' consensus forecasts. Its values for $k \leq 60$ days are 2.99% for our model and 3.11% for the consensus, and its values for $k \leq 90$ days are 2.98% and 3.12%, respectively, for the model and the consensus.

The last part of Table 5 shows the statistics for the binomial sign test. We find that model forecasts are more accurate than equity analysts' consensus forecasts in 51.82% and 52.17% cases in the two time windows, respectively. The difference from 50% is not statistically significant, with p-values of 0.32 and 0.28, respectively, for $k \leq 60$ and 90 days for the model. Thus, the number of firms for which the model yields lower APE than the consensus forecast is statistically indistinguishable from 50%. In other words, for a randomly picked firm, it is equally likely that either the model forecast or the consensus forecast may have the lower APE.

Table 5 Comparison of Forecast Accuracy Against Analysts' Consensus Forecasts for 2004–2007

	$k \leq 60$ days		$k \leq 90$ days	
Statistics	Model	Consensus	Model	Consensus
MAPE (%)	4.09	4.40	4.13	4.49
Median of the APE (%)	2.99	3.11	2.98	3.12
Minimum of the APE (%)	0.06	0.04	0.06	0.04
Maximum of the APE (%)	28.79	28.81	28.79	28.81
Difference in the MAPE	0.31%		0.36%	
(t-stat.; skewness adj. t-stat.)	(1.46*; 1.51*)		(1.58*; 1.67**)	
% of observations where model gives lower APE	51.82 (0.32)		52.17 (0.28)	
(p-value for binomial sign test)				
N	220		230	

Notes. Skewness adjusted t-statistic refers to Johnson's skewness-adjusted t-statistic. The reported t-statistics are from one-tailed tests, where the null hypothesis is that the MAPE of forecasts from reduced form model is greater than or equal to that of consensus forecasts. The binomial sign test is used to test if analysts' forecasts are more likely to produce lower APE than model forecasts.

*$p < 0.1$; **$p < 0.05$.

Combining the results from these two criteria, we see that the model performs better than the analysts' consensus on one criterion and equally as well as the analysts' consensus on the second criterion. The question arises whether the average magnitude of improvement in the MAPE is large enough to warrant managerial attention, even if it is statistically significant. To this, we note that forecasting sales accurately is important because small changes in sales expectations can have large consequences on firm valuation for retailers. For instance, consider the effect of sales forecasts on earnings forecasts. Suppose that the sales forecast of a retailer is increased by 1%. The profit after tax will then increase by 1% if all costs are variable, but by much more than 1% if some of the costs are fixed. In retailing, this effect is magnified by the fact that retailers' fixed costs are often a large fraction of sales and profit after tax is a small fraction, about 1%–5%, of sales. Thus, the impact on the value of the firm is substantial. This argument has been illustrated in many contexts in the operations literature; see Fisher and Raman (2010) for a

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

recent reference. Now, consider the effect of sales forecast on the forecast of sales growth rate. The sales forecast for a year gives a projection of sales growth rate for that year, which is then used to extrapolate sales growth rates for subsequent years. Because sales growth rate has a multiplier effect on valuation, valuation can change dramatically with small changes in sales forecast. In an example given in Lundholm et al. (2010), if a firm had an expected constant return on equity of 20%, a cost of equity capital of 10%, and an expected perpetual growth rate of 5%, then raising the growth rate to 6% would increase the firm's value 17%, and lowering the growth rate to 4% would decrease the firm's value 11%.

We analyze the sensitivity of the results shown in Table 5 to the chosen values of j. Thus far, the consensus forecasts are computed from analysts' forecasts made during $j = 1, \ldots, 5$ days after the EAD. We vary the terminal date from 6 days to 10 days and recompute consensus forecasts and forecast accuracy. The results for $j \in [1, 6]$, $[1, 8]$, and $[1, 10]$ are shown in Table 6, with $k \leq 90$ days; results for other terminal values for j are similar and, hence, not reported. First, note that as expected, the sample size of firms as well as the average number of analysts forecasting per firm increase with j. Second, the results are consistent with those in Table 5. Our model yields lower MAPE in each scenario, and the differences are significant at $p < 0.05$. Our model yields lower APE in more than 50% of the observations in *every* scenario but, as before, the difference from 50% for either the analysts or the model is not statistically significant.

We repeat the previous analysis for each individual year 2004–2007. We find that the median and mean of the APE of model forecasts are lower than those of consensus forecasts in each of those years. For example, the median (mean) of the APE of model forecasts and analysts' forecasts in 2007 were 3.22% (4.87%) and 4.17% (5.21%), respectively. Thus, we can conclude that our aggregate results are not driven by any particular year in our sample.

5.1. Performance Difference Explained by Inventory and Gross Margin Residuals

We now investigate the above results in detail to understand why our model gives lower MAPE relative to analysts' consensus. Recall that the model is based on three equations specifying the relationships among sales, inventory, and gross margin. Thus, the question is whether the lower MAPE of the model is due to analysts not incorporating the relationship among sales, inventory, and gross margin captured in our model in their forecasts. We address this question in the following way. Our results of the SEM indicate that retailers may increase sales by using inventory and margin. We identify retailers where we expect this relationship to be salient, i.e., we identify retailers for whom inventory and margin may have played a large role in driving sales growth historically. Then, we compare the forecast accuracy of our model with consensus forecasts for this subset of retailers in the test data set.

Specifically, suppose that we are generating sales forecasts for a given firm for year T using the model. Consider the residuals of the inventory and gross margin simultaneous equations (5) and (6) for year $T-1$ for this firm. We call these residuals *inventory residual* and *margin residual*, respectively. They represent the deviations of actual values from model predictions. In other words, we interpret the inventory (margin) residual as the amount by which the retailer

Table 6 Longitudinal Change in Forecast Accuracy for 2004–2007 When the Time Window for Generating Analysts' Consensus Forecasts Is Varied; $k \leq 90$ Days

	$j \in [1, 6]$ days		$j \in [1, 8]$ days		$j \in [1, 10]$ days	
Statistics	Model	Consensus	Model	Consensus	Model	Consensus
MAPE (%)	4.08	4.49	4.03	4.46	4.09	4.46
Median of the APE (%)	2.98	3.12	2.98	3.13	3.08	3.22
Minimum of the APE (%)	0.06	0.04	0.06	0.03	0.06	0.03
Maximum of the APE (%)	28.79	29.04	28.79	28.79	28.79	28.79
Difference in the MAPE	0.41%		0.42%		0.35%	
(t-stat.; skewness adj. t-stat.)	(1.69**; 1.77**)		(1.95**; 1.68**)		(1.68**; 1.75**)	
% of observations where model gives lower APE	52.38		52.25		51.58	
(p-value from binomial sign test)	(0.26)		(0.26)		(0.33)	
n	231		245		252	
Average number of analysts per firm	4.45		4.57		4.58	

Notes. Skewness adjusted t-statistic refers to Johnson's skewness-adjusted t-statistic. The reported t-statistics are from one-tailed tests, where the null hypothesis is that the MAPE of forecasts from reduced form model is greater than or equal to that of consensus forecasts. The binomial sign test is used to test if analysts' forecasts are more likely to produce lower APE than model forecasts.

*$p < 0.1$; **$p < 0.05$.

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

has higher or lower growth in inventory per store (change in gross margin) than predicted for given values of all other variables. Note that these residuals can be computed on the EAD of the given firm without look-ahead because they are based on information up to year $T-1$ only.

We posit that our model will yield more accurate forecasts than analysts' consensus for year T when one or more of these residuals for year $T-1$ is large in magnitude. That is, the residuals for year $T-1$ will predict the forecasting accuracy of the model relative to analysts' consensus for year T. Thus, we use these residuals to identify retailers whose sales growth would have been influenced by inventory and margin to a large extent in year $T-1$. A retailer with high (low) inventory residual is called overinventoried or OI (underinventoried or UI). A retailer with high (low) margin residual is called overpriced or OP (underpriced or UP). Our thesis is that an OI retailer may have obtained sales growth in year $T-1$ by having too much inventory, and extrapolating this sales growth into the future would implicitly assume further abnormal increases in inventory, which might not be sustainable due to increasing costs of such actions. Therefore, if analysts did not incorporate historical inventory residual in their sales forecasts, then their consensus sales forecast would be less accurate than our model. We apply analogous reasoning to UI, OP, and UP retailers.

Thus, inventory residual and margin residual allow us to identify subsamples where model forecasts can be expected to perform better than analysts' forecasts. We identify these subsamples as follows:

(i) Estimate Equations (4)–(6) using data for all retailers up to year $T-2$. This estimation gives us coefficients estimates. It also gives us residuals from Equations (5) and (6) for all observations up to year $T-2$. We compile empirical distributions of inventory and margin residuals for year $T-2$.

(ii) Using the coefficients estimated in step (i), compute inventory residual and gross margin residual for each firm for year $T-1$.

(iii) Using the residuals computed in step (ii) and the empirical distributions compiled in step (i), classify retailers in the following way: A retailer is said to be OI (UI) in year $T-1$ if its inventory residual for year $T-1$ lies in the top (bottom) pth percentile for this residual in the empirical distribution for inventory residual in year $T-2$ computed in step (i).[7] Likewise, a retailer is said to be OP (UP) in year $T-1$ if its margin residual lies in the top (bottom) pth percentile for this residual.

If a retailer is classified as belonging to one (or two) of these four categories, OI, UI, OP, and UP, then it is included in the subsample for further analysis. We conduct this analysis using values of the cutoff percentile $p = 5, 10, 15$, and 20, with k fixed at 90 days after the FYE. When $p = 5$, there are 24 firm-year observations across 2004–2007 in the subsample consisting of OI, UI, OP, and UP observations; this represents 10% of the total sample of 230 firms for $k =$ 90 days. The number of observations increases to 58 (25% of sample) for $p = 10$, 89 (39% of sample) for $p = 15$, and 130 (57% of the sample) for $p = 20$. To assess how firms in these categories compare with each other and to the rest of the test sample, we compute average values of inventory and margin residuals for the case of $p = 10$. We find that the average value of the inventory residual for the OI firms is 0.17. This means that OI firms in our data set have on average 17% more inventory growth than predicted by the model. In comparison, the average value of the inventory residual for the UI observations is -0.12 and for the rest of the firms is 0.01. Thus, OI category captures firms whose inventory grew by much more than the model prediction in year $T-1$, and UI category captures firms whose inventory grew by much less than the model prediction. Likewise, the margin residual has average values of 0.05, -0.07, and 0.00 for OP, UP, and the rest of the retailers, respectively, when $p = 10$. This means that the gross margin for OP firms grew by 5% more than the model prediction, and that for UP firms grew by 7% less than the model prediction on average.

Table 7 presents the results of the comparison of model forecasts against consensus forecasts for these subsamples of retailers. We find that the model forecasts give significantly lower MAPE ($p < 0.1$ or 0.05) in all cases. Moreover, the model gives lower APE in about 56% to 62.5% of the observations. The binomial sign test is statistically significant with 90% confidence for $p = 10$, 15, and 20.

Furthermore, we find that the magnitude of difference between the model and the consensus forecasts is larger for each subsample than for the entire test sample shown in Table 5. The differences in the MAPE are 1.39%, 0.92%, 0.70%, and 0.76% of sales for $p = 5$, 10, 15, and 20, respectively. The differences in median of the APE are 2.43%, 1.55%, 0.53%, and 0.41% in the same sequence. As expected, the forecasting advantage of the model decreases with increasing cutoffs. Thus, the model forecasts are more valuable for higher cutoffs because analysts' errors tend to be larger.

As before, we test the sensitivity of these results to the time window on j within which analysts' forecasts are issued. Whereas the results reported in Table 7

[7] Here we use the distribution of residuals from year $T-2$ for the same timing reason as discussed in §3.2. At the EAD of retailer i for fiscal year $T-1$, the financial statements of other retailers for year $T-1$ need not be available so that the distribution of residuals for $T-1$ would not be known.

Table 7 Comparison of Forecast Accuracy for OI, UI, OP, and UP Retailers Against Consensus Forecasts for 2004–2007; $k \leq 90$ Days

	Cutoff percentile, $p = 5\%$		$p = 10\%$		$p = 15\%$		$p = 20\%$	
Statistics	Model	Consensus	Model	Consensus	Model	Consensus	Model	Consensus
MAPE (%)	6.23	7.62	4.67	5.59	4.22	4.92	4.02	4.78
Median of the APE (%)	4.04	6.47	3.23	4.78	3.26	3.79	3.17	3.58
Minimum of the APE (%)	0.55	0.32	0.16	0.17	0.16	0.16	0.16	0.16
Maximum of the APE (%)	18.78	20.76	18.79	20.76	18.78	20.76	18.78	21.39
Difference in the MAPE	1.39%		0.92%		0.70%		0.76%	
(t-stat.; skewness adj. t-stat.)	(1.57*; 1.58*)		(1.87**; 1.88**)		(1.83**; 1.85**)		(2.20**; 2.26**)	
% of sample firms where model forecasts have lower APE	62.5		60.34		58.42		56.15	
(p-value for binomial sign test)	(0.15)		(0.07)		(0.07)		(0.09)	
n	24		58		89		130	
(% of total test sample of 230 firms)	(10.4)		(25.2)		(38.7)		(56.5)	

Notes. Skewness adjusted t-statistic refers to Johnson's skewness-adjusted t-statistic. The reported t-statistics are from one-tailed tests, where the null hypothesis is that the MAPE of forecasts from reduced form model is greater than or equal to that of consensus forecasts. The binomial sign test is used to test if analysts' forecasts are more likely to produce lower APE than model forecasts.

*$p < 0.1$; **$p < 0.05$.

used a time window of $j \in [1, 5]$ days after EAD, in Table 8 we show corresponding results for time windows of $j \in [1, 6]$, $[1, 8]$, and $[1, 10]$ with the cutoff percentile p set to 10 and $k \leq 90$ days. The sample size increases slightly and the qualitative inferences are identical. In fact, when we increase the time window further to $j \in [1, 45]$ days keeping p and k unchanged, the sample size increases to 78 and the MAPE of model forecasts remains significantly lower than the MAPE of analysts' consensus forecasts ($p < 0.1$). The former is equal to 4.66%, and the latter 5.32%. The percentage of retailers where model forecasts have lower APE remains significantly higher than 50% ($p < 0.1$). Hence, the model forecasts continue to be more accurate than analysts' forecasts for a long time after the earnings are released. Only when the time window for j increases to about 60 days after the EAD does the difference in the MAPE between the model and consensus forecasts lose statistical significance.

We conclude that the performance of the model relative to the analysts' consensus improves as the magnitude of either the inventory or the margin residual increases. This suggests that the model is able to improve forecast accuracy because analysts do not fully incorporate historical inventory and gross margin values in their sales forecasts. It also suggests that the model is complementary to the analysts. When historical inventory and/or margin residuals are large in magnitude, the model does significantly better than the analysts' consensus, whereas when both residuals are small in magnitude, the analysts' consensus has the lower APE. Because this difference is based on historical residuals, it can be predicted in advance.

Table 8 Longitudinal Change in Forecast Accuracy for Retailers Identified as OI, UI, OP, or UP for 2004–2007 When the Time Window for Generating Analysts' Consensus Forecasts Is Varied; $k \leq 90$ Days, Cutoff Percentile $p = 10$

	$j \in [1, 6]$ days		$j \in [1, 8]$ days		$j \in [1, 10]$ days	
Statistics	Model	Consensus	Model	Consensus	Model	Consensus
MAPE (%)	4.64	5.57	4.57	5.53	4.77	5.56
Median of the APE (%)	3.26	4.72	3.28	4.50	3.38	4.72
Minimum of the APE (%)	0.16	0.17	0.16	0.03	0.16	0.03
Maximum of the APE (%)	18.79	20.76	18.79	20.67	18.79	20.67
Difference in the MAPE	0.93%		0.96%		0.79%	
(t-stat.; skewness adj. t-stat.)	(1.91**; 1.91**)		(2.02**; 2.04**)		(1.73**; 1.73**)	
% of sample firms where model forecasts have lower APE	61.01		59.09		59.42	
(p-value from binomial sign test)	(0.06)		(0.09)		(0.07)	
n	59		66		69	
Average number of analysts per firm	4.37		4.52		4.42	

Notes. Skewness adjusted t-statistic refers to Johnson's skewness-adjusted t-statistic. The reported t-statistics are from one-tailed tests, where the null hypothesis is that the MAPE of forecasts from reduced form model is greater than or equal to that of consensus forecasts. The binomial sign test is used to test if analysts' forecasts are more likely to produce lower APE than model forecasts.

*$p < 0.1$; **$p < 0.05$.

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.

5.2. Sensitivity Analysis

We perform a few additional sensitivity analyses that we describe briefly. We examine the effect of length of history of data used to calibrate our model on forecast accuracy. Using historical time periods varying from 8 to 13 years, we find that our results are not sensitive to the length of the time period of data. For example, when we forecast for 2007, we could use 13 years of data from 1993–2005 to calibrate the model. When we calibrate the model with different years of data, we find that the MAPE of model forecasts vary randomly between 4.13% (13 years of data) and 4.04% (8 years of data) for $k = 90$ days. We also test the impact of ignoring endogeneity in estimating our equations by assuming that all the explanatory variables in the simultaneous equations (4)–(6) are exogenous. We generate the residuals and classify retailers as OI, UI, OP, and UP based on these new residuals. Our comparison of forecast accuracy of model forecasts against equity analysts' forecasts for this sample shows that the relative advantage of the model forecasts over analysts' forecasts diminishes when the residuals are computed by ignoring endogeneity when estimating the simultaneous equations. For example, for the case where cutoff percentile p is set to 10 and $k \leq 90$ days we find that the MAPE of model forecasts when endogeneity is ignored is 4.92%, and it reduces to 4.67% when endogeneity is taken into account. In both cases, the MAPE of the model is significantly lower than that of analysts' forecasts ($p < 0.05$). Finally, we add that ignoring endogeneity among these variables would lead to inconsistent estimators that could lead to unpredictable results in different samples, so we recommend taking the endogeniety into account for the analysis.

6. Conclusions

Our paper develops and tests a forecasting model motivated by the idea that historical sales of retailers are endogenous with their contemporaneous inventories and gross margins. It shows that historical inventory and margin data are extremely valuable for forecasting sales at the firm-year level. Moreover, the model performs well against a benchmark of consensus analysts' forecasts issued after the earnings announcement. It beats consensus forecasts with respect to the MAPE and performs as well as the consensus forecasts on the binomial sign test. On a large subsample of observations constructed based on prior-year residuals, the model beats the consensus forecasts on both metrics by significant margins.

Our results suggest that analysts do not adequately take into account historical inventory and gross margin data in their sales forecasts. This is consistent with recent research in accounting. For example, Ivkovic and Jegadeesh (2004, p. 462) state that "Indeed, the value of analysts' earnings forecasts and recommendations stems more from their independent collection of information than from their interpretation of public information." In this sense, our model complements analysts' forecasts by generating more accurate forecasts when prior-year inventory or gross margin residuals are large in magnitude.

Our findings raise further questions that may be investigated in future research. First, although we compare with analysts' forecasts, it would be valuable to study similar questions using forecasts of sales by retail managers. For example, whether they incorporate historical inventories and gross margin in their forecasts of sales. Such a study could be beneficial to not only equity analysts but also retail managers.

Second, it may be possible to use sales, gross margin, and inventory forecasts from our model to improve equity analysts' earnings forecasts. We could determine if equity analysts' earnings forecasts also contained large errors for OI, UI, OP, and UP retailers. This would allow opportunities to improve earnings forecasts of analysts.

Third, although we have compared the forecasts of our model with analysts' consensus forecasts, one may study if some subgroups of analysts, identified from industry surveys or ex ante performance, incorporate the principles of our model in their sales forecasts. One could study if the consensus forecast accuracy is affected by movement of analysts from one firm to another. One could also study if the length of time for which an analyst has covered a retailer affects that analysts' forecast accuracy. Analysts who have spent more time with a retailer might understand that retailer's operations well and be able to incorporate this knowledge in their forecasts.

7. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://mansci.journal.informs.org/.

Acknowledgments

The authors are grateful to the Baker Research Services, Harvard Business School, for helping with data collection for this study. The authors are also grateful for the feedback given by several individuals including Nicole DeHoratius, Marshall Fisher, Aleda Roth, Roy Shapiro, Vinod Singhal, and especially the associate editor and three anonymous reviewers for *Management Science*.

References

Balakrishnan, A., M. S. Pangburn, E. Stavrulaki. 2004. "Stack them high, let 'em fly": Lot-sizing policies when inventories stimulate demand. *Management Sci.* **50**(5) 630–644.

Carroll, C. D., J. C. Fuhrer, D. W. Wilcox. 1994. Does consumer sentiment forecast household spending? If so, why? *Amer. Econom. Rev.* **84**(5) 1397–1408.

Chen, H., M. Z. Frank, O. Q. Wu. 2005. What actually happened to the inventories of American companies between 1981 and 2000? *Management Sci.* **51**(7) 1015–1031.

Conroy, R., R. Harris. 1987. Consensus forecasts of corporate earnings: Analysts' forecasts and time series methods. *Management Sci.* **33**(6) 725–738.

Corbett, C. J., M. J. Montes-Sancho., D. A. Kirsch. 2005. The financial impact of ISO 9000 certification in the United States: An empirical analysis. *Management Sci.* **51**(7) 1046–1059.

Dana, J. D., Jr., N. C. Petruzzi. 2001. Note: The newsvendor model with endogenous demand. *Management Sci.* **47**(11) 1488–1497.

Ertimur, Y., J. Linvat, M. Martikainen. 2003. Differential market reactions to revenue and expense surprises. *Rev. Accounting Stud.* **8**(2–3) 185–211.

Fisher, M., A. Raman. 2010. *The New Science of Retailing*. Harvard Business School Press, Boston.

Gaur, V., S. Kesavan. 2009. The effects of firm size and sales growth rate on inventory turnover performance in the U.S. retail sector. N. Agrawal, S. Smith, eds. *Retail Supply Chain Management*. Springer Science+Business Media, New York, 25–52.

Gaur, V., M. L. Fisher, A. Raman. 2005. An econometric analysis of inventory turnover performance in retail services. *Management Sci.* **51**(2) 181–194.

Hall, J., E. Porteus. 2000. Customer service competition in capacitated systems. *Manufacturing Service Oper. Management* **2**(2) 144–165.

Hendricks, K. B., V. R. Singhal. 2008. The effect of product introduction delays on operating performance. *Management Sci.* **54**(5) 878–892.

Hendricks, K. B., V. R. Singhal. 2009. Demand-supply mismatches and stock market reaction: Evidence from excess inventory announcements. *Manufacturing Service Oper. Management* **11**(3) 509–524.

I/B/E/S Glossary 2001. Thomson Financial, New York.

Ivkovic, Z., N. Jegadeesh. 2004. The timing and value of forecast and recommendation revisions. *J. Financial Econom.* **73**(3) 433–463.

Johnson, N. J. 1978. Modified *t* tests and confidence intervals for asymmetrical populations. *J. Amer. Statist. Assoc.* **73**(363) 536–544.

Kekre, S., K. Srinivasan. 1990. Broader product line: A necessity to achieve success? *Management Sci.* **36**(10) 1216–1231.

Liu, T.-C. 1960. Underidentification, structural estimation, and forecasting. *Econometrica* **28**(4) 855–865.

Lundholm, R. J., S. E. McVay, T. Randall. 2010. Forecasting sales: A model and some evidence from the retail industry. Working paper, http://ssrn.com/abstract=486162.

Maestri, N. 2010. Wal-Mart holiday sales dip: Investors eye traffic. Reuters.com (February 18), http://www.reuters.com/article/idUSTRE61H2TG20100218.

Mest, D. P., E. Plummer. 2003. Analysts' rationality and forecast bias: Evidence from sales forecasts. *Rev. Quant. Finance Accounting* **21**(2) 103–122.

Palepu, K. G., P. M. Healy, V. L. Bernard. 2004. *Business Analysis & Valuation: Using Financial Statements*, 3rd ed. South-Western College Pub, Mason, OH.

Raman, A., V. Gaur, S. Kesavan. 2005. David Berman. HBS Case 605-081, Harvard Business School, Boston.

Rumyantsev, S., S. Netessine. 2007. What can be learned from classical inventory models? A cross-industry exploratory investigation. *Manufacturing Service Oper. Management* **9**(4) 409–429.

Swaminathan, S., J. Weintrop. 1991. The information content of earnings, revenues, and expenses. *J. Accounting Res.* **29**(2) 418–427.

van Ryzin, G., S. Mahajan. 1999. On the relationship between inventory costs and variety benefits in retail assortments. *Management Sci.* **45**(11) 1496–1509.

Wooldridge, H. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.

INFORMS holds **copyright** to this article and distributed this copy as a courtesy to the author(s).
Additional information, including rights and permission policies, is available at http://journals.informs.org/.