Forecast UPC-Level FMCG Demand, Part II: Hierarchical Reconciliation

Dazhi Yang*, Gary S. W. Goh, Siwei Jiang and Allan N. Zhang
Singapore Institute of Manufacturing Technology (SIMTech)
*Agency for Science, Technology and Research (A*STAR)*
Singapore, Singapore
*Email: *yangdz@simtech.a-star.edu.sg*
yangdazhi.nus@gmail.com

Orkan Akcan
Antuit
Singapore, Singapore

***Abstract*—In a big data enabled environment, manufacturers and distributors may have access to previously unobserved retailer-level demand related information. This additional information can be considered in demand forecasting to produce more accurate forecasts, and thus enable better stock-outs management. In Part II of this two-part paper, we explore the hierarchical nature of fast moving consumer goods (FMCG) demand (represented by sales) time series and produce one week ahead rolling forecasts on universal product code (UPC) level (or distributor level, as per our definition below). We show that the hierarchical forecasting framework has significant accuracy improvement over the conventional univariate forecasting methods. The main reason of the observed improvements is due to the price and promotion information available at the retailer level, which is assumed to be unknown to the distributor. To reconcile forecasts according to the hierarchy, only the forecast values at retailer level are needed, the business strategies of individual retailers remain proprietary. A freely available dataset is considered to encourage further exploration. Data exploratory analysis and visualization tools are discussed in Part I of the paper.**

***Keywords*-FMCG; forecasting; hierarchical reconciliation; visualization**

I. Introduction

Demand forecasting at various horizons is essential for sales and operations planning (S&OP) process. Good forecasts provide strong decision support for operation management tasks such as capacity planning, inventory management and planning & scheduling [1]. We are interested in forecasting the demand of fast moving consumer goods (FMCG) on universal product code (UPC) level in this paper.

In a big data enabled environment, manufacturers and distributors may have access to previously unobserved retailer-level demand related information. This additional information enables potential integration of the hierarchical demand. The hierarchy should be considered in demand forecasting in order to produce more accurate forecasts at various levels of the hierarchy. Before we introduce hierarchical forecasting, literature review on demand forecasting for a single product in a single market is performed.

A. Literature Reivew

Generally speaking, the demand for a product is governed by economic theory. A typical retailer would collect panel data over several dimensions including items, stores, markets, categories. In addition, product characteristics and demographics are also frequently being recorded and updated. Theoretically, when individual consumer characteristics are matched to the goods they purchased, the complex and evolving relationships between consumer behavior and demand can be studied.

Econometrics theory has potential in integrating the above-mentioned data for demand prediction[1]. One such example is the random-coefficients logit model [2]. It considers the utility u_{ijt} of an individual consumer i from a product j in a market t:

$$u_{ijt} = x_{jt}\beta_i - \alpha_i p_{jt} + \xi_{jt} + \epsilon_{ijt} \tag{1}$$

where x and ξ are observed and unobserved product characteristics, p is the price of product, ϵ is a zero mean stochastic term and α and β are interpretable model coefficients. It can be seen from Eq. (1) that prices of different products in different markets, observed and unobserved product characteristics and demographics are used to model u_{ijt}. The demand can thus be estimated by integrating over the probability mass of the consumers who choose brand j in market t (see Ref. [2] for details). However, such econometrics models are mostly used for market share prediction. Furthermore, if we consider a particular market at different time t, Eq. (1) suggests that in order to forecast future utility, the future values of the predictors are needed. Due to the theoretical nature of the method and the complexity of the formulation, much work along this track is on-going, manufacturers, distributors and retailers often rely on simplistic but effective demand forecasting methods in their day-to-day operations.

[1]Prediction is more general than forecasting.

Alternative to econometrics demand models are time series models (or extrapolative methods). These models consider the evolution and dependence of the temporal process of demand. Such demand forecasting models rely on discerning demand patterns and are constructed based on previous observations. Simple exponential smoothing (SES) is one of the well-accepted models; it is shown that SES is efficient in capturing the level component in demand over time [3]. Given a demand time series $\{y_t : t \in D_t\}$, where $D_t = \{0, 1, 2, \cdots\}$ is the time indices, the weighted average form of SES is given by:

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha)\hat{y}_t \tag{2}$$

where symbol ˆ denotes an estimate and $0 \leq \alpha \leq 1$ is the smoothing parameter. Eq. (2) shows that the forecast at time $t+1$ is a weighted average of the most recent observation y_t and the most recent forecast $\hat{y}_t$. The deficiency of SES in FMCG demand prediction is thus apparent: when the characteristics of the demand series change due to special events such as promotions or holidays, the fitted parameter can no longer describe the series. Exogenous inputs or multivariate methods are needed to explain those special events [4].

The so-called "base-times-lift" model is one of the simplistic models that use exogenous variables [5]; it is commonly adopted by the industry [3]. The base-times-lift models first generates a baseline forecast using a model, such as the SES, for non-promoted time periods. A "lift effect" L is added to the baseline forecast during the promoted periods. Ref. [5] models L using promotional index $\mathcal{I}$, i.e., $\hat{L}_c = (\mathcal{I}_c/\mathcal{I}_p)L_p$, where subscripts c and p represent current and previous promotions. However, such linear assumption on lift effect has been shown to be sub-optimal in FMCG forecasting; it often performs worse than SES model [3]. Better representations of lift effect are thus desired.

In the literature, promotional demand is related to various factors. Ref. [6] suggests that the promotional demand, and thus sales, are mainly affected by promotion style, item promotion history and store promotion history. Ref. [7] considers deal types (promotion style), depth of deals, as well as variations across categories, time and brands. Among several alternatives, Ref. [3] presents a more general formulation of FMCG demand with the following predictors: (1) past values of demand of the focal product, (2) price of the focal product, (3) promotional index of the focal product, (4) price of the competitor products, (5) promotional indices of the competitor products, (6) monthly dummy variables and (7) dummies for calendar events; all future information (other than demand itself) is assumed to be known to the forecaster. These predictors are related to the future demand of the focal product through the autoregressive distributed lag (ADL) model:

$$\begin{aligned} \ln(y_{0,t}) =& \alpha_0 + \sum_{j=1}^{\ell} \alpha_j \ln(y_{0,t-j}) + \sum_{j=0}^{\ell} \beta_{0,j} \ln(p_{0,t-j}) \\ &+ \sum_{j=0}^{\ell} \gamma_{0,j}\mathcal{I}_{0,t-j} + \sum_{p=1}^{P}\sum_{j=0}^{\ell} \beta_{p,j} \ln(p_{p,t-j}) \\ &+ \sum_{q=1}^{Q}\sum_{j=0}^{\ell} \gamma_{q,j}\mathcal{I}_{q,t-j} + \sum_{d=1}^{12} \theta_d \text{monthly\_dummy}_d \\ &+ \sum_{c=1}^{C}\sum_{\nu=0}^{1} \delta_{c,\nu}\text{calendar\_event}_{c,t-\nu} + \varepsilon_t \end{aligned} \tag{3}$$

where ℓ is the maximum lag; P and Q are numbers of competitor products in terms of price and promotional index respectively; $\nu = 0$ corresponds to the calendar event week; $\nu = 1$ corresponds to the week before the calendar event week; C is the number of distinct calendar events in a year; the remaining symbols in Eq. (3) are self-explanatory. The values of P and Q are selected via the lasso (least absolute shrinkage and selection operator). It was shown that the ADL model outperforms both the SES and base-times-lift models.

The method proposed in Ref. [3] can be considered as the state-of-the-art forecasting model for FMCG. It reflects a strong belief in the effect of predictor variables in predicting the response variable. However, the ADL model shown in Eq. (3) has two major deficiencies for operational forecast:

1) The first deficiency lies within the modeling part. Suppose $C = 9$ and $P = Q = \ell = 2$ (which are the typical values [8]), the total number of predictors is 50. There is a high chance that the model is misspecified (e.g., retaining null of the t-test). This would lead to an additional regression parameter selection stage beside the original lasso step, i.e., lasso determines P and Q, and the additional parameter selection step reduces number of predictors from 50 to a smaller number. Without the additional parameter selection stage the insignificant predictors only contribute to the variance of the predicted value.
2) The exogenous factors such as promotion information is usually available at store level or retail enterprise level but not available to manufacturers and distributors. Therefore the ADL model is not applicable to manufacturers and distributors; they have to make forecasts based on the historical order placement information provided by the retailers.

Beside the above two major deficiencies, other issues such as missing data handling, scalability of the algorithm, lasso regression design and error evaluations in Ref. [3] also have room for improvements. However, we mainly address the

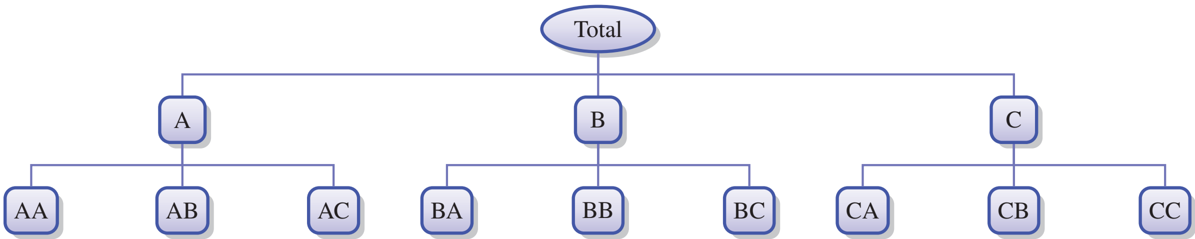

Figure 1. A three-level hierarchical time series structure.

second issue in this paper, i.e., improving FMCG unit sales[2] forecasting on universal product code (UPC) level, without using the retailer-level strategic information (such as an upcoming promotion).

B. Hierarchical nature of FMCG sales

FMCG unit sales data can be modeled as a hierarchical structure as according to how the UPCs are being distributed in a supply chain; Fig. 1 gives such an illustration. Level 0 (Total) denotes the completely aggregated distributor-level time series. Level 1 (A, B and C) shows the first level of disaggregation which could represents various UPCs handled by the distributor. Down to level 2 (AA, AB, $\cdots$, CC), it contains the most disaggregated time series, e.g. the demand at individual outlets. The dataset used in this paper is disaggregated first by product type then by geography, although other configurations of the hierarchical structure could be defined.

In a typical supply chain, distributors receive orders from the retailers and thus make forecasts based on historical orders. If we assume independent decision making among the retailers, price and promotion information is unobservable by the distributors. The forecasts made by the distributors are thus often based solely on the aggregated sales time series. However, such distributor-side forecasts are inconsistent with the sum of all retailer-side forecasts, due to the different forecast generating processes used by various parties. In an ideal environment where every retailer produce their own forecasts based on their business strategy and feed the results to the distributor, the forecasts in the supply chain need to be "aggregate consistent". Forecast reconciliation is therefore needed.

In fact, the problem herein discussed is related to information sharing in supply chains. In the field of management science, information sharing is not new [10], [11]. However, most studies are theoretical; much attention is focused on the effects of information sharing on operation management issues such as inventory management. Ref. [12] reviewed the current state-of-the-art on the value of sharing demand information. It is concluded that whether to share demand information depends on whether the downstream demand process can be inferred. At this stage, we consider the case where the downstream demand cannot be inferred. In addition, we note that the forecasts at retailer level are not actual orders. If the forecasts are the same as orders, providing such information to the distributed would have no practical relevance.

In this paper, we show that the hierarchical reconciliation using the retail-level data would improve UPC-level forecast accuracy significantly. The proposed forecasting framework needs collaborative efforts among the individual retailers and the distributor. Such collaboration protects proprietary information as only the forecasts are needed, without disclosure of business strategies. The paper is organized as follows: Section II introduces formulation and variants of the hierarchical reconciliation. As the name suggests, the method reconciles the forecasts instead of producing the forecasts; base forecasts are generated independently by various distributors and retailers prior to reconciliation. Section III illustrates our assumptions and approaches to produce base forecasts. Detailed results and discussions are shown in Section IV and Section V concludes the paper.

II. Hierarchical Reconciliation

In this section, hierarchical forecasting methods will be discussed based on the illustrative example shown in Fig. 1. More general description of the methods can be found in Refs. [13]–[15]. We note that the number of hierarchies and aggregation/disaggregation interpretation may vary based on the granularity of the available research data.

The underlying principle of hierarchical prediction is originated from a summing matrix $\boldsymbol{S}$. Suppose $\boldsymbol{Y}_{i,t}$ denotes all observations at level i and time t and $\boldsymbol{Y}_t = (Y_{0,t}, \boldsymbol{Y}_{1,t}, \boldsymbol{Y}_{2,t})^\top$, then:

$$\boldsymbol{Y}_t = \boldsymbol{S}\boldsymbol{Y}_{2,t} \tag{4}$$

[2]Despite the differences between demand and sales [9], we consider using unit sales to represent demand and evaluate various models' performance based on sales data.

where

$$
\boldsymbol{S}=\begin{pmatrix}
1&1&1&1&1&1&1&1&1\\
1&1&1&0&0&0&0&0&0\\
0&0&0&1&1&1&0&0&0\\
0&0&0&0&0&0&1&1&1\\
1&0&0&0&0&0&0&0&0\\
0&1&0&0&0&0&0&0&0\\
0&0&1&0&0&0&0&0&0\\
0&0&0&1&0&0&0&0&0\\
0&0&0&0&1&0&0&0&0\\
0&0&0&0&0&1&0&0&0\\
0&0&0&0&0&0&1&0&0\\
0&0&0&0&0&0&0&1&0\\
0&0&0&0&0&0&0&0&1
\end{pmatrix} \tag{5}
$$

The summing effect is apparent. The bottom-level time series will sum into various time series in the hierarchy based on various rows in $\boldsymbol{S}$. Suppose $\widehat{\boldsymbol{Y}}_{t+h} = (\widehat{Y}_{0,t+h}, \widehat{\boldsymbol{Y}}_{1,t+h}, \widehat{\boldsymbol{Y}}_{2,t+h})^\top$ denotes the h-step-ahead base forecasts (see below), all hierarchical forecasting methods can be written as:

$$\widetilde{\boldsymbol{Y}}_{t+h} = \boldsymbol{S}\boldsymbol{P}\widehat{\boldsymbol{Y}}_{t+h} \tag{6}$$

where $\boldsymbol{P}$ is a matrix of choice to reconcile forecasts, $\widetilde{\boldsymbol{Y}}_{t+h}$ is the final revised hierarchical forecasts.

In general, there are four ways to reconcile the forecasts, namely, the top-down approach, the bottom-up approach, the middle-out approach and optimal reconciliation [14]. For example, if $\boldsymbol{P} = (\boldsymbol{0}_{9\times4}|\boldsymbol{I}_9)$, where $\boldsymbol{0}_{i\times j}$ is the $i\times j$ null matrix and $\boldsymbol{I}_9$ is a size 9 identity matrix, Eq. (6) represents the bottom-up approach. In this case $\boldsymbol{SP} = (\boldsymbol{0}_{13\times4}|\boldsymbol{S})$, the times series at levels 0 and 1 are simple sums of the bottom-level series. Similarly, the top-down approach corresponds to $\boldsymbol{P} = (\boldsymbol{p}|\boldsymbol{0}_{9\times12})$ where $\boldsymbol{p} = (p_1, \cdots, p_9)^\top$ is a vector of "weights" of the bottom-level series.

As compared to the bottom-up and top-down approaches, much effort is needed to derive the representation for optimal reconciliation. We refer the interested readers to Ref. [13] for details; we only state the main results here. The general idea is derived from the representation of the h-step-ahead base forecasts by a linear regression model:

$$\widehat{\boldsymbol{Y}}_{t+h} = \boldsymbol{S}\boldsymbol{\beta}_h + \boldsymbol{\epsilon}_h \tag{7}$$

where $\boldsymbol{\epsilon}_h$ has zero mean and covariance $\boldsymbol{\Sigma}_h$; $\boldsymbol{\beta}_h = \mathbb{E}(\widehat{\boldsymbol{Y}}_{2,t+h})$, i.e., $\boldsymbol{\beta}_h$ is the expectation of the base forecasts. Ref. [13] showed that under a reasonable assumption[3], the optimal reconciled forecasts are given by the generalized least square (GLS) solution:

$$\widetilde{\boldsymbol{Y}}_{t+h} = \boldsymbol{S}\widehat{\boldsymbol{\beta}}_h = \boldsymbol{S}(\boldsymbol{S}^\top\boldsymbol{S})^{-1}\boldsymbol{S}^\top\widehat{\boldsymbol{Y}}_{t+h} \tag{8}$$

[3]Sometimes, it is reasonable to assume $\boldsymbol{\epsilon}_h \approx \boldsymbol{S}\boldsymbol{\epsilon}_{2,h}$, where $\boldsymbol{\epsilon}_{2,h}$ is the level 2 forecast errors in our case, or more generally, the bottom-level forecast errors.

which is equivalent to the ordinary least square (OLS) solution. Note that $\widehat{\boldsymbol{\beta}}_h = (\boldsymbol{S}^\top\boldsymbol{S})^{-1}\boldsymbol{S}^\top\widehat{\boldsymbol{Y}}_{t+h}$ in Eq. (8) is the best linear unbiased estimator [14]. In this case, $\boldsymbol{P}$ in Eq. (6) is $(\boldsymbol{S}^\top\boldsymbol{S})^{-1}\boldsymbol{S}^\top$. Therefore, we only need to construct the summing matrix and produce the base forecasts in order to obtain the optimal reconciled forecasts. On the other circumstances, if the earlier assumption does not hold, GLS solution is:

$$\widetilde{\boldsymbol{Y}}_{t+h} = \boldsymbol{S}(\boldsymbol{S}^\top\boldsymbol{\Sigma}_h^\dagger\boldsymbol{S})^{-1}\boldsymbol{S}^\top\boldsymbol{\Sigma}_h^\dagger\widehat{\boldsymbol{Y}}_{t+h} \tag{9}$$

where $\boldsymbol{\Sigma}_h^\dagger$ is the generalized inverse of the error covariance matrix. However, $\boldsymbol{\Sigma}_h$ is not known and very difficult[4] (or impossible) to estimate for large hierarchies [15]. The weighted least squares (WLS) solution given by:

$$\widetilde{\boldsymbol{Y}}_{t+h} = \boldsymbol{S}(\boldsymbol{S}^\top\boldsymbol{\Lambda}_h\boldsymbol{S})^{-1}\boldsymbol{S}^\top\boldsymbol{\Lambda}_h\widehat{\boldsymbol{Y}}_{t+h} \tag{10}$$

may be appropriate. $\boldsymbol{\Lambda}_h$ in Eq. (10) is a diagonal matrix with elements equal to the inverse of the variances of $\boldsymbol{\epsilon}_h$.

III. Producing Base Forecasts

Following Eq. (6), the collection of base forecasts is needed to reconcile the forecast according to the hierarchy. We consider the Dominick's database in this paper; the dataset is provided by the James M. Kilts Center, University of Chicago Booth School of Business with a collaborative effort by the Dominick's Finer Food (DFF). In Part I of this two-part paper [16], data exploratory analysis, visualization and preprocessing are described in details. However, we would make necessary reiterations in Part II, so that the present paper can be relatively independent. At this stage, we note that 37 UPCs (level 1 time series) from the bottled juice category are selected to demonstrate hierarchical forecasting.

A. General methods versus specific methods

Univariate statistical forecasting methods such as the autoregressive integrated moving average models (ARIMA) and exponential smoothing state space models (ETS) are general; they can arguably be applied to any time series and generate forecasts (see Refs. [17], [18] for details of ARIMA and ETS respectively). These univariate methods do not consider the effects of exogenous factors. In FMCG forecasting, the exogenous factors such as price and promotion are well-studied. There is little argument we can make on not using these factors when data are available. Forecasting methods which consider domain knowledge are specific methods.

[4]In general, given the symmetrical structure of $\boldsymbol{\Sigma}_h$, there are $n(n+1)/2$ parameters to be estimated in a GLS setting, which is difficult especially when n is large. This is usually counteracted by imposing some structure on $\boldsymbol{\Sigma}_h$, so that the number of parameters to be estimated is much smaller. Even though, the appropriateness of such structure needs to be carefully examined.

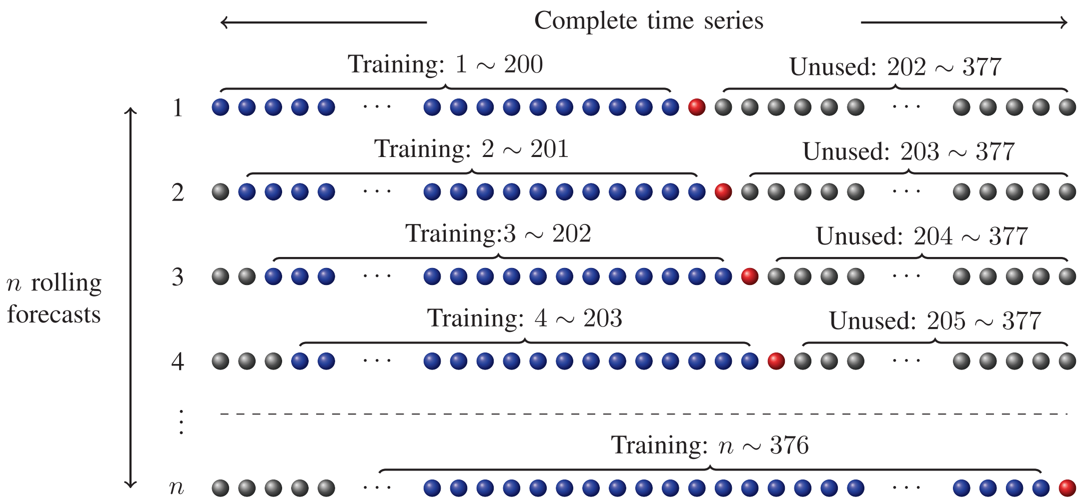

Figure 2. Experimental design: 177 ($n = 177$ in this case) rolling forecasts are performed for each time series. The data point being forecast is shown in red. The size of training window (shown in blue) is fixed at 200.

In this paper, we consider using SES (equivalent to an ARIMA(0,1,1) model) to represent the general methods. This is because our time series do not process strong seasonal and trend components. For specific methods, we consider using the ADL model. Instead of following the fully specified Eq. (3), a reduced form is used:

$$\begin{aligned}\ln(y_{0,t}) =&\alpha_0 + \sum_{j=1}^{\ell} \alpha_j \ln(y_{0,t-j}) + \sum_{j=0}^{\ell} \beta_{0,j} \ln(p_{0,t-j}) \\ &+ \sum_{j=0}^{\ell} \gamma_{0,j} \mathcal{I}_{0,t-j} + \varepsilon_t \end{aligned} \tag{11}$$

To support our model of choice, we use a data visualization technique in Ref. [16]. It is shown that the calendar events (9 holidays) and 12 months in a year have minimal correlation with the demand fluctuation. The monthly dummy variables and dummy variables for calendar event in Eq. (3) are thus dropped.

Besides the dummy variables, the full ADL model shown in Eq. (3) also considers the promotional index and price of the competitor products as predictors. In a separate analysis we found that given the 37 selected UPCs, including the lasso-identified competitors only add to the prediction variance but not accuracy. An obvious reason for such observation is that our products are from the same category, the complimentary effect (e.g. bread and butter) of FMCG sales is not applicable. The substitution effect (e.g. eggs from different farms) is also hypothesized to be minimal due to the relatively persistent price over the years, and thus the consumer loyalty. Therefore, we use Eq. (11) in this paper to optimize the trade-off between prediction variance and accuracy by only considering the promotional index and price information of the focal product.

B. *The pitfalls of using averaged information*

Ideally, if all the retailer-level business strategies for a particular UPC is available to the distributor, the distributor would make an informed decision on overall price and promotional index, and thus calculate the price curves and promotion stripes shown in Fig. 2 of Ref. [16]. The UPC-level forecasts can then be made using ADL models such as Eq. (11). However, as mentioned in point 2) Section I, such information is usually not available to distributors; applying ADL on UPC level may not be practically viable. In addition, as all the stores in the DFF dataset belong to the same retailer group, the retailers tend to adopt a uniform pricing strategy where they increase or decrease the price for all the stores at the same time [19]. When the pricing strategies are decided independently and the number of independent retailers is large, the price and promotion information on UPC level may be redundant (e.g., each and every week there may be several retailers making promotions so that the overall promotional index is "flat" throughout a year). In such scenarios, the distributor would rely on simple time series models to make forecasts. Following this assumption, we only use SES to forecast the UPC-level sales, i.e., forecast the level 1 time series; whereas for the retailer-level sales where price and promotion information is more readily available, ADL is used.

C. *Experimental design*

An ad hoc data preprocessing sequence is used in Ref. [16] to yield a total of 1971 store-level time series

Table I
FORECAST ERRORS (MAPE) OF 37 SELECTED UPCS FROM THE BOTTLED JUICE CATEGORY. THE ERRORS ARE IN PERCENTAGE.

	UPC	**Product name**	**No. of series**	**Univariate methods**		**Hierarchical methods**		
				Persistence	SES	Bottom-up	Optimal (OLS)	Optimal (WLS)
1	7045011484	SUNSWEET PRUNE JUIC 48Oz	55	10.16	10.74	9.02	10.73	**8.65**
2	5300015154	REALEMON PLASTIC LEM 45Oz	61	11.17	9.64	9.00	11.03	**8.92**
3	5300015108	REALEMON LEMON JUICE 8Oz	64	15.11	13.05	10.89	14.98	**10.89**
4	3828103123	HH LEMON JUICE 32Oz	61	16.74	16.92	18.32	**16.55**	17.50
5	3828103091	HH APPLE JUICE 128Oz	66	31.81	37.66	20.92	31.40	**20.86**
6	3828103025	DOM APPLE JUICE 32Oz	52	39.11	41.96	21.97	38.40	**21.68**
7	3828103021	HH APPLE JUICE 48Oz	63	62.84	182.55	34.23	61.84	**33.88**
8	3120027407	OS PINK GRAPEFRUIT 64Oz	42	29.99	34.58	17.61	30.04	**17.51**
9	3120027007	OS GRAPEFRUIT JUICE 64Oz	53	28.16	30.34	12.68	27.94	**12.59**
10	3120026134	OS LC CRANRASP 48Oz	52	26.11	24.66	**12.26**	26.22	12.33
11	3120021007	OS CRANAPPLE DRINK 64Oz	62	25.18	27.01	12.21	24.74	**12.12**
12	3120020035	OS LO CAL CRANBERRY 48Oz	63	21.33	20.29	10.49	21.37	**10.43**
13	3120020007	OS CRNBRY JCE COCKTA 64Oz	65	26.15	30.91	11.67	25.86	**11.51**
14	3120020005	OS CRANBERRY COCKTA 48Oz	66	34.65	50.54	12.46	34.38	**12.41**
15	1480031656	MOTTS NATURAL APPLE 64Oz	65	74.67	55.48	20.62	73.62	**20.57**
16	1480000034	MOTTS REGULAR APPLE 64Oz	67	85.04	73.67	**26.49**	83.91	27.72
17	7045011329	SUNSWEET PRUNE JUICE 32Oz	61	12.28	11.99	12.10	12.32	**11.70**
18	5300015132	REALEMON LEMON JUICE 32Oz	66	16.55	17.49	17.05	16.57	**16.05**
19	4850000193	TROP TWSTR ORGCRAN 46Oz	34	30.35	29.55	16.27	30.06	**16.20**
20	4180022700	WELCHS WHITE GRAPE J 64Oz	54	10.49	11.11	16.51	10.36	**15.14**
21	4180020750	WELCHS GRAPE JUICE 64Oz	65	11.21	11.11	14.87	11.41	**13.89**
22	4176000394	INDIAN SUMMER APPLE 64Oz	64	141.24	262.59	31.19	139.11	**30.83**
23	3828103017	HH APPLE JUICE 64Oz	67	66.32	76.37	25.33	65.51	**25.16**
24	3828103009	HH PRUNE JUICE 40Oz	40	15.72	15.93	13.47	16.14	**12.91**
25	3120027005	OS GRAPEFRUIT JUICE 48Oz	39	29.75	30.56	16.69	29.61	**16.60**
26	3120026107	OS CRANRASPBERRY DR 64Oz	56	28.70	33.27	12.07	28.38	**11.97**
27	3120026105	OS CRANRASPBERRY 48Oz	45	43.58	71.13	**20.12**	43.51	20.30
28	3120021005	OS CRANAPPLE DRINK 48Oz	41	41.11	55.02	18.67	41.08	**18.65**
29	3828103115	DOM CRANBERRY RSPBRY 48Oz	7	33.58	33.57	27.12	37.00	**26.60**
30	3828103033	HH CRANBERRY JUICE C 48Oz	61	23.28	24.27	18.61	22.94	**18.45**
31	3828103005	DOM PRUNE JUICE 32Oz	41	16.49	15.89	15.48	16.62	**14.63**
32	3120027405	OS PINK GRAPEFRUIT 48Oz	26	31.49	36.20	19.40	31.97	**19.33**
33	1480000032	MOTTS APPLE JUICE P 32Oz	62	9.57	**9.07**	11.71	9.62	10.60
34	5300015407	REAL LIME JUICE 8Oz	7	24.35	22.16	22.72	25.99	**21.44**
35	7045011402	SUNSWEET PRUNE JCE W 40Oz	58	12.53	14.76	11.56	12.43	**11.00**
36	3120020000	OS CRANBERRY COCKT 32Oz	59	10.99	**9.96**	10.50	11.06	10.20
37	1480051324	MOTTS CLAMATO JUICE 32Oz	61	14.10	**12.69**	15.27	14.10	14.75
	overall		1971	31.40	39.59	16.96	31.32	**16.65**

for 37 UPCs over a period of 377 weeks. For each of the 2009 time series (1, 37 and 1971 time series at levels 0, 1 and 2 respectively), we produce 177 rolling forecasts with a moving window of 200 weeks of training data. In other words, the first 200 data points are used in model fitting to produce the first forecast; when the 'new data' becomes available, we refit the model using the new window of 200 weeks and produce another forecast. An illustration of the experimental design is depicted in Fig. 2. An alternative to the present approach is to use a fixed model throughout the evaluation period. It is found that some time series have a gradual change in level component; the iterative approach could capture the gradual change and thus make better forecasts. We also note that it is possible to perform experiments to investigate the effect of training length on forecasting accuracy, however, we limit our choice of training length of 200 following Ref. [3]. In forecasting studies, another commonly evaluated parameter is the forecast horizon. For the DFF dataset, it was shown that the difference between a 1-week-ahead and 12-weeks-ahead forecasting accuracies is small [3].

There is a rich literature on choice of error metrics in forecasting. In this study we choose the scale-independent mean average percentage error (MAPE), which is most widely used in practice [20].

IV. Results and discussion

We apply both SES and persistence models on the 37 UPC-level aggregated series. Persistence model assumes the forecast is equal to the current observation; it is often included as a naive benchmark. Table I shows the results. It is observed that the overall performance of the persistence model is better than that of the SES, although SES gives improved results for some UPCs. Fig. 3 shows

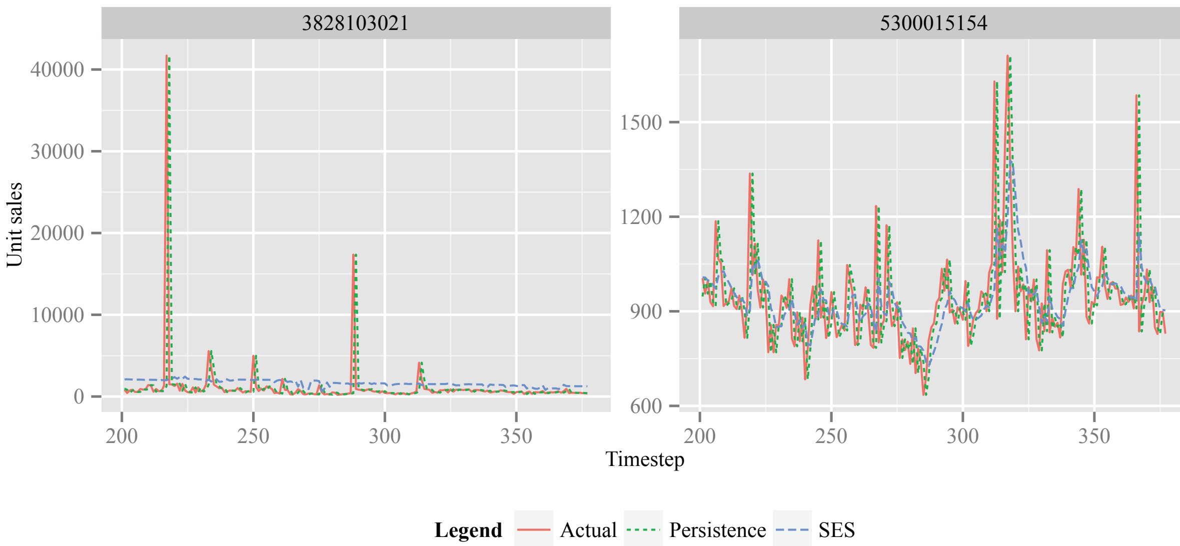

Figure 3. Time series plot of forecast values using univariate models for 2 example UPCs.

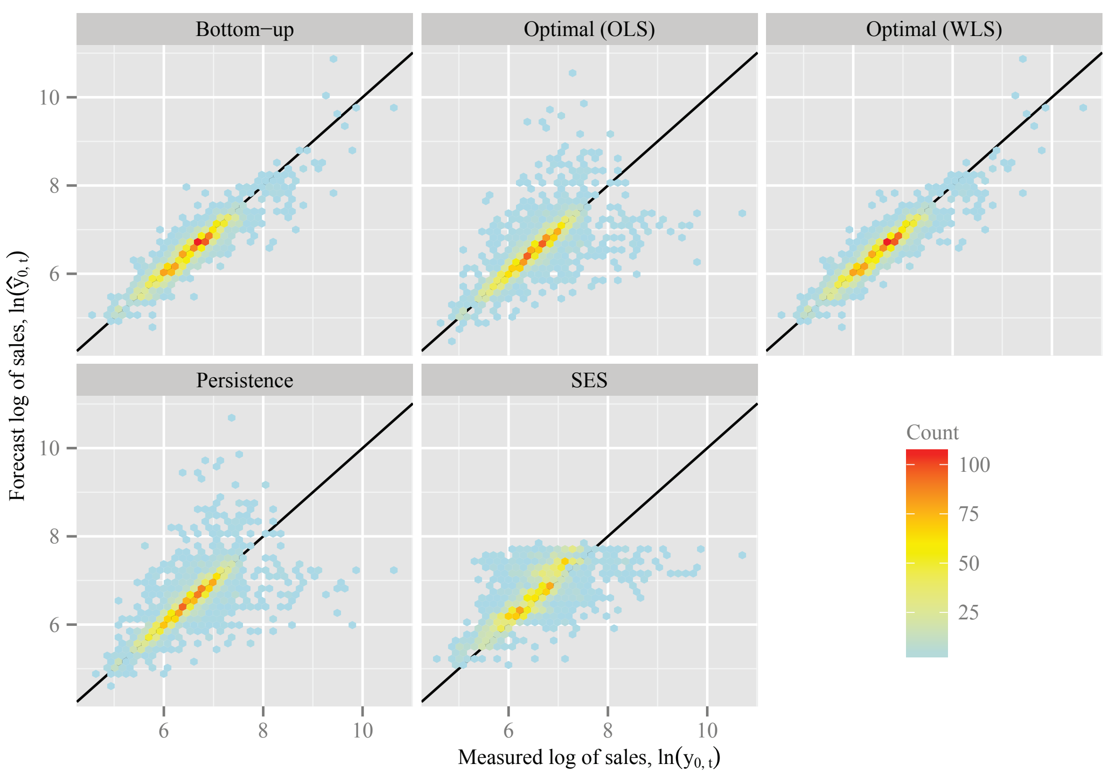

Figure 4. Scatter plots of the measured and forecast log of unit sales. For each forecasting method, all forecast points from the 37 UPCs are included ($177 \times 37 = 6549$ points in each sub-plot). Hexagon binning algorithm is used for visualization.

the forecast results time series plot for 2 selected UPCs. The relatively big SES forecast error (182.55% for SES; 62.84% for persistence) for UPC-3828103021 originates from the misidentified level component (see left plot). The misidentification is caused by the large sales peaks present in the data. On the other hand (right plot), when the data fluctuates about a level, SES gives slightly better results owing to its smoothing property.

Recall in Section II, we introduced several variants of hierarchical forecasting, including top-down, bottom-up and optimal approaches. For FMCG data, it is well-known that information loss is substantial in aggregation and therefore the top-down method may not be suitable. Furthermore, given our hierarchical structure, it is not easy to assign weights to the middle- and bottom-level series. For example, we need to know the market share of each UPC in order to assign weights to the middle-level series. We therefore consider bottom-up and optimal approaches for hierarchical forecasting. Both OLS and WLS but not GLS are used in optimal reconciliation for obvious reasons stated in Section II. Since persistence is superior to SES for our particular dataset, we use results from the persistence model (for levels 0 and 1 only; level 2 series is forecast using ADL models) during reconciliation. The results are tabulated in Table I. To further understand the errors, the scatter plot of measured versus forecast unit sales for each forecasting method is shown in Fig. 4. Note that both the ordinate and abscissa are on log-scale; hexagon binning algorithm is used for visualization.

From Table I and Fig. 4, both bottom-up and optimal (WLS) approaches show significant improvements over the univariate methods, whereas the optimal (OLS) reconciliation gives similar results to the persistence model. The unsatisfactory performance of OLS-based reconciliation is caused by the heteroscedasticity in the data. The bottom-up and optimal (WLS) approaches do not distinguish from each other in this case study. It is believed that the good performance of the bottom-up approach can be attributed to the nature of the data. Even at the very bottom level, the data is well behaved, with price and promotion driven unit sales for most series. It is also hypothesized that when the number of time series and number of levels in the hierarchy scale up, the optimal reconciliation would be more advantageous than the bottom-up approach.

V. Conclusion

We perform one-week-ahead UPC-level FMCG sales forecasting by considering the inherent hierarchical structure of the distributor-retailer relationship in a supply chain. The optimally reconciled hierarchical forecasts using WLS is shown to be the most promising as compared to the other benchmarking models. The WLS approach is able to achieve an overall of 58% lower MAPE over SES. With these more accurate forecasts and without the need for either distributors or retailers to divulge sensitive business information, it depicts a win-win situation for this supply chain cooperative model to be both efficient and practical.

In future, we will further explore hierarchical forecasting (for supply chain applications) by considering an extended hierarchy, i.e., including manufacturers/suppliers and/or products from different categories. To improve scalability, the recursive calculation of $(\boldsymbol{S}^\top \boldsymbol{\Lambda}_h \boldsymbol{S})^{-1}$ as shown in Ref. [15] can be employed. Other exogenous information such as sentiment scores from online product reviews as well as marketing and advertising strategies will be explored to evaluate whether they can improve forecast accuracy. In addition, probabilistic forecasts instead of the current point forecasts will be investigated. We will also attempt to address the computational challenge in estimating the variance-covariance structure of the base forecast, so that a GLS solution to the hierarchical reconciliation can be employed.

Acknowledgment

This work is partially supported under the A*STAR TSRP fund 1424200021 and Antuit–SIMTech Supply Chain Analytics Lab.

References

[1] D. Samson and P. J. Singh, *Operations Management*. Cambridge University Press, 2008, cambridge Books Online. [Online]. Available: http://dx.doi.org/10.1017/CBO9781139150002

[2] A. Nevo, "Measuring market power in the ready-to-eat cereal industry," *Econometrica*, vol. 69, no. 2, pp. 307–342, Mar. 2001.

[3] T. Huang, R. Fildes, and D. Soopramanien, "The value of competitive information in forecasting FMCG retail product sales and the variable selection problem," *European Journal of Operational Research*, vol. 237, no. 2, pp. 738–748, 2014.

[4] R. Fildes, K. Nikolopoulos, S. F. Crone, and A. A. Syntetos, "Forecasting and operational research: a review," *Journal of the Operational Research Society*, vol. 59, no. 9, pp. 1150–1172, May 2008.

[5] O. G. Ali, S. Sayin, T. van Woensel, and J. Fransoo, "SKU demand forecasting in the presence of promotions," *Expert Systems with Applications*, vol. 36, no. 10, pp. 12 340–12 348, Dec. 2009.

[6] L. G. Cooper, P. Baron, W. Levy, M. Swisher, and P. Gogos, "Promocast$^{\mathrm{TM}}$: A new forecasting method for promotion planning," *Marketing Science*, vol. 18, no. 3, pp. 301–316, 1999.

[7] S. Bogomolova, S. Dunn, G. Trinh, J. Taylor, and R. J. Volpe, "Price promotion landscape in the US and UK: Depicting retail practice to inform future research agenda," *Journal of Retailing and Consumer Services*, vol. 25, pp. 1–11, Jul. 2015.

[8] T. Huang, personal communication, 2015.

[9] S. A. Conrad, "Sales data and the estimation of demand," *Operational Research Quarterly (1970-1977)*, vol. 27, no. 1, pp. 123–127, 1976.

[10] C. Terwiesch, Z. J. Ren, T. H. Ho, and M. A. Cohen, "An Empirical Analysis of Forecast Sharing in the Semiconductor Equipment Supply Chain," *Management Science*, vol. 51, no. 2, pp. 208–220, Feb. 2005.

[11] H. L. Lee, K. C. So, and C. S. Tang, "The Value of Information Sharing in a Two-Level Supply Chain," *Management Science*, vol. 46, no. 5, pp. 626–643, May 2000.

[12] M. M. Ali and J. Boylan, "On the value of sharing demand information in supply chains," in *OR56 Annual Conference*, London, 2014, pp. 44–56.

[13] R. J. Hyndman, R. A. Ahmed, G. Athanasopoulos, and H. L. Shang, "Optimal combination forecasts for hierarchical time series," *Computational Statistics & Data Analysis*, vol. 55, no. 9, pp. 2579–2589, Sep. 2011.

[14] G. Athanasopoulos, R. A. Ahmed, and R. J. Hyndman, "Hierarchical forecasts for Australian domestic tourism," *International Journal of Forecasting*, vol. 25, no. 1, pp. 146–166, Jan. 2009.

[15] R. J. Hyndman, A. Lee, and E. Wang, "Fast computation of reconciled forecasts for hierarchical and grouped time series," 2014, working paper.

[16] D. Yang, G. S. W. Goh, C. Xu, and A. N. S. Zhang, "Forecast UPC-level FMCG demand, Part I: Exploratory analysis and visualization," in *Big Data (Big Data), 2015 IEEE International Conference on*, Oct 2015.

[17] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*. Englewood Cliffs-New Jersey: Prentice Hall, Inc., 1994.

[18] R. J. Hyndman, A. B. Koehler, J. K. Ord, and R. D. Snyder, *Forecasting with Exponential Smoothing*. Deblik, Berlin, Germany: Springer, 2008.

[19] S. J. Hoch, B.-D. Kim, A. L. Montgomery, and P. E. Rossi, "Determinants of store-level price elasticity," *Journal of Marketing Research*, vol. 32, no. 1, pp. 17–29, 1995.

[20] R. Fildes and P. Goodwin, "Fine judgements: do organizations follow best practice when applying management judgement to forecasting?" *Interfaces*, vol. 37, pp. 570–576, 2007.