Available online at www.sciencedirect.com
ScienceDirect
Int. J. Production Economics 108 (2007) 74–83
international journal of production economics
www.elsevier.com/locate/ijpe

A model for selecting the appropriate level of aggregation in forecasting processes

Giulio Zotteri[a], Matteo Kalchschmidt[b,*]

[a]Dipartimento di Sistemi di Produzione ed Economia Aziendale, Politecnico di Torino, C.so Duca degli Abruzzi 24, Torino, Italy
[b]Dipartimento di Ingegneria Gestionale, Università degli Studi di Bergamo, Via Marconi 5, 24044 Dalmine (BG), Italy

Available online 16 December 2006

Abstract

Demand forecasting is a major issue in several industrial sectors. A relevant choice for companies is the proper level of forecast aggregation. Forecasters need to properly identify what is the object of the forecasting process, in terms of time bucket (e.g., forecasts are produced on a daily level or on weekly one), set of items the demand refers to (e.g., single item or group of items), set of locations the demand refers to (e.g., single store or chain of stores). Managers can follow two basic approaches: on the one hand they can adopt a *detailed forecasting approach*, i.e., they can forecast demand for the item at the store by simply looking at the demand for the specific item/store; on the other hand they can adopt an *aggregated forecasting approach*.

In this paper, we aim at figuring out what is the balance between the strengths and weaknesses of these two options, and to identify the contingent variables that might lead managers to adopt one approach rather than the other. In this paper we study the aggregation across locations by evaluating the components of forecasting error under the assumption of stationary demand.

Finally, we suggest metrics that one can adopt to support the choice of the appropriate forecasting process, thus providing help to managers in defining the proper level of aggregation for a specific situation.

Keywords: Demand forecasting; Aggregation level; Quantitative model

1. Introduction

A forecast of final demand is one of the key inputs for all planning activities as most decisions aim to make supply meet demand. The forecasting problem is properly set once the output of such a process is defined. In particular, we shall define demand over three dimensions:

(a) one shall define the *market* he/she tries to forecast; e.g., one retailer might want to forecast demand at the single store level, while a manufacturer might be interested in the demand for the overall region or country; clearly the former forecasting problem is harder to tackle than the latter;
(b) one shall define the *product* the demand refers to; e.g., for a given retailer it might be fairly

*Corresponding author. Tel.: +39 35 205 2360; fax: +39 035 562 779.
E-mail address: matteo.kalchschmidt@unibg.it (M. Kalchschmidt).

0925-5273/$ - see front matter
doi:10.1016/j.ijpe.2006.12.030

difficult to predict the demand for a given product at the style-colour-size-packaging level, whereas forecasting the total turnover for a given product category might not be that hard (Wacker and Lummus, 2002);

(c) finally one needs to define the *time frame* of the forecasting problem, i.e., one shall define the *time bucket* and the forecasting horizon; indeed forecasting demand at the day level is much more complex than forecasting total yearly demand; also forecasting what is going to happen tomorrow is simpler than forecasting what is going to happen in 60 days.

In the remainder of the paper we will refer to these three dimensions as the *level of aggregation of the forecasting problem*. The smaller the market, the more detailed the definition of the product and the smaller the time bucket, the more the forecasting problem is detailed.

One might think that the definition of these parameters is up to the forecasting manager; on the contrary, they really depend on the decision-making process forecasting is supporting. One shall set the level of aggregation of the forecasting problem according to the level of aggregation of the decision making one. If the decision-making problem is a fairly aggregate one, the forecasting is going to be aggregate as well; whereas a detailed decision-making problem requires a detailed forecast. E.g., when one is budgeting total production costs for next year an aggregate yearly demand (for all products and all markets) is enough; on the contrary, when planning inventories, demand forecast needs to be very detailed, probably down to the single store, single Stock Keeping Unit (SKU), and probably for a single week. Also when one replenishes stores weekly the demand forecast shall be at the least weekly (if not daily) and a monthly demand rate is simply not enough to drive the replenishment process.

Thus, there is a fairly tight relationship between the level of aggregation of decision making and the level of aggregation of the forecasting problem. This leads some forecasting managers to believe that they have very little latitude over the choice of the proper level of aggregation of the *forecasting process*.

While forecasting managers have little to say on the level of aggregation of the forecasting problem, they can choose the *level of aggregation of the forecasting process.*

Often practitioners and academicians are led to believe that an aggregate output (e.g. forecasting total yearly turnover) requires a correspondingly aggregate process that looks at total yearly sales over time and vice versa. Empirical evidence shows that this is not the case as many companies choose a level of aggregation of the forecasting process that differs from the level of aggregation of the forecasting problem.

In the consumer goods sector, a couple of large manufacturers in Italy need to plan the inventory levels at their warehouses. Each warehouse serves a set of customers that are large retail chains. To create a demand forecast for the each sku carried in the warehouse they look at the demand for each single customer to better understand the effects of trade promotions as they are retailer-specific. They forecast demand for each retail chain-sku combination and then aggregate them at the warehouse-sku level. In other words in the case of these manufacturers the forecasting process is more *detailed* than the forecasting problem on the *market* dimension (see Caniato et al., 2002). A major European retailer needs to plan inventory levels at the store-sku-day level and faces the challenge of managing promotions. This retailer looks at the aggregate demand to understand the ''lift'' factors that a promotion creates and then uses such lift factors at the single store level. In this case, the forecasting process is more *aggregate* than the forecasting problem on the *market* dimension (see Zotteri et al., 2005)[1]. Finally, in the now classic Sport Obermeyer Case (Hammond and Raman, 1994), Wally Obermeyer needs to plan the production of parkas at the style-colour-size-level. However, the early demand from the Las Vegas fair is analysed at the style–colour level. Only after style–colour predictions are generated they are broken down at the size level. In the case of Obermeyer the forecasting process is more *aggregate* than the forecasting problem on the *product* dimension.

To show that the choice of the level of aggregation is a relevant issue we resort to a simple example described in Table 1. The company has two stores, A and B, and one product. We assume the company

[1]Literature refers to a forecasting process that is more detailed than the forecasting problem as *bottom-up*: initially forecasts are generated at a detailed level and then they are added up (i.e., aggregated). On the contrary, literature refers to situations where the forecasting problem is more detailed than then forecasting process as top-down: first aggregate forecasts are generated and then they are broken down at a more detailed level.

Table 1
Demand data: differences between aggregate and detailed forecasting processes

Store	Period 1	Period 2	Period 3	Period 4	Period 5
A	1	0	1	0	1
B	0	2	2	4	4
Chain	1	2	3	4	5

has decided to adopt linear regression to estimate demand and needs to forecast demand for the product both at the chain level to plan purchases and at the store level to plan distribution.

In a detailed process we first forecast demand at the store level and then aggregate forecasts at the chain level. Demand forecast for the next period would be 0.5 units for store A and 5.4 units for store B, adding up to 5.9 units for the chain (this is often called the bottom-up approach). If we adopt an aggregate process we first create a forecast for the chain (6 units) and then break it down at the store level. If we break down the forecast according to the sales rate, the forecast for store A is 1.2 and the forecast for store B is 4.8 (this is often called the top-down approach).

Thus, not only companies choose different levels of aggregation of the forecasting process, but this choice matters as, with a given algorithm, the outcome may change significantly (Fildes and Beard, 1992).

Clearly the decisions of a company depend on many factors including the degree of centralization of the organization, the availability of detailed data, algorithms adopted (Mentzer and Cox, 1984; Mentzer and Kahn, 1995), availability and flexibility of human resources (Wacker and Lummus, 2002) and features of the software used to forecast demand. While we acknowledge that these issues are relevant, in our paper we will only investigate how the *nature of demand* influences the choice of the appropriate level of aggregation.

2. Literature review

Forecasting has received significant attention from academicians but literature on aggregation is relatively sparse. Several contributions on this issue focus on the use of aggregation to estimate seasonality curves (Dalhart, 1974; Withycombe, 1989; Bunn and Vassilopoulos, 1993, 1999; Dekker et al., 2004). These works provided evidence that aggregating correlated time series can be helpful to better estimate seasonality since it can reduce casual variability. Other works focus on the selection of the proper level of data aggregation (e.g., Chan, 1993; Gonzales, 1992; Weiss, 1984). Some authors argue that the top-down approach can be helpful as it is more efficient and more accurate in times of stable demand (Theil, 1954; Grunfeld and Griliches, 1960; Lapide, 1998). Other authors reply that the bottom-up approach is needed when there are differences across time series (Orcutt et al., 1968; Zellner and Tobias, 2000; Weatherford et al., 2001).

Finally, a third group of papers (Miller et al., 1976; Barnea and Lakonishok, 1980; Fliedner, 1999) seems to take a more contingent approach and shows that the choice between the aggregate and detailed approach depends on correlation among time series.

Our contribution belongs to this third cluster and tries to highlight demand contingencies that (should) drive the decision of managers. However, this paper differs from previous contributions in two ways:

- Previous literature is based on data analysis rather than modelling; this makes the comparison of papers and findings hard as different results can be due to different data as well as differences in algorithms or aggregation processes. These analyses provided the community with empirical evidence of the relevance of the issue but we still lack a model that defines relevant variables and provides a general framework.
- Previous papers focus on aggregation across items, whereas this paper provides a model on the aggregation across stores, i.e., geographical locations. However, our model can be applied to the case of different items as well, as it can be used to investigate the aggregation over the *market* dimension.

3. The model

3.1. Assumptions

Our model investigates a single item i, so we assume that there is no relevant relationship (e.g., no correlation) among products.

Product i is sold through a chain of J stores and the demand process is stationary. Also, for the sake of simplicity we assume there is no correlation among stores and over time, note that this assumption could be easily relaxed at the expense of clarity and simplicity.

The demand x_{ijt} of item i at store j at time t follows a generic distribution $d(m_{ij},\sigma_{ij})$, where:

- m_{ij} is the expected level of demand for item i at store j in any time bucket;
- σ_{ij} is the standard deviation of demand for item i at store j and it is proportional to the mean demand $\sigma_{ij} = \mathrm{CV}m_{ij}$, where CV is the coefficient of variation of demand that measures variability.

We assume to have observed demand at all J stores over T periods of time. T is the number of periods during which we could observe a stationary demand process, thus it can be limited for several reasons:

1. the company/product can be fairly new and thus the amount of observations on past demand could be limited;
2. the company might not store the data for a long period of time and thus a reduced set of observations might be available, even if ICTs are making data storage cheaper and cheaper;
3. the demand process could experience dramatic changes that make (a maybe long) post history irrelevant. E.g., consider a product that has recently been set on promotion and thus the off promotion history is somehow irrelevant (and vice versa) or a store that has re-opened after substantial restructuring. Actually, the idea behind a very traditional forecasting technique such as moving average is that only the last T observations are stationary and thus can be used to forecast future demand.

Also we assume that the managers have no priors on the parameters of the distribution that will be estimated only by looking at past observations. We consider the forecasting managers to have no information on future demand draws (e.g., no orders are collected before time t) other than past observations.

In the end, we assume that item i does not sell evenly at all J locations and thus m_{ij} is distributed according to $f(m_j;\ h\cdot m_j)$ where h is the degree of heterogeneity of stores, i.e., the differences in the success of the product among the stores. We assume that the company does not have any prior on the differences in selling rates among stores so we assume stores to be a priori equal.[2]

In our model we assume m_{ij} to be unknown parameters we want to estimate. So we will compare the alternatives we are going to investigate in terms of ability to properly estimate m_{ij}. Indeed, in a stationary process the expected demand is the forecast that minimizes the (expected squared) error. Thus an improved estimate of the expected demand leads to a better forecast accuracy on average. Given the i.i.d. assumption it is easy to show that

$$E[(x_{ijt}-\hat{x}_{ijt})^2] = E[(x_{ijt}-m_{ij})^2] + E[(m_{ij}-\hat{x}_{it})^2] = \sigma_{ij}^2 + E[(m_{ij}-\hat{x}_{it})^2], \quad (1)$$

where $\hat{x}_{ijt}$ is the forecast of demand of item i at store j at time t. Thus the total forecasting error, in this context, is equal to the variability of the process plus the estimation error of the parameter. The first portion of the equation is an exogenous variable for the forecasting managers (though it can be to some extent endogenous for the company at large). On the contrary, forecasting managers can somehow improve the second term through better estimates of the expected demand. In the remainder of the paper we will focus on this term (error of estimate of the expected demand) as it is the only portion of the forecasting error that, given our assumptions, can actually be improved through better forecasting techniques.

3.2. Alternatives

We assume the forecasting problem to be detailed, i.e., we assume that managers need to forecast demand at the sku-store level. E.g., managers might plan inventories for each single store in the chain; thus they might need to forecast demand at the item/store level. To do so, they can follow two basic approaches:

- on the one hand, they can adopt a very *detailed forecasting approach*, i.e., they can forecast future demand for the item at a store by simply looking at the past demand at the specific store;

[2] Also one might take store size into account by separating the estimation of store traffic from the estimation of closing rate (percentage of customers buying the item out of the customers that entered the store); when estimating closing rates even very different stores might have a very similar prior on selling rates of a given product.

- on the other hand, they can adopt a more aggregated forecasting approach. One might assume that the demand rate for the item is constant across all locations (maybe adjusting for the size of the stores in such a way that one only needs to assume that the customers' tastes are constant across locations) and thus look at the overall demand for the chain to get a more reliable estimate.

As literature suggests, these two approaches have contrasting pros and cons. The first approach can very well capture the uniqueness and specificity of the demand at a given store; the flipside however is the inability to enjoy large samples to draw statistically significant conclusions (sample size is T). The model faces a very small specification error (i.e., will be very adequate conceptually) but faces a substantial sampling error as the number of parameters is very large (J). The second approach, on the other hand, can be considered to be fairly crude, since often different stores have different customers that tend to like different products and create different demand patterns. However, the second approach has the advantage of enjoying a larger sample (sample size is JT). So conclusions might be conceptually crude as the specification error will be substantial, but the sampling error will be limited.

Under our assumptions, managers need to set inventory levels for each single store and thus need to forecast demand at this fairly detailed level. Hence, we compare the ability of the aggregated and detailed forecasting process to create an accurate forecast at this level of detail. Also notice that under our assumptions both alternatives provide the same aggregate forecast, thus in this specific context comparing the two approaches in terms of their ability to forecast aggregate demand does not make sense.

Given the assumptions we have made, the average of the past T observations is the best estimator available (unbiased minimum variance estimator—UMVE) of the average demand (i.e., the estimator that minimizes the right-hand size in Eq. (1)). Thus, we deploy this estimator both in the aggregate and in the detailed processes to check which process performs best.

If the detailed forecasting process is used then the estimate of the average demand for item i at store j is $\hat{m}_{ij}$

$$\hat{m}_{ij} = \frac{\sum_{t=1}^{T} x_{ijt}}{T} \sim g\left(m_{ij}; \frac{\sigma_{ij}}{\sqrt{T}}\right) = g\left(m_{ij}; \frac{\mathrm{CV} \cdot m_{ij}}{\sqrt{T\cdot}}\right).$$

Thus, the standard error of estimate of the detailed process (seed) is

$$\mathrm{seed}_{m_{ij}} = \frac{\mathrm{CV}\, m_{ij}}{\sqrt{T}}.$$

If the aggregate process is followed then the demand rate estimate is $\tilde{m}_{ij}$

$$\tilde{m}_{ij} = \frac{\sum_{j=1}^{J}\sum_{t=1}^{T} x_{ijt}}{TJ} \sim g\left(m_i; \frac{\sigma_{ij}}{\sqrt{J}\sqrt{T}}\right) = g\left(m_i; \frac{\mathrm{CV}\, m_i \sqrt{1+h^2}}{\sqrt{J}\sqrt{T}}\right).$$

In this case it can be shown that standard error of estimate of the aggregate process (seea) is (see Appendix A for all relevant proofs)

$$\mathrm{seea}_{m_{ij}} = \sqrt{\frac{\mathrm{CV}^2(1+h^2)m_i^2}{JT} + (m_{ij} - m_i)^2}.$$

The above formulas[3] highlight the basic trade off we are facing: in the detailed approach we have only the sampling error that might be relatively large, as data are limited. On the contrary, the aggregated approach will enjoy a relatively smaller sampling error (on average J times smaller because we sum the demand from all the J stores); however, in the aggregated approach we also face a specification error ($(m_{ij}–m_i)^2$), i.e., the model we are building is an oversimplification of reality and assumes that the demand is constant across all stores. Indeed, this error does not vanish as we increase the number of observations T.

Clearly these are very detailed errors at the item/store level and one approach might work better for some stores while the second might work better for others. Actually when choosing the aggregate rather than the detailed approach one should not look at whether one method performs better at a specific store but rather at how accurate the two alternatives are on the average.

We suggest that one might be willing to minimize the sum of all errors across all stores. Indeed the sum of all errors is likely to correlate with costs due to as safety stock and stock outs. In particular, we suggest to measure the sum of all quadratic errors that is consistent with the adoption of standard

[3]Note that this is the UMVE of the average (over the store locations) expected demand m_i; so other estimators would increase the variance $(\mathrm{CV}^2(1+h^2)m_i^2)/JT$ of the estimate or the average (over all locations J) of the difference between the expected demand and the expected value of the estimator .

error of estimate at the single store level (see Syntetos et al., 2005, for a similar approach in a different forecasting problem).

Thus, one might compare the total error for the detailed approach (ted)

$$\text{ted}_i = \sum_{j=1}^{J} \text{seed}^2_{m_{ij}} = \frac{\text{CV}^2}{T} J(1+h^2)m_i^2$$

with the error of the aggregate one (tea) (see Appendix A for proofs)

$$\text{tea}_i = \sum_{j=1}^{J} \text{seea}^2_{m_{ij}} = \frac{\text{CV}^2(1+h^2)}{T} m_i^2 + J(h\, m_i)^2.$$

Finally, the approach with the smaller total (expected) error is selected. Thus, the detailed approach is going to be selected if

$$\frac{\text{CV}^2}{T} J(1+h^2) < \frac{\text{CV}^2(1+h^2)}{T} + J\, h^2. \qquad (2)$$

4. Analysis of results

The above formulas can be used to investigate the performance of the two alternative approaches for various combinations of the contingent variables CV, h, J, T. Also, Eq. (2) enables to identify threshold values for the above variables.

Eq. (2) suggests that the detailed approach shall be chosen in cases of low demand variability, since Eq. (2) is verified when

$$\text{CV} < \sqrt{\frac{J}{J-1} \frac{h^2}{h^2+1} T}.$$

Indeed, variability makes the estimation of expected demand for each store hard; a high degree of variability forces the company to aggregate data at the chain level to gain a large enough sample.

In the case of the chain size J, Eq. (2) suggests to adopt the detailed solution if

$$\text{if } \frac{\text{CV}^2}{T} \frac{1+h^2}{h^2} \leqslant 1 \quad \forall J,$$

$$\text{if } 1+h^2 \frac{\text{CV}^2}{T} \frac{1+h^2}{h^2} > 1, \quad J < 1 + \frac{1}{\frac{\text{CV}^2}{T}\frac{1+h^2}{h^2} - 1}.$$

In other words, we suggest to use the detailed process where the chains are relatively small as the advantage one gains by aggregating at the chain level is limited.

In the case of heterogeneity, the detailed model shall be used if

$$\text{if } \frac{\text{CV}^2}{T} \frac{J-1}{J} < 1 \text{ and } h > \sqrt{\frac{1}{1 - \frac{\text{CV}^2}{T}\frac{J-1}{J}} - 1}.$$

In other words, the solution suggests to adopt the aggregate model only in case of homogeneous chains, as in other cases the error one makes by providing the same estimate for all stores is too large. Also we shall notice that the detailed alternative is not even an option in case of variable demand, small samples and large chains.

Eq. (2) also suggests to select the detailed model in case the product has been around (and/or demand has been stable) for a relatively long period of time

$$T > \text{CV}^2 \frac{J-1}{J} \frac{h^2+1}{h^2}.$$

This adds a product life cycle flavour to the model. Indeed, for a given product we suggest to use an aggregate demand model early on in the product lifecycle (or just after a sudden change in demand such as the start of a promotion) and then use a detailed process once enough evidence to support store-specific forecasts is gathered. This result is consistent with results provided by previous works (e.g., Roberts, 1998). Also this finding is very consistent with the practise of a European grocery retailer that early on in the promotion looks at the general demand trend but, as more sales data are collected, looks at the selling rate of each single store independently. Indeed, early in the product (or promotion) life cycle, what matters the most is to gain a large enough sample to reduce the sampling error. Thus, the aggregate process can work better than the detailed one. Later on, the sampling error is reduced: while this leads the error of the detailed process to zero, the aggregate process is still left with the specification error that does not disappear with larger samples, as shown in Fig. 1.

Finally, we shall notice that in all our derivations of the thresholds on the four contingent variables we introduced three functions:

- $(\text{CV}^2/T) = v$ measures the extent to which the sample size enables to provide accurate estimates, given the variability of demand; in other words this function captures the sampling error;
- $(J{-}1/J) = J_1$, $0.5 \leqslant J_1 < 1$ $(J \geqslant 2)$ captures the effect of the chain size, and in particular this

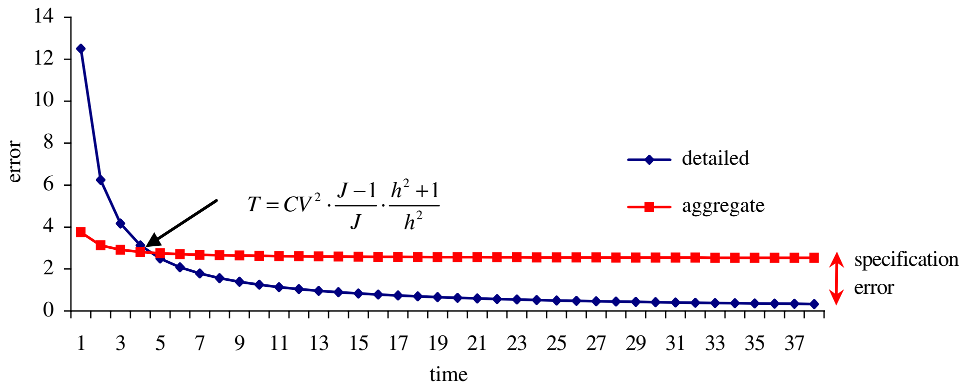

Fig. 1. Errors as a function of the number of periods T observed (case of $J = 10$, $\mathrm{CV} = 1$, $h = 0.5$).

measures the advantage that the aggregate process has over the detailed one as it can increase the sample size and thus reduce the sampling error;

- $h^2/(h^2+1) = h_1$, $0 \leqslant h_1 < 1$ is the factor that captures the effect of heterogeneity on the specification error of the aggregate approach.

Indeed, all above equations can be re-written as a function of v, J_1 and h_1.

5. Estimation of parameters

The model presented in this paper provides insights into the variables that underlie the choice of the level of aggregation of the forecasting process. However, when one wants to move from the interpretative and modelling stage to the prescriptive one, he/she needs to estimate the parameters of the model.

While the estimation of parameter J is self explanatory, and for T we refer to the third section of this paper, other variables deserve further discussion. In particular, the estimate of h and CV is somehow tricky. Clearly when one wants to estimate the degree of difference across the stores he/she needs to look at demand at the store/item level, i.e., estimate different values of m_{ij} with a detailed process. However, in case different values $\hat{m}_{ij}$ are observed, one shall wonder whether such differences are due to actual differences in m_{ij} across stores (and thus shall be interpreted as heterogeneity) or might be due to the variability of demand. In the latter case, differences in m_{ij} across stores might simply be due to sampling errors as the estimates might be different while the parameters are equal (or have little differences).

A simple example will make the issue more clear. Let us assume that all the demands at the J stores are equal but the demand is rather variable and the amount of information available is rather limited (i.e., large CV and small T). In this case $\hat{m}_{ij}$ are likely to be fairly different and thus one could be led to believe that the demand is heterogeneous. On the contrary, we shall tell the effect of variability from the effect of heterogeneity by looking at the variance across $\hat{m}_{ij}$, to estimate whether the parameters (rather than their estimates) are actually heterogeneous.

In other words, one might be tempted to estimate h as

$$\hat{h}^2 = \frac{\sum_{j=1}^{J}(\hat{m}_{ij} - \hat{m}_i)^2}{J\,\hat{m}_i^2}.$$

However, in Appendix B we show that

$$E\left(\sum_{j=1}^{J}(\hat{m}_{ij} - \hat{m}_i)^2\right) = m_i^2\left(\frac{(1+h^2)(J-1)\mathrm{CV}^2}{T} + J\,h^2\right) \tag{3}$$

and thus what seems to be an intuitive estimate of h is actually biased because of variability.

Thus, Eq. (3) can be used to estimate h, once an estimate of CV is derived.

It is fairly easy to estimate CV by looking at the demand for a given product at a given location over time as follows:

$$\hat{\mathrm{CV}} = \frac{1}{J}\sum_{1}^{J}\sqrt{\frac{\sum_{t=1}^{T}(x_{ijt} - \hat{m}_{ij})^2}{(T-1)\hat{m}_i^2}}.$$

Thus, once the estimation of the variability is derived from time series we can tell the contribution of variability from the contribution of heterogeneity in the variance of the estimates of mean selling rates at stores (see Eq. (3)).

6. Conclusions

This work evaluates the trade-off between specification and sampling error when choosing the level of aggregation of the forecasting process. This paper identifies the contingent variables that might lead managers to adopt a detailed rather than an aggregate approach. This is achieved by evaluating the components of forecasting error under the assumption of non-correlated (over time and across products or markets), stationary demand and by considering aggregation only on locations.

The contribution may be significant in real applications, since it may help companies to understand what the best approach to forecast future demand is. Thus we also suggest metrics that can be adopted to support the choice of the appropriate forecasting process. However the current model is limited from several perspectives and thus future development is needed to properly investigate this issue.

A first set of directions for further developments regards the simplistic assumptions of current model, so additional efforts should be devoted to building the theory.

This work assumes demand to be stationary, however in many different contexts this is not the case (e.g., spare parts, fashion goods, and so on). Thus it would be useful to model the case of non-stationary demand. A first step could be to assume linear trend and use linear regression to generate future forecasts. Also such a change might lead to differences in aggregate forecasts between the detailed and the aggregate forecasting process (see Zotteri et al., 2005); thus one might have to compare the performance of various alternatives both at the chain and store level.

A second interesting issue is to explore the case of demand correlated over time and/or products and/or across locations; nevertheless we argue that we would probably get similar results. Indeed, correlation across stores simply reduces the benefits of aggregation as the behaviour of stores over time is similar and thus the sampling error is not reduced by larger samples (i.e., aggregation).

Moreover, attention has been paid here on aggregation on the market dimension (e.g., forecast at store or chain level). It would be important to add product dimension, so more options are available and we can aggregate over products and/or stores.

Also the current model assumes no priors on parameters and namely heterogeneity and thus the expected level of demand at different stores. Actually this seems to be a fairly strong assumption as even before a product is launched the managers might have some ideas on both the absolute level of demand (m_i) and its distribution across stores. Indeed, store size or traffic could be a significant (and only partially unknown) drivers of demand at the store level (m_{ij}).

Finally, it would be interesting to model clustering of stores (or items) as an intermediate alternative between a very detailed and a very aggregate model (see Zotteri et al., 2005; Caniato et al., 2005). This would help companies to identify the level of aggregation that better solves the trade-off between specification and sampling error. This would probably entail the modelling of both intra-cluster and inter-cluster heterogeneity.

Also the current research requires some additional efforts on the empirical side to test the current (and eventually future) theory.

The robustness of the current model shall be tested by checking the sensitivity to both the fairly strong assumption made and errors in estimate of parameters. This can be achieved by performing sensitivity analyses through Monte Carlo simulation. Also the actual ability to select the appropriate aggregation of the forecasting process should be tested with real data that are currently being collected.

Acknowledgments

Authors have contributed jointly to the present work; however Giulio Zotteri has edited the following paragraphs: "The model", "Analysis of results", "Estimation of parameters" and "Conclusions". Matteo Kalchschmidt has written the remaining paragraphs "Introduction" and "Literature review". Also we wish to thank three anonymous referees for their helpful guidance and insightful comments and suggestions.

Appendix A. Proofs

Given assumptions we know that

$$\frac{\sum_{j=1}^{J} m_{ij}}{J} = m_i,$$

$$\frac{\sum_{j=1}^{J} (m_{ij} - m_i)^2}{J} = h^2 m_i^2,$$

hence we can derive that

$$\frac{\sum_{j=1}^{J} (m_{ij} - m_i)^2}{J} = \frac{\sum_{j=1}^{J} (m_{ij})^2 - \sum_{j=1}^{J} (m_i)^2}{J} = \frac{\sum_{j=1}^{J} (m_{ij})^2 - J(m_i)^2}{J} = h^2 m_i^2$$

and thus

$$\sum_{j=1}^{J} m_{ij}^2 = J\, m_i^2 (1 + h^2).$$

Also we can derive the properties of the estimator $\tilde{m}_{ij}$

$$E(\tilde{m}_{ij}) = \frac{\sum_{j=1}^{J} \sum_{t=1}^{T} E(x_{ijt})}{T\,J} = m_i,$$

$$\begin{aligned}\sigma^2_{\tilde{m}_{ij}} &= E[(\tilde{m}_{ij} - m_i)^2] \\ &= \frac{\sum_{j=1}^{J} \sum_{t=1}^{T} \sigma^2_{x_{ijt}}}{T^2 J^2} = \frac{\sum_{j=1}^{J} \sum_{t=1}^{T} (\mathrm{CV}\, m_{ij})^2}{T^2 J^2} \\ &= \frac{\mathrm{CV}^2 \sum_{j=1}^{J} \sum_{t=1}^{T} m_{ij}^2}{T^2 J^2} = \frac{\mathrm{CV}^2\, T \sum_{j=1}^{J} m_{ij}^2}{T^2 J^2} \\ &= \frac{\mathrm{CV}^2\, J\, m_i^2 (1 + h^2)}{T\, J^2} = \frac{\mathrm{CV}^2\, m_i^2 (1 + h^2)}{T\, J}.\end{aligned}$$

In the derivation process we first used the assumption on independence of demands across stores and over time and then the distributive assumption on the variance of m_{ij}.

$$\begin{aligned}\mathrm{seea}_{m_{ij}} &= E[(\tilde{m}_{ij} - m_{ij})^2] \\ &= E[((\tilde{m}_{ij} - m_i) + (m_i - m_{ij}))^2] \\ &= E[(\tilde{m}_{ij} - m_i)^2] + E[(m_i - m_{ij})^2] \\ &\quad + 2E[(\tilde{m}_{ij} - m_i)(m_i - m_{ij})] \\ &= \frac{\mathrm{CV}^2\, m_i^2 (1 + h^2)}{T\, J} + (m_i - m_{ij})^2.\end{aligned}$$

We can compute the total error one makes by using the detailed approach as

$$\begin{aligned}\mathrm{ted}_i &= \sum_{j=1}^{J} \mathrm{seed}_{ij}^2 \\ &= \sum_{j=1}^{J} \frac{\mathrm{CV}^2\, m_{ij}^2}{T} = \frac{\mathrm{CV}^2}{T} \sum_{j=1}^{J} m_{ij}^2 \\ &= \frac{\mathrm{CV}^2}{T} J (1 + h^2) m_i^2.\end{aligned}$$

And can compute the total error one makes by using the aggregate approach as

$$\begin{aligned}\mathrm{tea}_i &= \sum_{j=1}^{J} \mathrm{seea}_{mj}^2 \\ &= \sum_{j=1}^{J} \left[\frac{\mathrm{CV}^2\, m_i^2 (1 + h^2)}{T\, J} + (m_i - m_{ij})^2 \right] \\ &= \sum_{j=1}^{J} \left[\frac{\mathrm{CV}^2\, m_i^2 (1 + h^2)}{T\, J} \right] + \sum_{j=1}^{J} (m_i - m_{ij})^2 \\ &= \frac{\mathrm{CV}^2 (1 + h^2)}{T} m_i^2 + J (h\, m_i)^2,\end{aligned}$$

$$\begin{aligned}\mathrm{ted}_i &= \sum_{j=1}^{J} \mathrm{seed}_{m_{ij}}^2 = \sum_{j=1}^{J} \frac{\mathrm{CV}^2\, m_{ij}^2}{T} \\ &= \frac{\mathrm{CV}^2}{T} \sum_{j=1}^{J} m_{ij}^2 = \frac{\mathrm{CV}^2}{T} m_i^2 (h^2 + 1).\end{aligned}$$

Appendix B. Estimation proofs

$$\begin{aligned}&E\left(\sum_{j=1}^{J} \frac{(\hat{m}_{ij} - \hat{m}_i)^2}{J} \right) \\ &= \frac{1}{J} \sum_{j=1}^{J} E\{[(\hat{m}_{ij} - m_{ij}) + (m_{ij} - \hat{m}_i)]^2\} \\ &= \frac{1}{J} \sum_{j=1}^{J} \{E(\hat{m}_{ij} - m_{ij})^2 + E(m_{ij} - \hat{m}_i)^2 \\ &\quad + 2E[(\hat{m}_{ij} - m_{ij})(m_{ij} - \hat{m}_i)]\},\end{aligned}$$

where by definition

$$E(\hat{m}_{ij} - m_{ij})^2 = \frac{\mathrm{CV}^2 m_{ij}^2}{T}$$

and

$$\begin{aligned}E(m_{ij} - \hat{m}_i)^2 &= E(m_{ij} - m_i + m_i - \hat{m}_i)^2 \\ &= (m_{ij} - m_i)^2 + \frac{\mathrm{CV}^2 (1 + h^2) m_i^2}{J\, T} + 0,\end{aligned}$$

$$
\begin{aligned}
&E[(\hat{m}_{ij} - m_{ij})(m_{ij} - \hat{m}_i)] \\
&= E\left[\left(\frac{\sum_{t=1}^{T} x_{ijt}}{T} - m_{ij}\right)\left(m_{ij} - \frac{\sum_{j=1}^{J}\sum_{t=1}^{T} x_{ijt}}{JT}\right)\right] \\
&= E\left[\left(\frac{\sum_{t=1}^{T} x_{ijt}}{T} - m_{ij}\right)\left(m_{ij} - \frac{\sum_{k\neq j}\sum_{t=1}^{T} x_{ijt}}{J\,T} - \frac{\sum_{t=1}^{T} x_{ijt}}{J\,T}\right)\right] \\
&= 0 + E\left[\left(\frac{\sum_{t=1}^{T} x_{ijt}}{T} - m_{ij}\right)\left(-\frac{\sum_{t=1}^{T} x_{ijt}}{J\,T}\right)\right] \\
&= E\left[m_{ij}\frac{\sum_{t=1}^{T} x_{ijt}}{J\,T}\right] - E\left[\frac{\sum_{t=1}^{T} x_{ijt}}{T}\frac{\sum_{t=1}^{T} x_{ijt}}{J\,T}\right] \\
&= \frac{m_{ij}^2}{J} - E\left[\frac{\sum_{t=1}^{T} x_{ijt}^2}{J\,T^2} + \frac{\sum_{t=1}^{T} x_{ijt}\sum_{u\neq t} x_{iju}}{J\,T^2}\right] \\
&= \frac{E(x_{ijt})^2}{J} - \frac{T\,E(x_{ijt}^2)}{J\,T^2} - \frac{T(T-1)E(x_{ijt})^2}{J\,T^2} \\
&= \frac{E(x_{ijt})^2}{J\,T} - \frac{E(x_{ijt}^2)}{J\,T} = -\frac{\sigma_{x_{ijt}}^2}{J\,T} = -\frac{\mathrm{CV}^2 m_{ij}^2}{J\,T}.
\end{aligned}
$$

Thus

$$
\begin{aligned}
&\sum_{j=1}^{J}\left[\frac{\mathrm{CV}^2 m_{ij}^2}{T} + (m_{ij} - m_i)^2 \right.\\
&\quad\left. + \frac{\mathrm{CV}^2(1+h^2)m_i^2}{J\,T} - 2\frac{\mathrm{CV}^2 m_{ij}^2}{J\,T}\right] \\
&= \frac{J\,\mathrm{CV}^2(1+h^2)m_i^2}{T} + J\,h^2 m_i^2 \\
&\quad + \frac{\mathrm{CV}^2(1+h^2)m_i^2}{T} - 2\frac{\mathrm{CV}^2(1+h^2)m_i^2}{T} \\
&= \frac{(J-1)\mathrm{CV}^2(1+h^2)m_i^2}{T} + J\,h^2 m_i^2.
\end{aligned}
$$

References

Barnea, A., Lakonishok, J., 1980. An analysis of the usefulness of disaggregated accounting data for forecasts of corporate performance. Decision Sciences 11, 17–26.

Bunn, D.W., Vassilopoulos, A.I., 1993. Using group seasonal indices in multi-item short-term forecasting. International Journal of Forecasting 9, 517–526.

Bunn, D.W., Vassilopoulos, A.I., 1999. Comparison of seasonal estimation methods in multi-item short-term forecasting. International Journal of Forecasting 15, 431–443.

Caniato, F., Kalchschmidt, M., Ronchi, S., Verganti, R., Zotteri, G., 2002. Forecasting demand fluctuations due to promotional activities: a case in the fresh food industry. In: Proceedings of POMS Conference, San Francisco.

Caniato, F., Kalchschmidt, M., Ronchi, S., Verganti, R., Zotteri, G., 2005. Clustering customers to forecast demand. Production Planning and Control 16, 32–43.

Chan, W., 1993. Disaggregation of annual time-series with common seasonal patterns. Journal of Econometrics 55, 173–200.

Dalhart, G., 1974. Class seasonality—a new approach. In: Proceedings of 1974 Conference of American Production and Inventory Control Society, Reprinted in Forecasting, 2nd ed. APICS, Washington, DC, pp. 11–16.

Dekker, M., Van Donselaar, K., Ouwehand, P., 2004. How to use aggregation and combed forecasting to improve seasonal demand forecast. International Journal of Production Economics 90, 151–167.

Fildes, R., Beard, C., 1992. Forecasting systems for production and inventory control. International Journal of Operations and Production Management 12 (5), 4–27.

Fliedner, G., 1999. An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation. Computers & Operations Research 26, 1133–1149.

Gonzales, P., 1992. Temporal aggregation and systemic sampling in structural time-series models. Journal of Forecasting 11, 271–281.

Grunfeld, Y., Griliches, Z., 1960. Is aggregation necessarily bad? Review of Economics and Statistics 42, 1–13.

Hammond, J., Raman, A., 1994. Sport Obermeyer Ltd., Case number 9-695-022, HBS Publishing, Boston, MA.

Lapide, L., 1998. New developments in business forecasting. The Journal of Business Forecasting Summer, 28–29.

Mentzer, J.T., Cox Jr., J., 1984. Familiarity, application, and performance of sales forecasting techniques. Journal of Forecasting 3, 27–36.

Mentzer, J.T., Kahn, K.B., 1995. Forecasting technique familiarity, satisfaction, usage, and application. Journal of Forecasting 14, 465–476.

Miller, J.G., Berry, W., Lai, C-Y.F., 1976. A comparison of alternative forecasting strategies for multi-stage production inventory systems. Decision Sciences 7, 714–724.

Orcutt, G., Watts, H.W., Edwards, J.B., 1968. Data aggregation and information loss. American Economic Review 58, 773–787.

Roberts, J.H., 1998. Marketing approaches to forecasting problems. Journal of Forecasting 17, 169–174.

Syntetos, A.A., Boylan, J.E., Croston, J.D., 2005. On the categorisation of demand patterns. Journal of the Operational Research Society 56 (5), 495–503.

Theil, H., 1954. Linear Aggregation of Economic Relations. North-Holland, Amsterdam.

Wacker, J.G., Lummus, R., 2002. Sales forecasting for strategic resource planning: practical implications and research directions. International Journal of Production and Operations Management 22 (9), 1014–1031.

Weiss, A.A., 1984. Systematic sampling and temporal aggregation in time-series models. Journal of Econometrics 12, 547–552.

Weatherford, L.R., Kimes, S.E., Scott, D.A., 2001. Forecasting for hotel revenue management—testing aggregation against disaggregation. Cornell Hotel and Restaurant Administration Quarterly August, 53–64.

Withycombe, R., 1989. Forecasting with combined seasonal indices. International Journal of Forecasting 5, 547–552.

Zellner, A., Tobias, J., 2000. A note on aggregation, disaggregation and forecasting performance. Journal of Forecasting 19, 457–469.

Zotteri, G., Kalchschmidt, M., Caniato, F., 2005. The impact of aggregation level on forecasting performance. International Journal of Production Economics 93–94, 479–491.