What is the Optimal Strategy of Aggregation for Forecasting Sales? Time Series Forecast Reconciliation by Region, Product Category, and Channel

Autoria

Carla Freitas Silveira Netto - carla.netto@gmail.com
Prog de Pós-Grad em Admin/Esc de Admin - PPGA/EA/UFRGS - Universidade Federal do Rio Grande do Sul

Rob J. Hyndman - Rob.Hyndman@monash.edu
Department of Econometrics & Business Statistics/Monash University

Vinicius Andrade Brei - brei@ufrgs.br
Prog de Pós-Grad em Admin/Esc de Admin - PPGA/EA/UFRGS - Universidade Federal do Rio Grande do Sul

Agradecimentos

The first and the third authors have a scholarship from CNPq - National Council for Scientific and Technological Development - Brazil. We also would like to acknowledge Chair Tramontina Eletrik that provided the data.

Resumo

While some companies still struggle to gather, store and analyze data necessary to make better predictions, others are worried about increased requirements for data minimization and anonymization. This scenario raises questions about which variables are important to gather, and the resources necessary to do so. An important topic for academia and practice is how to improve forecasts, having limited access to data. In this paper, we consider such difficulties and propose forecasting strategies based on sales data in different aggregations criteria and structures. We compare aggregation criteria using both hierarchical and grouped time series structures, applying data that most companies already have access, marketing mix variables. Our paper indicates whether product category, channel type or region (geographic location) works best alone or combined when using the optimal reconciliation approach. This research suggests how to run sales forecasting more efficiently, using open-source tools. The method is also generalizable to all types of goods.

What is the Optimal Strategy of Aggregation for Forecasting Sales? Time Series Forecast Reconciliation by Region, Product Category, and Channel

Abstract: While some companies still struggle to gather, store and analyze data necessary to make better predictions, others are worried about increased requirements for data minimization and anonymization. This scenario raises questions about which variables are important to gather, and the resources necessary to do so. An important topic for academia and practice is how to improve forecasts, having limited access to data. In this paper, we consider such difficulties and propose forecasting strategies based on sales data in different aggregations criteria and structures. We compare aggregation criteria using both hierarchical and grouped time series structures, applying data that most companies already have access, marketing mix variables. Our paper indicates whether product category, channel type or region (geographic location) works best alone or combined when using the optimal reconciliation approach. This research suggests how to run sales forecasting more efficiently, using open-source tools. The method is also generalizable to all types of goods.

Keywords: sales; forecasting; marketing; hierarchical time series; grouped time series

1 INTRODUCTION

To plan and deliver products and services, it is necessary to know what the future might hold. Sales data are among the most important types of data related to future business performance, and sales forecasts are important to the most basic processes in any organization (ARMSTRONG, 2001; FILDES et al., 2008; SEAMAN, 2018). Therefore, forecasting accuracy is not just an important topic for academics, to develop, test, and improve their methods. It is also a major concern for practitioners and for marketing strategy. For example, Seaman (2018) illustrates that an accurate forecast is important for retailer pricing and distribution strategies. Errors in sales forecasts can lead to products being out of stock (leading customers to buy from competitors) or produced in excess (leading to holding costs and price promotions to increase sales). The consequences of out-of-stock products or excess price promotions are well-known, especially in the marketing literature. They include harms in brand reputation, quality perception, customer loyalty, repurchase intentions, and satisfaction. It is arguable that being able to forecast future purchase totals is more valuable to marketing than for other fields in social sciences (CHINTAGUNTA; NAIR, 2011).

Much of the marketing literature focuses on disaggregate analysis using data sets from companies that have access to individual-level information (CHINTAGUNTA; NAIR, 2011). However, most companies do not have easy access to this type of individual-level data. Gathering, storing, and analysing this type of data is complex, expensive, and time-consuming. This difficulty is perceived by small and medium-sized enterprises (SMEs), that do not have the resources to handle such data. This is also true for offline business that do not have access to the end customer data, and B2B companies, such as low-involvement, durable goods producers (CHEN; STECKEL, 2012; HANSSENS, 1998). Not to mention the increasing concerns and costs associated with ensuring the privacy of consumer information, a task that is becoming a legal responsibility and a strategic asset in building relationships with customers (WEDEL; KANNAN, 2016). This scenario of lack of access, coupled with increased requirements for data minimization and anonymization, makes access to individual-level data increasingly difficult. A key theoretical and empirical question is how to make better plans and decisions, having limited access to resources and, possibly, to the data itself. By considering such difficulties, this paper proposes a sales forecasting strategy based on proprietary data to which most companies already have access, that is, sales data aggregated, using different criteria and structures.

The decision of which strategy of data aggregation to use has been a subject of discussion for many years in different fields of knowledge, such as marketing (e.g. ABHISHEK; HOSANAGAR; FADER, 2015; RUSSELL; KAMAKURA, 1994; TELLIS; FRANSES, 2006), operations research (FLIEDNER; LAWRENCE, 1995), statistics (DUNN; WILLIAMS; DECHAINE, 1976), and economics (ZOTTERI; KALCHSCHMIDT; CANIATO, 2005). Sales can also be forecasted at many levels of aggregation, using different criteria to divide those levels, such as geographic considerations, product categories, types of channels, among others. This is necessary to plan organizational budgets more precisely (KREMER; SIEMSEN; THOMAS, 2015) and to help managers to diagnose forecast errors in more detail (DIVAKAR; RATCHFORD; SHANKAR, 2005). This will determine not only the level of production and distribution for each product, store and location, but also the allocation of promotion resources, sales representatives, and even the amount of energy firms will spend on each business partner. As Fliedner (2001) and Seaman (2018) state, different departments and hierarchical levels in a firm have different, but related, interests when it comes to sales forecasts. For example, a retail chain store needs to access the forecasts of its own departments. On the other hand, its head office needs a more aggregated level forecast to plan the chains' strategy. The different forecasting needs must be delivered as part of a forecasting system. Further, they should be "coherent", meaning that disaggregated forecasts should add-up to the total forecasts, in the same way as the historical data (HYNDMAN et al., 2011). These coherent forecasts, either by level or groups, can add accuracy to marketing efforts on those products, partners, departments, regions, channels, that are more likely to be profitable. Therefore, theory should demonstrate which type of data and aggregation strategy produce the best sales forecasting results.

In this paper, we show which information (product category, channel type, or geographic location) combined with different structures (hierarchical or grouped time series) improves forecast accuracy the most. This, to the best of our knowledge, has not yet been addressed. Thus, the goal of this paper is to test which type of marketing variables works best when using the optimal reconciliation approach to forecast sales. The paper makes the following contributions. First, we advance the knowledge about sales forecasting by comparing different levels of aggregation and reconciling them through trace minimization (WICKRAMASURIYA; ATHANASOPOULOS; HYNDMAN, 2018). We propose an aggregation strategy that can lead to better decisions regarding budget allocation, leading to more precise actions at points-of-sales and improved results. Second, we show that such a strategy can be used to improve sales forecast accuracy, by combining information that is easily available to firms. We demonstrate that it can be applied across many levels of aggregation, is easy to implement, and is relevant to practice. We also contribute to the management literature by providing a forecasting approach that uses what is one of the most basic type of variables existing in our field, that is, marketing mix variables. Third, we demonstrate our approach through a large-scale forecast reconciliation study, with several aggregation criteria using both hierarchical and grouped time series. Wedel and Kannan (2016) state that future studies should focus on models that are easy to implement in practice. Our last contribution follows this recommendation, by using automatic forecasting tools implemented in open-source software, based on data easily accessible to companies. Our strategy is generalizable to all types of goods.

This paper unfolds as follows. Section 2 presents the concepts and notations of grouped and hierarchical time series, the different aggregation criteria studied in literature, and the methods for forecast reconciliation. Section 3 presents the data and details of the methods' implementation. Section 4 describes how forecast accuracy of the different combinations of structures and aggregation criteria was evaluated and show the results. We conclude in section

5, by offering a discussion of the results and its implications to practice, followed by suggestions for future research.

2 THEORY

Marketing research largely focuses on cross-sectional, structured, quantitative, and causal models that can help explain how the market responds to changes in the marketing mix. One of the most studied models in marketing is the discrete-choice class of models. They are popular because they can be causally interpreted, but also because much of micro data in marketing is about consumer choice from a fixed set of alternatives within a category (CHINTAGUNTA; NAIR, 2011).

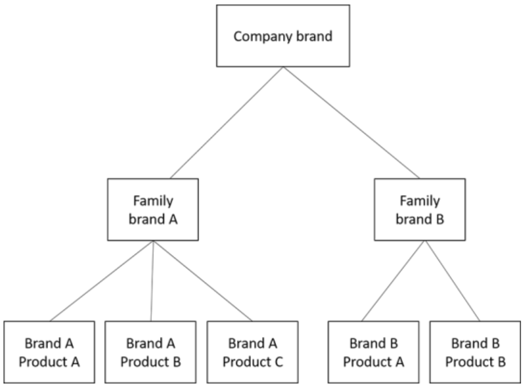
Even when research in marketing focus on prediction rather than explanation, the results are challenging to marketing practice since their implementation is not easy. As stated by Fildes et al. (2008, p.1165), "evidence of improved accuracy is lacking [… and] there remains a need for simple operational models that include key marketing instruments and that are downwardly compatible in the product hierarchy (from category to brand to SKU)". Because forecasting accuracy is more important in practice than understanding the impact of each variable, practitioners choose forecasting methods that are simple to implement (FILDES et al., 2008). For example, time series models which identify patterns (trend, seasonality, cyclical, and randomness) and extrapolate them into the future. These techniques have higher predictive accuracy in stable markets, are simple to develop, and require a limited amount of data (CHASE JR, 2013). However, marketing has not given much attention to time series. This is attributed, among other reasons, to the resistance of marketing scientists to data-driven approaches and to the lack of adequate data sources (DEKIMPE; HANSSENS, 2000).

The context of previous studies in marketing are, usually, companies that produce and distribute consumer packaged goods on a global scale (e.g., DIVAKAR; RATCHFORD; SHANKAR, 2005). This category of products is commonly used, since it tends to provide richer data sets (CHEN; STECKEL, 2012). Nevertheless, one may face some challenges regarding the available data and the needs of a durable good manufacturer. Scanner data are uncommon for infrequently purchased (durable) items, so they rely on aggregate sales data. Also, it is more difficult to forecast sales of durable goods, leading to lower forecast accuracy than for other types of goods (WACKER; LUMMUS, 2002). Despite these challenges, our aim is to offer a tool that can be generalized for any company, regardless of the product or category, using aggregated data.

In forecasting literature, the issue of aggregation has been studied under two research streams (DEKKER; VAN DONSELAAR; OUWEHAND, 2004). One is the kind of aggregation, what we call in this paper "aggregation criteria", that is related to how the forecaster chooses to divide the levels of data. For example, one can divide the data by products or by similarity of seasonal patterns and then structure those levels as hierarchies or groups. The other is focused on how to adjust the forecasts done in different levels of aggregation till they add up. The later stream is called reconciliation (HYNDMAN; ATHANASOPOULOS, 2018).

Another issue to consider is the structure of the aggregation. Grouped time series are those that can be aggregated based on some criteria such as product characteristics, geographic regions, customer characteristics, and so on. When these criteria can also be represented in a tree structure, it is called a "hierarchical time series", as shown in figure 1. In a marketing context, hierarchies commonly occur due to geography (where total sales are disaggregated by state, region, city, and store) and due to product classification (where brands are disaggregated into groups, sub-groups, and finally products). An example of grouped time series is when both geographic and product hierarchies are used simultaneously, such as when one wants to forecast for different products in different regions.

Figure 1 - Hierarchy example



Source: the authors (2019)

2.1 Aggregation criteria

There are several decisions to make when setting up a forecasting system with hierarchical or grouped time series, which can influence the system performance (FLIEDNER; LAWRENCE, 1995). One of the first decisions is what aggregation criteria to use. Coherent forecasts grouped by product or distribution channels can help determine companies' effort allocation. This can be done not only by product or channel, but also by partner, department, or region, giving information about which of those are more likely to be profitable.

Fliedner and Mabert (1992) studied the influence of different grouping criterion on the performance of hierarchical forecasts. They conclude that the criteria used to determine the groups for forecasting are determinant to the success of a forecasting system. However, the number or size of the groups does not have a significant impact. Series can be aggregated based on their similarities using clustering methods (see DEKKER; VAN DONSELAAR; OUWEHAND, 2004; FLIEDNER; LAWRENCE, 1995; ZOTTERI; KALCHSCHMIDT; CANIATO, 2005) or by temporal aggregation (see KOURENTZES; ROSTAMI-TABAR; BARROW, 2017). However, when the groups do not correspond to market-related criteria, they are less useful for budget plans and strategy development.

In a later study, comparing different groups based on volume but generated by cluster techniques, Fliedner and Lawrence (1995) found no evidence that the added sophistication improved forecast performance. For them, grouping items is responsible for improved forecast performance, but not the process of group formation. Divakar, Ratchford and Shankar (2005) on the other hand focused on forecasting for existing products by channels. Still, the comparison between different criteria based on marketing mix variables was not addressed, and the present paper intends to fill this gap.

2.2 Reconciliation approaches

Another decision in a hierarchical or grouped forecast system is related to the reconciliation approach. Following the notation of Hyndman and Athanasopoulos (2018), we denote the data at the most aggregate level at time t by y_t ($t = 1,\ldots,T$). More disaggregated data are denoted by $y_{j,t}$, with j corresponding to the "node" of the observation. In this way, the time series of figure 1 can be written as follows.

Bottom-level: $y_t = y_{AA,t} + y_{AB,t} + y_{AC,t} + y_{BA,t} + y_{BB,t}$ (1)
Middle level: $y_{A,t} = y_{AA,t} + y_{AB,t} + y_{AC,t}$ (2)
$y_{B,t} = y_{BA,t} + y_{BB,t}$ (3)
Top-level: $y_t = y_{A,t} + y_{B,t}$. (4)

While the data will naturally add up appropriately, following the hierarchical structure, the forecasts may not. This can cause confusion when firms use the forecasts to plan their actions. For that reason, hierarchical forecasts must be "coherent" (WICKRAMASURIYA; ATHANASOPOULOS; HYNDMAN, 2018), that is, they must add up in the same way as the historical data.

We let $\hat{\boldsymbol{y}}_h$ denote the vector of forecasts for all nodes at horizon h, stacked in the same order as $\boldsymbol{y}_t$. These forecasts can come from any appropriate model. They are created independently for each node without regard for the hierarchical or grouped structure of the data. For example, the forecasts of the aggregate may be obtained from an ARIMA model, while forecasts for the most disaggregated series might come from a Delphi process for each sales division. We call these "base" forecasts.

Because we require coherent forecasts, these base forecasts must be reconciled; that is, they are adjusted to ensure they add up appropriately. Let $\tilde{\boldsymbol{y}}_h$ denote the reconciled forecasts. These can be expressed as

$$\tilde{\boldsymbol{y}}_h = \boldsymbol{R}\hat{\boldsymbol{y}}_h. \quad (5)$$

where $\boldsymbol{R}$ is a reconciliation matrix that can be decomposed as $\boldsymbol{R} = \boldsymbol{SG}$, and $\boldsymbol{S}$ denotes a summing matrix representing the aggregation structure (groups or hierarchies) of the data. The matrix $\boldsymbol{G}$ depends on the reconciliation approach to be used.

The most common approach to obtaining coherent forecasts is known as "bottom-up" forecasting. It involves simply summing the most disaggregated forecasts to obtain forecasts for the other series of the structure. This corresponds to setting $\boldsymbol{G}$ equal to an identity matrix in the right-hand columns, and all zeros to the left. That is,

$$\begin{bmatrix} \tilde{y}_h \\ \tilde{y}_{A,h} \\ \tilde{y}_{B,h} \\ \tilde{y}_{AA,h} \\ \tilde{y}_{AB,h} \\ \tilde{y}_{AC,h} \\ \tilde{y}_{BA,h} \\ \tilde{y}_{BB,h} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{y}_h \\ \hat{y}_{A,h} \\ \hat{y}_{B,h} \\ \hat{y}_{AA,h} \\ \hat{y}_{AB,h} \\ \hat{y}_{AC,h} \\ \hat{y}_{BA,h} \\ \hat{y}_{BB,h} \end{bmatrix}.$$

While this approach has low computational costs, it leads to less accurate forecasts (HYNDMAN et al., 2011), since the bottom-level series are typically noisy. Sales at item level are more erratic, with more variation and may have insufficient data to construct reliable forecasts (DEKKER; VAN DONSELAAR; OUWEHAND, 2004).

Two other traditional approaches to obtain coherent forecasts apply only to hierarchical time series: top-down and middle-out. Top-down starts the forecast from the total and then divides it into different levels, commonly, by historical proportions. However, historical proportions might change over time, leading to less accurate forecasts. The middle-out approach forecasts some appropriate middle level of the hierarchy. It sums the forecasts to generate predictions for the higher levels and disaggregates the forecasts to obtain predictions for the lower levels.

Hyndman et al. (2011) introduced a fourth approach, the optimal reconciliation method, later refined by Wickramasuriya, Athanasopoulos and Hyndman (2018). Unlike the other methods, this approach considers the structure of the groups or hierarchies, using more information than the traditional methods, and therefore tends to be more accurate. In this approach, the matrix ***G*** is estimated by minimizing the forecast error variance of the coherent forecasts.

The weights that form the matrix ***G*** depend on the hierarchical structure and are the result of a linear regression problem. Hyndman et al. (2011) called it "optimal" because the difference between the reconciled forecasts and the incoherent base forecasts is minimized. Wickramasuriya et al. (2018) showed that the trace of the forecast error covariance matrix is minimized using optimal reconciliation, provided the base forecasts are unbiased, and so the method is now called "MinT" or minimum trace. A variation on MinT is WLS (weighted least squares) which uses only the diagonal of the covariance matrix, setting all other values to zero. In the present paper we apply both MinT and WLS approaches to obtain reconciled forecasts.

3 MATERIAL AND METHODS

The steps we followed in our empirical evaluation were: (1) selecting the grouping criteria; (2) setting the forecast horizon; and (3) selecting the time series forecast method.

All data analysis, forecasting and output was conducted using R (R CORE TEAM, 2018) with the following packages: tidyverse (WICKHAM, 2017), forecast (HYNDMAN et al., 2018; HYNDMAN; KHANDAKAR, 2008), hts (HYNDMAN; LEE; WANG, 2017), lubridate (GROLEMUND; WICKHAM, 2011), and tsibble (WANG; COOK; HYNDMAN, 2018). The code for reproducing our analysis is available at request.

3.1 Data and selection of aggregation strategy

We used a dataset provided by a major manufacturer of electrical components that has 10 factories in Brazil and is present in more than 120 countries. Our intention (like DEKKER; VAN DONSELAAR; OUWEHAND, 2004; and NENOVA; MAY, 2016) is to test the accuracy of our proposal using complex real data and not merely on simulations that may not portray the challenges and circumstances that are incorporated on real data. The dataset consists of the history of sales from one company of electrical components (plugs and light switches). Each record is a stock keeping unit (SKU) sold from the industry to the channels located in São Paulo, Brazil, comprising over seven years of sales records, from July 2010 to September 2017. No individual customer information and transactions were used. The database refers only to channels (points-of-sale) purchases and their characteristics (type of channel, size, revenue, etc.). The total number of observations (channels purchases) is 13,719.

We constructed three different structures by aggregating product categories, channel characteristics, and geographic considerations (table 1). The 339 products were categorized in three types: plugs, light switches, and others. The database comprised 220 points-of-sale, which were categorized in four types: distributor, retail, warehouse, and others. Finally, the geographic hierarchy comprised the city zones (centre, east, west, north, and south), and 54 of the 96 city districts (the missing districts had no sales records during the database time frame). At the most disaggregated level, there were 339 x 220 = 74,580 product-store combinations. However, such disaggregated data are too noisy to be useful, so we do not consider them.

Table 1 - Geographic hierarchy, product categories and channel categories

Levels	Series
Geography	
Total	1
Zones	5
Districts	54
Products	
Total	1
Product categories	3
Products	339
Channels	
Total	1
Channel categories	4
Stores	220

Source: the authors (2019)

The selection of such different hierarchies was based on both theoretical and managerial reasons. For the managers, being able to access forecasts of these levels and characteristics gives them better knowledge to plan budgets and strategies with partners. Theoretically, it can help to give insight on which marketing mix variables can be used as grouping criteria, improving accuracy of sales forecasts. Our goal was to evaluate the forecast accuracy of reconciled forecasts generated from the hierarchies. To achieve such a goal, we used different marketing variables, combining them in a grouped time series forecast.

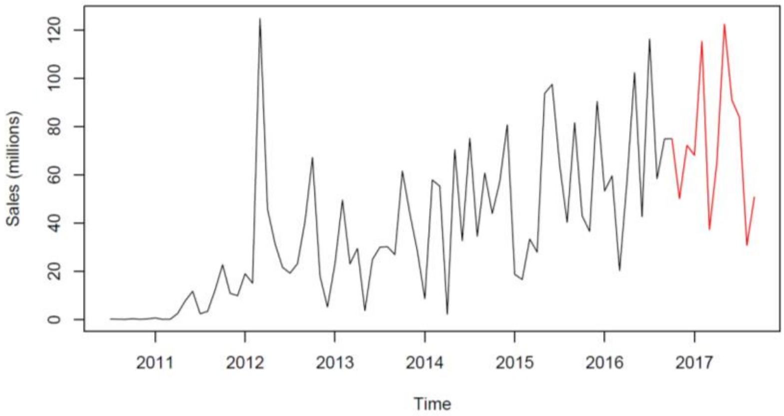
The first step was to combine different data sets provided by the company, including: (1) sales made by the company's distribution centre and direct by the factory; (2) information about the channels' characteristics; and (3) location data of the channels. Next, we filtered the data corresponding to sales of electrical components made for the city of São Paulo. The categories of products and channels were created based on the descriptions provided by the data sets.

Further aggregation of the products, channels and geographic regions was carried out so that the amount of missing data would not make the forecasting infeasible. According to Seaman (2018), handling missing data is one of the most important parts of the analysis of forecasting accuracy. While imputation methods are available, they do not allow the accuracy of the actual sales to be measured.

The literature on hierarchical forecasting suggests how to distribute the forecasts throughout different levels (HYNDMAN; ATHANASOPOULOS, 2018). Historically, the major theoretical question has been whether to forecast aggregated data and divide it among different levels, or to produce disaggregated forecasts that are then added up. These two common strategies are known as "top-down" and "bottom-up", as reviewed in section 2.2. In this paper, we apply an alternative approach proposed in forecasting research, known as "optimal reconciliation" (HYNDMAN et al., 2011), also reviewed in section 2.2.

We executed five different analyses. First, we estimated the forecasts for the aggregates based on the three grouping variables (geography, products, and channels) separately, as a hierarchical time series. Next, we combined the product and geographic hierarchies. Finally, we integrated all three hierarchies. In each case, we divided the time series into a training set of 75 months for model estimation, and a test set of 12 months for post-sample evaluation, as figure 2 shows for the total series. Thus, the maximum forecast horizon we consider is 12 months.

Figure 2 - Total sales



Source: the authors (2019)[1]

3.2 Time series method

Exponential smoothing models were used (HYNDMAN et al., 2008), and the selected model was ETS(A,N,N). The ETS(A,N,N) model is a non-stationary linear state space model with additive homoscedastic errors, no trend, and no seasonality. The model is defined via an observation equation that establishes the relationship between unobserved states and the observations,

$$y_t = l_{t-1} + \varepsilon_t \, , \tag{6}$$

and a state equation, that establishes the evolution of states over time,

$$\mathrm{l}_t = \mathrm{l}_{t-1} + \alpha\varepsilon_t \, , \tag{7}$$

Where l_t represents the anticipated sales volume and $\varepsilon_t \sim \mathrm{NID}(0, \sigma^2)$, represents the unanticipated sales volume. The latter is usually assumed to be from a Gaussian white noise process with variance σ^2. The smoother parameter α denotes the degree of change in the sales volume over time. The forecast can be expressed as a linear function of historical observations.

The algorithm implemented by the forecast package selects the best model according to Akaike's Information Criteria (AIC); optimizes the parameters using maximum likelihood estimation; forecasts using the selected model for the nominated horizon; and calculates the associated prediction intervals (HYNDMAN; KHANDAKAR, 2008).

The name "exponential smoothing" is related to the fact that weights given to observations in the forecast function decrease exponentially with time (the older the observation is, the less weight it will have). ETS models were fitted to the series at each node in each of the five grouping structures. These were then reconciled using the WLS and MinT methods as described in section 2.2.

When applied on a large scale, a natural concern is the computational costs. Because our approach separates the generation of forecasts from their reconciliation, it is easily parallelizable. Forecast for all nodes can be computed in parallel, and the reconciliation step can then be calculated using sparse matrix algebra. The computational time for generating the forecasts is much more substantial than the time required to reconcile them. We have

[1] Training data are shown in black and test data in red.

successfully reconciled forecasts from 5 million nodes in less than 30 seconds using a standard laptop computer.

To reduce the computational time required to generate individual forecasts, data for similar products or regions can be clustered. This also helps when the time series contain numbers close to zero or short time series, that would make forecasts infeasible. Our approach also uses programming languages that are open and accessible for use by any organization. It also uses information that is easily available to companies and is generalizable and easy to implement.

4 RESULTS

The hierarchical and grouped time series forecast accuracy measures are shown in tables 2 - 4. The various grouping structures make little difference to the accuracy of total aggregate sales but can make a substantial difference to some of the disaggregated forecasts. To evaluate the different reconciliation methods and grouping criteria we used the scaled error proposed by Hyndman and Koehler (2006) in order to remove the effect of the scale of the series at each node.

MASE is an evaluation measurement that can be used for comparing forecasts with different horizons, time frames or even different time series (HYNDMAN; KOEHLER, 2006). Percentage errors are also unit-free but have the disadvantage of not being useful when observations are zero or close to zero. In our study the horizon was kept fixed, but the different grouping criteria created different time series to be compared.

For seasonal time series, MASE is defined by Hyndman and Athanasopoulos (2018) as:

$$\text{MASE} = \text{mean}(|q_j|),$$

where $q_j = e_j/Q, e_j$ is a forecast error,

$$Q = \frac{1}{T-m}\sum_{t=m+1}^{T} |y_t - y_{t-m}|$$

is the scaling factor computed on the training data, and m is the number of observations per year. MASE will return a value smaller than one if the out-of-sample forecast error is smaller than the in-sample one-step forecast MAE of the seasonal naïve method and will be greater than one otherwise. The values of MASE for the different grouping criteria and methods in our study can be seen in table 2. Each level was divided into a training set with 75 months of data for model estimation and a test set with 12 months, for post-sample evaluation.

Table 2 - MASE by level

Level	MASE
Products & geography	
Total	0.9
Products	0.83
Zones	1.01
Districts	1.27
All groups	
Total	0.9
Products	0.83

Zones	1
Districts	1.27
Channels	0.81
Geography	
Total	0.9
Zones	1.02
Districts	1.31
Channels	
Total	0.9
Channels	0.82
Products	
Total	0.91
Products	0.84

Source: the authors (2019)

Table 2 also gives more detailed information about the MASE at each level. At the aggregate level, "total", forecast accuracy is higher than in each of the disaggregated levels, only when geographic considerations are used. On the hierarchical and grouped structures, the level of products and channels has smaller errors than at the most aggregate level. If the interest is on forecasting by product category, either of the grouped structures have a slightly better performance. If the aim is to forecast only the total, aggregate level, adding product category to the hierarchical structure was not enough to reduce the error. However, if the interest is to forecast by channels or geographic areas, the grouped structure with all the marketing mix variables considered gives better performance for the channels level, but requires additional information to be collected, stored and analysed.

Table 3 - MASE by products, channels and geography

Districts	GTS		HTS		
	Products & geography	All groups	Geography	Products	Channels
Total	0.9	0.9	0.9	0.91	0.9
Products					
I	1.07	1.07		1.07	
O	0.44	0.44		0.49	
T	0.84	0.84		0.85	
Channels					
DT		0.31			0.28
OT		1.82			1.69
RT		0.81			0.83
WS		1.43			1.44
Zones					
Centre	0.47	0.47	0.49		
East	2.06	2.05	2.05		
North	1.15	1.15	1.16		
South	2.79	2.77	2.81		
West	1.29	1.25	1.32		

Source: the authors (2019)

Overall, the structure with the best MASE values is the one that combines all the groups. Tables 3 and 4 give the MASE values by group at the lowest level of aggregation. Again, the structure with all the groups provides the best forecasts in most of the nodes. Of all the nodes, 65% (43 of 66) had a better performance with the structure that uses the information of all the groups. The grouped time series forecast with the combined information about product categories and geographic considerations had a better performance in 45% (28 of 62) of the nodes. The hierarchical forecast based on channel characteristics, was the best in 3 of 5 nodes (60%). The forecasts based on geography and product categories performed better in 20% (12 of 59) and 25% (1 of 4) of the cases, respectively.

Table 4 - MASE by district

Districts	GTS		HTS
	Products & geography	All groups	Geography
Água Rasa	0.47	0.48	0.48
Aricanduva	1.31	1.27	1.33
Artur Alvim	1.63	1.63	1.64
Barra Funda	1.06	1	0.95
Belém	6.6	6.59	6.58
Brasilândia	1.75	1.76	1.8
Cachoeirinha	3.34	3.34	3.37
Campo Grande	0.87	0.67	0.8
Campo Limpo	4.27	4.3	4.34
Cangaíba	1.73	1.72	1.77
Capão Redondo	4.95	4.98	4.92
Casa Verde	1.1	1.09	1.1
Cidade Ademar	3.14	2.87	3.26
Cidade Dutra	5	4.94	5
Cidade Líder	2.08	2.07	2.08
Freguesia do Ó	0.51	0.54	0.55
Grajaú			
Iguatemi	0.87	0.84	0.93
Itaim Bibi	12.86	12.81	12.85
Itaquera	0.49	0.48	0.53
Jabaquara	0.98	1.06	1.36
Jaçanã	2.26	2.27	2.3
Jaraguá	0.9	0.91	0.94
Jardim Helena	2.24	2.21	2.41
Jardim São Luís	1.5	1.5	1.52
Mandaqui	1.72	1.72	1.74
Moema	0.67	0.65	0.98
Parelheiros	1.99	2.08	2.4
Pari	36.28	35.96	36.97
Pedreira	0.49	0.46	0.54
Penha	1.82	1.79	1.88

Perus	9.76	9.76	9.76
Pinheiros	4.63	5.08	7.01
Pirituba	1.25	1.25	1.25
Raposo Tavares	4.25	4.26	4.31
República	0.55	0.55	0.56
Rio Pequeno	11.26	11.65	10.76
Sacomã	1.71	1.75	1.68
Santa Cecília	0.51	0.51	0.53
Santana	0.61	0.61	0.65
Santo Amaro	12.48	12.45	12.4
São Domingos	2.9	2.87	3.08
São Mateus	1.6	1.55	1.59
São Rafael	2.08	2.04	2.24
Sapopemba	0.67	0.64	0.76
Tatuapé	3.1	3.67	4.37
Tremembé	1.22	1.23	1.28
Vila Curuçá	0.69	0.58	0.76
Vila Leopoldina	0.38	0.36	0.29
Vila Maria	2.04	2.03	2.04
Vila Mariana	20.96	20.93	20.84
Vila Matilde	1.67	1.67	1.66
Vila Prudente	1.1	1.09	1.12
Vila Sônia	0.27	0.19	0.26

Source: the authors (2019)

For one of the districts (Grajaú), the MASE was undefined because all the historical observations on the training set where equal. That is the only circumstance under which a MASE is undefined. When it is necessary to disaggregate the forecast, our results provide evidence that it is best to add more information in a grouped structure. The results also suggest that, when forecasting sales, geographic considerations are important to improve accuracy.

5 DISCUSSION AND CONCLUSION

Even though researchers have been studying the aggregation issue for a long time, there is still no clear consensus about the criteria to determine its levels (ABHISHEK; HOSANAGAR; FADER, 2015). The aggregation criteria studied to date are either based on the similarity of the time series using clustering methods (see FLIEDNER; LAWRENCE, 1995 and ZOTTERI; KALCHSCHMIDT; CANIATO, 2005), by temporal aggregation (see KOURENTZES; ROSTAMI-TABAR; BARROW, 2017), or product, using volume of sales. However, comparing criteria based on different marketing mix variables is a key question that the theory should answer, due to its relevance for management and, more specifically, marketing practice. In this paper we have compared different structures of hierarchical and grouped forecasting, using different marketing variables.

This paper contributes to the literature in a few ways. First, we recommend the variables necessary to be collected and managed to allow predictive analysis. The results bring insights into which marketing mix variables are valuable to predict sales more accurately and subsequently plan actions. Second, our empirical application provides evidence that it is best to add more information about the marketing mix in a grouped structure, rather than in a hierarchical structure. Combining the information of product categories, channel characteristics, and geographic considerations led to more accurate forecasts than choosing only one criterion. The results also suggest that geographic considerations are most important to improve the accuracy of sales forecasts. Also, it suggests that grouping time series based on marketing variables improves accuracy more than considering the hierarchical structure.

We have demonstrated how forecast reconciliation can be used in a large marketing application involving several types of aggregation criteria. The resulting forecasts have the advantage of coherency and accuracy, thus providing retailers with substantial useful information. The channel characteristics and geographic considerations added useful information when considered in a grouped structure, improving the forecasts at the product level. The information about different product categories or channel characteristics alone were not enough to improve the accuracy of the forecasting models, but when added to other information regarding the grouping structure, it led to better sales forecast.

The forecasting strategy we propose allows firms to communicate consistent information at all hierarchical levels and departments, leading the efforts in the same direction. The system scales easily and provides consistency for retailers and manufacturers. Many retailers have thousands of stores spread geographically, while manufacturers may have thousands of business partners in different locations selling their products and requiring tailored forecasts. Our strategy provides coherent forecast with almost no human effort, by using an optimal reconciliation approach. It can be adapted for goals, horizons, update necessity and levels of aggregation.

One clear limitation of our research is the lack of information about other marketing variables, such as promotion or price strategies. This can be addressed in future research. Also, given the importance of geographic disaggregation in our analysis, future research can explore greater geographic disaggregation. Location data are now abundant due to mobile devices and applications that store location meta-data. It has been used, for example, in models of market response concerning hotels (ZHANG; KALRA, 2014), gas stations (CHAN; PADMANABHAN; SEETHARAMAN, 2007), fuel adoption (SHRIVER, 2015), drug prescriptions (STREMERSCH; LANDSMAN; VENKATARAMAN, 2013), solar panels (BOLLINGER; GILLINGHAM, 2012), organic products (SRIDHAR; BEZAWADA; TRIVEDI, 2012), and the car industry (ALBUQUERQUE; BRONNENBERG, 2012; BUCKLIN; SIDDARTH; SILVA-RISSO, 2008; NARAYANAN; NAIR, 2013).

Another research opportunity regarding geographic information is to explore the concentration of stores in a specific area. Sales of low involvement and low-cost durables such as plugs, and switches tend to be concentrated in areas of higher retail activity. This is present in our data. Sales are correlated (0.7) to the retail activity index (calculated by the Central Bank of Brazil) of each district of São Paulo. This is consistent with agglomeration theory (for a marketing application and revision see Liu, Steenkamp and Zhang, 2018). This theory states that for consumers it is more convenient when stores from the same category are concentrated on a specific geographic area. This way they can compare alternatives and get more information about products. That should be especially important for products that typically are not bought and do not have reviews shared online. Agglomeration of companies in a certain geographic area will, for that reason, have a positive impact on sales, explaining demand more than the agglomeration of consumers (LIU; STEENKAMP; ZHANG, 2018).

Marketing research efforts have focused on explaining the impact of the marketing mix on market response, and less attention has been given to forecast accuracy and which marketing instruments can help to improve it. However, accurate forecasts can help plan budgets and production levels, and influence brand image, price perception, customer satisfaction, and many other areas of marketing interest. We hope that the forecast reconciliation tool will help address this issue and prove useful in many marketing analyses.

REFERENCES

ABHISHEK, Vibhanshu; HOSANAGAR, Kartik; FADER, Peter S. Aggregation bias in sponsored search data: The curse and the cure. **Marketing Science**, [s. l.], v. 34, n. 1, p. 59–77, 2015.
ALBUQUERQUE, Paulo; BRONNENBERG, Bart J. Measuring the impact of negative demand shocks on car dealer networks. **Marketing Science**, [s. l.], v. 31, n. 1, p. 4–23, 2012.
ARMSTRONG, Jon Scott. **Principles of forecasting: a handbook for researchers and practitioners**. [s.l.] : Springer Science & Business Media, 2001. v. 30.
BOLLINGER, Bryan; GILLINGHAM, Kenneth. Peer effects in the diffusion of solar photovoltaic panels. **Marketing Science**, [s. l.], v. 31, n. 6, p. 900–912, 2012.
BUCKLIN, Randolph E.; SIDDARTH, Sivaramakrishnan; SILVA-RISSO, Jorge M. Distribution intensity and new car choice. **Journal of Marketing Research**, [s. l.], v. 45, n. 4, p. 473–486, 2008.
CHAN, Tat Y.; PADMANABHAN, V.; SEETHARAMAN, P. B. An econometric model of location and pricing in the gasoline market. **Journal of Marketing Research**, [s. l.], v. 44, n. 4, p. 622–635, 2007.
CHASE JR, Charles W. **Demand-driven forecasting: a structured approach to forecasting**. [s.l.] : John Wiley & Sons, 2013.
CHEN, Yuxin; STECKEL, Joel H. Modeling credit card share of wallet: Solving the incomplete information problem. **Journal of Marketing Research**, [s. l.], v. 49, n. 5, p. 655–669, 2012.
CHINTAGUNTA, Pradeep K.; NAIR, Harikesh S. Structural workshop paper-discrete-choice models of consumer demand in marketing. **Marketing Science**, [s. l.], v. 30, n. 6, p. 977–996, 2011.
DEKIMPE, Marnik G.; HANSSENS, Dominique M. Time-series models in marketing:: Past, present and future. **International journal of research in marketing**, [s. l.], v. 17, n. 2–3, p. 183–193, 2000.
DEKKER, Mark; VAN DONSELAAR, Karel; OUWEHAND, Pim. How to use aggregation and combined forecasting to improve seasonal demand forecasts. **International Journal of Production Economics**, [s. l.], v. 90, n. 2, p. 151–167, 2004.
DIVAKAR, Suresh; RATCHFORD, Brian T.; SHANKAR, Venkatesh. Practice Prize Article—CHAN4CAST: A Multichannel, Multiregion Sales Forecasting Model and Decision Support System for Consumer Packaged Goods. **Marketing Science**, [s. l.], v. 24, n. 3, p. 334–350, 2005.
DUNN, DM; WILLIAMS, WH; DECHAINE, TL. Aggregate versus subaggregate models in local area forecasting. **Journal of the American Statistical Association**, [s. l.], v. 71, n. 353, p. 68–71, 1976.
FILDES, Robert et al. Forecasting and operational research: a review. **Journal of the Operational Research Society**, [s. l.], v. 59, n. 9, p. 1150–1172, 2008.
FLIEDNER, Eugene B.; LAWRENCE, Barry. Forecasting system parent group formation: An empirical application of cluster analysis. **Journal of Operations Management**, [s. l.], v. 12, n. 2, p. 119–130, 1995.
FLIEDNER, Eugene B.; MABERT, Vincent A. Constrained forecasting: some implementation guidelines. **Decision Sciences**, [s. l.], v. 23, n. 5, p. 1143–1161, 1992.
FLIEDNER, Gene. Hierarchical forecasting: issues and use guidelines. **Industrial Management & Data Systems**, [s. l.], v. 101, n. 1, p. 5–12, 2001.
GROLEMUND, Garrett; WICKHAM, Hadley. Dates and Times Made Easy with lubridate. **Journal of Statistical Software**, [s. l.], v. 40, n. 3, p. 1–25, 2011.

HANSSENS, Dominique M. Order forecasts, retail sales, and the marketing mix for consumer durables. **Journal of Forecasting**, [s. l.], v. 17, n. 3–4, p. 327–346, 1998.
HYNDMAN, Rob J. et al. **Forecasting with exponential smoothing: the state space approach**. [s.l.] : Springer Science & Business Media, 2008.
HYNDMAN, Rob J. et al. Optimal combination forecasts for hierarchical time series. **Computational Statistics & Data Analysis**, [s. l.], v. 55, n. 9, p. 2579–2589, 2011.
HYNDMAN, Rob J. et al. **forecast: Forecasting functions for time series and linear models**. [s.l: s.n.]. 2018. Disponível em: <http://pkg.robjhyndman.com/forecast>
HYNDMAN, Rob J.; ATHANASOPOULOS, George. **Forecasting: principles and practice**. Melbourne, Australia: OTexts, 2018.
HYNDMAN, Rob J.; KHANDAKAR, Yeasmin. Automatic time series forecasting: the forecast package for R. **Journal of Statistical Software**, [s. l.], v. 26, n. 3, p. 1–22, 2008.
HYNDMAN, Rob J.; KOEHLER, Anne B. Another look at measures of forecast accuracy. **International journal of forecasting**, [s. l.], v. 22, n. 4, p. 679–688, 2006.
HYNDMAN, Rob J.; LEE, Alan; WANG, Earo. **hts: Hierarchical and Grouped Time Series**. [s.l: s.n.]. 2017. Disponível em: <https://CRAN.R-project.org/package=hts>
KOURENTZES, Nikolaos; ROSTAMI-TABAR, Bahman; BARROW, Devon K. Demand forecasting by temporal aggregation: Using optimal or multiple aggregation levels? **Journal of Business Research**, [s. l.], v. 78, p. 1–9, 2017.
KREMER, Mirko; SIEMSEN, Enno; THOMAS, Douglas J. The sum and its parts: Judgmental hierarchical forecasting. **Management Science**, [s. l.], v. 62, n. 9, p. 2745–2764, 2015.
LIU, Angela Xia; STEENKAMP, Jan-Benedict EM; ZHANG, Jurui. Agglomeration as a Driver of the Volume of Electronic Word of Mouth in the Restaurant Industry. **Journal of Marketing Research**, [s. l.], 2018.
NARAYANAN, Sridhar; NAIR, Harikesh S. Estimating causal installed-base effects: A bias-correction approach. **Journal of Marketing Research**, [s. l.], v. 50, n. 1, p. 70–94, 2013.
NENOVA, Zlatana D.; MAY, Jerrold H. Determining an optimal hierarchical forecasting model based on the characteristics of the data set. **Journal of Operations Management**, [s. l.], v. 44, p. 62–68, 2016.
R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria: R Foundation for Statistical Computing, 2018. Disponível em: <https://www.R-project.org/>
RUSSELL, Gary J.; KAMAKURA, Wagner A. Understanding brand competition using micro and macro scanner data. **Journal of Marketing Research**, [s. l.], p. 289–303, 1994.
SEAMAN, Brian. Considerations of a retail forecasting practitioner. **International Journal of Forecasting**, [s. l.], 2018.
SHRIVER, Scott K. Network effects in alternative fuel adoption: Empirical analysis of the market for ethanol. **Marketing Science**, [s. l.], v. 34, n. 1, p. 78–97, 2015.
SRIDHAR, Karthik; BEZAWADA, Ram; TRIVEDI, Minakshi. Investigating the drivers of consumer cross-category learning for new products using multiple data sets. **Marketing Science**, [s. l.], v. 31, n. 4, p. 668–688, 2012.
STREMERSCH, Stefan; LANDSMAN, Vardit; VENKATARAMAN, Sriram. The relationship between DTCA, drug requests, and prescriptions: Uncovering variation in specialty and space. **Marketing Science**, [s. l.], v. 32, n. 1, p. 89–110, 2013.
TELLIS, Gerard J.; FRANSES, Philip Hans. Optimal data interval for estimating advertising response. **Marketing Science**, [s. l.], v. 25, n. 3, p. 217–229, 2006.
WACKER, John G.; LUMMUS, Rhonda R. Sales forecasting for strategic resource planning. **International Journal of Operations & Production Management**, [s. l.], v. 22, n. 9, p. 1014–1031, 2002.

WANG, Earo; COOK, Di; HYNDMAN, Rob J. **tsibble: Tidy Temporal Data Frames and Tools**. [s.l: s.n.]. 2018. Disponível em: <https://CRAN.R-project.org/package=tsibble>
WEDEL, Michel; KANNAN, P. K. Marketing analytics for data-rich environments. **Journal of Marketing**, [s. l.], v. 80, n. 6, p. 97–121, 2016.
WICKHAM, Hadley. **tidyverse: Easily Install and Load the "Tidyverse"**. [s.l: s.n.]. 2017. Disponível em: <https://CRAN.R-project.org/package=tidyverse>
WICKRAMASURIYA, Shanika L.; ATHANASOPOULOS, George; HYNDMAN, Rob J. Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. **Journal of the American Statistical Association**, [s. l.], p. 1–16, 2018.
ZHANG, Wei; KALRA, Ajay. A Joint Examination of Quality Choice and Satisfaction: The Impact of Circumstantial Variables. **Journal of Marketing Research**, [s. l.], v. 51, n. 4, p. 448–462, 2014.
ZOTTERI, Giulio; KALCHSCHMIDT, Matteo; CANIATO, Federico. The impact of aggregation level on forecasting performance. **International Journal of Production Economics**, [s. l.], v. 93, p. 479–491, 2005.