Forecasting item-level demands : an analytical evaluation of top-down versus bottom-up forecasting in a production-planning framework

Viswanathan, S.; Widiarta, Handik; Piplani, Rajesh

2008

Widiarta, H., Viswanathan, S., & Piplani, R. (2008). Forecasting item-level demands : an analytical evaluation of top-down versus bottom-up forecasting in a production-planning framework. IMA journal of management mathematics, 19(2), 207-218.

<https://hdl.handle.net/10356/100772>

<https://doi.org/10.1093/imaman/dpm039>

© 2008 The Author(s).

Downloaded on 26 Apr 2023 14:04:46 SGT

Forecasting item-level demands: an analytical evaluation of top–down versus bottom–up forecasting in a production-planning framework

HANDIK WIDIARTA AND S. VISWANATHAN[†]

*Nanyang Business School, Nanyang Technological University, Singapore 639798,
Republic of Singapore*

AND

RAJESH PIPLANI

*School of Mechanical and Aerospace Engineering, Nanyang Technological University,
Singapore 639798, Republic of Singapore*

[Received on 12 January 2007; accepted on 19 November 2007]

We compare the performance of top–down (TD) and bottom–up (BU) strategies for forecasting the demand of an item that belongs to a product family. The TD strategy forecasts the sum of the item demands and distributes it to the individual item based upon the historical demand proportion of each item in the family. The BU strategy forecasts each item demand individually using the historical demand data for the particular item. All the item demands, which may be correlated with each other, are assumed to follow a first-order univariate moving average process. As is common in a production-planning environment, the forecasting under both strategies is carried out using the exponential smoothing technique. We show that the performance of the two forecasting strategies is nearly identical, regardless of the coefficient of correlation between the item demands, the items' proportion in the family and the coefficient of the serial correlation term of the demand process. We further investigate the relative performance of the two strategies when a fixed (rather than the optimal) smoothing constant is used for forecasting the demand under both strategies.

Keywords: forecasting; stationary time series; moving average process; top–down forecasting; bottom–up forecasting; exponential smoothing.

1. Introduction

Good forecasts are a prerequisite for efficient and effective management of inventories in the manufacturing system and the supply chain. In a production-planning environment involving inventory management of a large number of items, forecasting is typically automated using the exponential smoothing technique and the use of hierarchical forecasting. A question that has engaged researchers as well as practitioners over the last few decades is the effectiveness of hierarchical forecasting.

In this paper, we analytically compare the performance of top–down (TD) and bottom–up (BU) strategies for forecasting the demand of an item that belongs to a product family. The demand for the items within the family may be correlated with each other and is assumed to follow a first-order moving average [MA(1)] process. The BU strategy forecasts each item individually using the historical demand data for the particular item. The TD strategy forecasts the aggregate family demand (or the sum of the item demands) and allocates the forecast to the individual item based upon the historical demand

[†]Email: vish@pmail.ntu.edu.sg

proportion of each item in the family. The rationale for the TD strategy is based on the intuition that the overall variability of demand is reduced when it is aggregated across several items, and therefore, it is easier to forecast the aggregate demand and then allocate it to the individual items (Fogarty *et al.*, 1991; DeLurgio, 1998; Kahn, 1998; Ballou, 1999).

The research literature on TD versus BU forecasting strategies can generally be divided into two streams based on the type of the forecasting technique utilized in both strategies. The first stream of research adopts an explanatory technique such as univariate and multivariate autoregressive integrated moving average (ARIMA). Theoretically, when one can perfectly identify the statistical properties of the data-generating process and the available data are free from measurement errors, the ARIMA produces unbiased forecasts (Wei, 1993, p. 86). This technique is relatively popular in the economics domain (Grunfeld & Griliches, 1960; Orcutt *et al.*, 1968; Dunn *et al.*, 1971; Shlifer & Wolff, 1979) and it was generally concluded that the BU strategy is superior to the TD strategy.

Most of the papers in the second stream of research study the issue of TD versus BU forecasting strategies in the context of a production-planning framework. Typically, in a production-planning framework, the forecasting of demand is carried out for a large number of items simultaneously and relatively simpler techniques such as exponential smoothing is used to perform the forecasting. The use of the exponential smoothing technique in a production-planning framework is justified by the fact that practitioners may not want to spend too much time and effort to examine and define the characteristics of the data-generating process (for a large number of items) prior to determining the forecasting model, as is required in the ARIMA. The literature in this stream is generally inconclusive as to whether the TD or BU strategy is superior. Dangerfield & Morris (1992), e.g. found that the BU forecasting outperformed the TD forecasting in nearly three out of four cases and that the relative superiority of the BU forecasting was more pronounced as the correlation between the two items increased and/or one item increasingly dominated the family class. These findings were also supported by Gordon *et al.* (1997), Weatherford *et al.* (2001) and Diebold (1998, p. 188). However, Gross & Sohl (1990) numerically found that the TD strategy (with a proper disaggregation method) provided better estimates than the BU forecasting in two out of three product lines examined. Other papers in this stream include Barnea & Lakonishok (1980), Miller *et al.* (1976), Zellner & Tobias (2000) and Zotteri & Kalchschmidt (2007).

The focus of the current study is on the second stream of research. Apart from the fact that the literature in a production-planning framework is inconclusive as to whether TD or BU forecasting strategies is superior, little effort has been devoted to analytically investigating this issue.

In a companion paper (Widiarta *et al.*, 2007), we analytically study the relative performance of the two strategies (TD and BU) when the item demands follow a first-order univariate autoregressive [AR(1)] process. We show in that paper that for a certain range of parameters of the AR(1) process, the BU strategy might consistently outperform the TD strategy. For the rest of the parameter range, there is no significant difference between the two forecasting strategies. In the current paper, we analytically examine the relative performance of the two strategies when the item demands follow a first-order moving average [MA(1)] process with identical coefficients of the serial correlation term. We further perform a simulation study to investigate the issue when the coefficient of the serial correlation term of the item demand processes is not identical.

We show that the performance of the two forecasting strategies is identical, regardless of the coefficient of correlation between the item demands and the items' proportion in the family when the coefficient of the serial correlation term of the item demands is identical. We further investigate the relative performance of the two strategies when a fixed smoothing constant (i.e. not necessarily the optimal) is used for forecasting the demand under both strategies. We also investigate through a simulation study the relative performance of the two strategies when the coefficient of the serial correlation term of the

item demands is not identical. The remainder of this paper is organized as follows: Section 2 describes the notation and assumptions used throughout the study. Section 3 presents the analytical analysis of the TD and BU forecasting strategies. Section 4 discusses the simulation findings. Finally, Section 5 presents some concluding remarks.

2. Notation and assumptions

We first define the following notations:

N = Number of items in the family.

α = Smoothing constant used in the exponential smoothing technique to forecast the aggregate demand, $0 < \alpha \leq 1$.

α_i = Smoothing constant used to forecast the demand for item i , $0 < \alpha_i \leq 1$, $i = 1, 2, \dots, N$.

$d_{i,t}$ = Demand for item i in period t , in units.

$D_t = \sum_{i=1}^N d_{i,t}$ = Aggregate family demand in period t , in units.

$\varepsilon_{i,t}$ = Error term of the demand process for item i in period t , normally distributed with zero mean [$E(\varepsilon_{i,t}) = 0$], variance σ_i^2 and zero autocovariance [$\text{Cov}(\varepsilon_{i,t}, \varepsilon_{i,(t-k)}) = 0$], for all $k \neq 0$.

$\hat{\varepsilon}_t = \sum_{i=1}^N \varepsilon_{i,t}$ = Error term of the aggregate demand in period t , where $\text{Cov}(\hat{\varepsilon}_t, \hat{\varepsilon}_{(t-k)}) = 0$, for all $k \neq 0$.

$F_{i,t}$ = Demand forecast for item i in period t , in units.

F_t = Forecast of the aggregate demand in period t , in units.

μ = Expected value of the aggregate demand, in units.

μ_i = Expected value of the demand for item i , in units.

p_i = Relative proportion of the value of the demand for item i over the value of the family demand, $\sum_i p_i = 1$.

ρ_{ij} = Coefficient of correlation between the error term of the demand process for items $i(\varepsilon_{i,t})$ and $j(\varepsilon_{j,t})$, $|\rho_{ij}| \leq 1$.

Due to its wide adoption in production-planning environments, simple exponential smoothing (SES) is used as the forecasting technique under both TD and BU strategies. Under the BU strategy, the demand forecast for the item i in period t is mathematically denoted as

$$F_{i,t} = \alpha_i d_{i,(t-1)} + (1 - \alpha_i) F_{i,(t-1)}, \quad (1)$$

which can also be written in an infinite form as

$$F_{i,t} = \sum_{k=1}^{\infty} \alpha_i (1 - \alpha_i)^{k-1} d_{i,(t-k)}. \quad (2)$$

Under the TD strategy, the demand forecast for the item i in period t is defined as

$$F_{i,t} = p_i F_t, \quad (3)$$

where $p_i = \mu_i / \mu$ is the relative proportion of item i in the family's expected aggregate demand. It is assumed that the relative proportion (p_i) is known with certainty or can be estimated with high accuracy using the historical demand data.

When the item demand time series follows an MA(1) process, the observation for the item i in period t is mathematically represented as

$$d_{i,t} = \mu_i + \varepsilon_{i,t} - \theta_i \varepsilon_{i,(t-1)}, \quad (4)$$

where θ_i is identified as a coefficient of the serial correlation term. In order to obtain a converging autoregressive representation (also known as ‘invertibility’ condition), we restrict $|\theta_i| < 1$ (Wei, 1993, p. 47). We also assume that $\theta_1 = \theta_2 = \dots = \theta_N$. This assumption is necessary because when $\theta_i \neq \theta_j$, $\forall i, j$, the detailed statistical properties of the aggregate family demand can only be obtained through approximation (Granger & Morris, 1976) and therefore, a precise analytical evaluation of the forecast error value would be difficult. This assumption, however, is relaxed in the simulation study which considers both identical and non-identical coefficients of the serial correlation term of the item demands.

We further assume that σ_i is significantly smaller than μ_i so that the probability of a negative demand value is negligible. It follows from (4) that (Wei, 1993, p. 47)

$$\text{Var}(d_{i,(t-k)}) = (1 + \theta_i^2)\sigma_i^2, \quad \text{for all } k, \quad (5)$$

$$\text{Cov}(d_{i,1}, d_{i,(1-k)}) = \text{Cov}(d_{i,2}, d_{i,(2-k)}) = \dots = \begin{cases} -\theta_i \sigma_i^2, & |k| = 1, \\ 0, & |k| > 1. \end{cases} \quad (6)$$

3. Analytical evaluation of TD and BU strategies

In this section, we analytically evaluate and compare the performance of the two forecasting strategies, TD and BU, based on the variance of forecast error. This performance measure is adopted as both TD and BU strategies provide unbiased estimates, meaning that in the long run, their expected forecast errors are equal to zero¹ [i.e. $E(d_{i,t} - F_{i,t}) = E(d_{i,t} - p_i F_t) = 0$]. The variance of the forecast error for the BU strategy (V_{BU}) is defined as

$$V_{\text{BU}} = \sum_{i=1}^N [\text{Var}(d_{i,t} - F_{i,t})], \quad (7)$$

while the variance of the forecast error for the TD strategy (V_{TD}) is determined as

$$V_{\text{TD}} = \sum_{i=1}^N [\text{Var}(d_{i,t} - p_i F_t)]. \quad (8)$$

We begin the analysis for V_{BU} by deriving the covariance between the demand of one item and its forecast. From (6), it can be shown that

$$\begin{aligned} \text{Cov}(F_{i,t}, d_{i,t}) &= \text{Cov} \left[\sum_{k=1}^{\infty} \alpha_i (1 - \alpha_i)^{k-1} d_{i,(t-k)}, d_{i,t} \right] \\ &= \alpha_i \sum_{k=1}^{\infty} (1 - \alpha_i)^{k-1} \text{Cov}(d_{i,(t-k)}, d_{i,t}) \\ &= -\alpha_i \theta_i \sigma_i^2. \end{aligned} \quad (9)$$

¹The proofs for this result is straightforward and available upon request.

Furthermore, the variance of the item demand forecast can also be obtained below:

$$\begin{aligned}\text{Var}(F_{i,t}) &= \text{Var} \left[\sum_{k=1}^{\infty} \alpha_i (1 - \alpha_i)^{k-1} d_{i,(t-k)} \right] \\ &= \lim_{M \rightarrow \infty} \left(\alpha_i^2 \sum_{k=1}^M (1 - \alpha_i)^{2(k-1)} \text{Var}(d_{i,(t-k)}) \right. \\ &\quad \left. + 2\alpha_i^2 \sum_{k=0}^{(M-2)} \sum_{j=1}^{(M-k-1)} (1 - \alpha_i)^{2k+j} \text{Cov}(d_{i,(t-1-k)}, d_{i,(t-1-k-j)}) \right).\end{aligned}$$

Taking the limit, $M \rightarrow \infty$, and using (5) and (6), we have

$$\text{Var}(F_{i,t}) = \frac{\alpha_i \sigma_i^2 [(1 - \theta_i)^2 + 2\alpha_i \theta_i]}{2 - \alpha_i}. \quad (10)$$

The variance of forecast error can thus be computed by substituting (5), (9) and (10) into $\text{Var}(d_{i,t} - F_{i,t}) = \text{Var}(d_{i,t}) + \text{Var}(F_{i,t}) - 2\text{Cov}(d_{i,t}, F_{i,t})$ and simplifying the resulting expression. We get

$$\text{Var}(d_{i,t} - F_{i,t}) = \frac{\sigma_i^2 (\theta_i^2 + \alpha_i \theta_i + 1)}{1 - 0.5\alpha_i}. \quad (11)$$

Note that the smoothing constant (α_i) is a parameter which is normally set to optimal by practitioners (based on minimizing the forecast error during the model-fitting period). Consequently, it is of interest to find the optimal α_i which minimizes the value of $\text{Var}(d_{i,t} - F_{i,t})$. From (11), it is obvious that $\text{Var}(d_{i,t} - F_{i,t})$ is monotonically increasing in α_i . Hence, α_i^* should be as small as possible; but in order to ensure that the forecast is an unbiased estimate, either the initial forecast has to be set equal to an unbiased estimate of the mean demand or α_i^* has to be set equal to a very small positive value $\delta > 0$. Finally, we can derive V_{BU} by substituting (11) into (7). We get

$$V_{\text{BU}} = \sum_{i=1}^N \frac{\sigma_i^2 (\theta_i^2 + \alpha_i^* \theta_i + 1)}{1 - 0.5\alpha_i^*}. \quad (12)$$

Note that if all the smoothing constants are optimized, they are all equal to a very small positive value $\delta > 0$. If in practice the smoothing constants are not optimized, but set to a convenient and reasonable small value such as 0.2, then it is reasonable to assume that all the smoothing constants are equal. In either event, $\alpha_1^* = \alpha_2^* = \dots = \alpha^*$. Using this and the fact that we assume $\theta_1 = \theta_2 = \dots = \theta_N = \hat{\theta}$, we can rewrite (12) as

$$V_{\text{BU}} = \frac{(\hat{\theta}^2 + \alpha^* \hat{\theta} + 1)}{1 - 0.5\alpha^*} \sum_{i=1}^N \sigma_i^2. \quad (13)$$

We now derive the variance of forecast error for the TD strategy. As explained earlier, we assume that $\theta_1 = \theta_2 = \dots = \theta_N = \hat{\theta}$. Accordingly, the family demand will also follow an MA(1) process with the

same value of the coefficient of the serial correlation term (Lutkepohl, 1984). Let the aggregate family demand be defined as

$$D_t = \mu + \hat{\varepsilon}_t - \hat{\theta}\hat{\varepsilon}_{t-1}, \quad (14)$$

where $\hat{\theta} = \theta_1 = \theta_2 = \dots = \theta_N$ and $\text{Var}(\hat{\varepsilon}_t) = \sum_{i=1}^N \text{Var}(\varepsilon_{i,t}) + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \text{Cov}(\varepsilon_{i,t}, \varepsilon_{j,t})$. The variance and autocovariance of D_t can then be determined as

$$\text{Var}(D_{t-k}) = (1 + \hat{\theta}^2) \left(\sum_{i=1}^N \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{ij} \sigma_i \sigma_j \right), \quad \text{for all } k, \quad (15)$$

$$\text{Cov}(D_t, D_{t-k}) = \begin{cases} -\hat{\theta} \left(\sum_{i=1}^N \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{ij} \sigma_i \sigma_j \right), & |k| = 1, \\ 0, & |k| > 1. \end{cases} \quad (16)$$

Based on (15) and (16), we can derive the covariance between the family demand and its forecast:

$$\begin{aligned} \text{Cov}(F_t, D_t) &= \text{Cov} \left[\sum_{k=1}^{\infty} \alpha (1 - \alpha)^{k-1} D_{t-k}, D_t \right] \\ &= -\alpha \hat{\theta} \left(\sum_{i=1}^N \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{ij} \sigma_i \sigma_j \right) \end{aligned} \quad (17)$$

and the variance of the family demand forecast:

$$\text{Var}(F_t) = [(1 + \hat{\theta}^2) - 2(1 - \alpha)\hat{\theta}] \alpha^2 \left(\sum_{i=1}^N \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{ij} \sigma_i \sigma_j \right) \left[\sum_{k=0}^{\infty} (1 - \alpha)^{2k} \right].$$

Simplifying the above expression further, we get

$$\text{Var}(F_t) = \frac{\alpha [(1 + \hat{\theta}^2) - 2\hat{\theta}(1 - \alpha)] \left(\sum_{i=1}^N \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{ij} \sigma_i \sigma_j \right)}{2 - \alpha}. \quad (18)$$

By substituting (15), (17) and (18) into $\text{Var}(D_t - F_t) = \text{Var}(D_t) + \text{Var}(F_t) - 2\text{Cov}(D_t, F_t)$ and simplifying, we get the variance of forecast error for the family demand as follows:

$$\text{Var}(D_t - F_t) = \frac{(\hat{\theta}^2 + \alpha\hat{\theta} + 1) \left(\sum_{i=1}^N \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{ij} \sigma_i \sigma_j \right)}{1 - 0.5\alpha}. \quad (19)$$

From (19), it is straightforward to demonstrate that the optimal smoothing constant (α^*) for the aggregate family demand is again the lowest possible value of α . That is, $\alpha^* = \delta > 0$.

We now derive the expressions for $\text{Cov}(d_{i,t}, D_{t-k})$ and $\text{Cov}(d_{i,t}, F_t)$ below:

$$\begin{aligned} \text{Cov}(d_{i,t}, D_{t-k}) &= \text{Cov}[p_i \mu + \varepsilon_{i,t} - \hat{\theta}\varepsilon_{i,t-1}, \mu + \hat{\varepsilon}_{t-k} - \hat{\theta}\hat{\varepsilon}_{t-k-1}] \\ &= \begin{cases} -\hat{\theta}\sigma_i \sum_{j=1}^N \rho_{ij} \sigma_j, & |k| = 1, \\ 0, & |k| > 1, \end{cases} \end{aligned} \quad (20)$$

and

$$\begin{aligned}\text{Cov}(d_{i,t}, F_t) &= \text{Cov}\left[d_{i,t}, \alpha \sum_{k=1}^{\infty} (1-\alpha)^{k-1} D_{t-k}\right] \\ &= -\alpha \hat{\theta} \sigma_i \sum_{j=1}^N \rho_{ij} \sigma_j.\end{aligned}\quad (21)$$

The variance of forecast error for the item i can thus be computed by substituting (5), (18) and (21) into $\text{Var}(d_{i,t} - p_i F_t) = \text{Var}(d_{i,t}) + p_i^2 \text{Var}(F_t) - 2p_i \text{Cov}(d_{i,t}, F_t)$:

$$\begin{aligned}\text{Var}(d_{i,t} - p_i F_t) &= (1 + \hat{\theta}^2) \sigma_i^2 + 2p_i \alpha \hat{\theta} \sigma_i \sum_{j=1}^N \rho_{ij} \sigma_j \\ &\quad + \frac{p_i^2 \alpha (1 + \hat{\theta}^2 - 2\hat{\theta} + 2\hat{\theta} \alpha) (\sum_{i=1}^N \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{ij} \sigma_i \sigma_j)}{2 - \alpha}.\end{aligned}\quad (22)$$

Finally, by substituting (22) into (8), we are able to determine the variance of forecast error for the TD strategy as shown below:

$$V_{\text{TD}} = \sum_{i=1}^N \left[(1 + \hat{\theta}^2) \sigma_i^2 + 2p_i \alpha^* \hat{\theta} \sigma_i \sum_{j=1}^N \rho_{ij} \sigma_j + \frac{p_i^2 \alpha^* (1 + \hat{\theta}^2 - 2\hat{\theta} + 2\hat{\theta} \alpha^*) (\sum_{i=1}^N \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \rho_{ij} \sigma_i \sigma_j)}{(2 - \alpha^*)} \right].$$

Rewriting the above, we get

$$V_{\text{TD}} = (1 + \hat{\theta}^2) \left(\sum_{i=1}^N \sigma_i^2 \right) + \frac{\varpi}{(2 - \alpha^*)}, \quad (23)$$

where

$$\begin{aligned}\varpi &= 2(2 - \alpha^*) \alpha^* \hat{\theta} \sum_{i=1}^N \left(p_i \sigma_i \sum_{j=1}^N \rho_{ij} \sigma_j \right) \\ &\quad + \alpha^* (1 + \hat{\theta}^2 - 2\hat{\theta} + 2\hat{\theta} \alpha^*) \left(\sum_{i=1}^N p_i^2 \right) \left(\sum_{l=1}^N \sigma_l^2 + 2 \sum_{l=1}^{N-1} \sum_{j=l+1}^N \rho_{lj} \sigma_l \sigma_j \right).\end{aligned}\quad (24)$$

As (23) and (24) are evaluated against (13) and recalling that $\alpha_1^* = \alpha_2^* = \alpha^* = \delta$, we get

$$V_{\text{TD}}/V_{\text{BU}} = \frac{(1 + \hat{\theta}^2)(2 - \alpha^*) + \varpi / \sum_{i=1}^N \sigma_i^2}{2(\hat{\theta}^2 + \alpha^* \hat{\theta} + 1)}.\quad (25)$$

We therefore have the following result.

THEOREM 1 If the time series of all the item demands follow an MA(1) process with $\theta_1 = \theta_2 = \dots = \theta_N$ and the smoothing constants used for forecasting the item demands under the BU strategy and the aggregate demand under the TD strategy are set to the optimal value, then the performance of the TD and BU strategy for forecasting the item demands is identical (i.e. $V_{TD} = V_{BU}$).

Proof. See Appendix. $\square$

In practice, managers may choose to work with a fixed value of smoothing constant (such as $\alpha = 0.2$ or $\alpha = 0.3$) rather than the optimal smoothing constant. We now analyse the relative performance of the two forecasting strategies when all the smoothing constants $\alpha_1 = \alpha_2 = \dots = \alpha_N = \alpha$ are set at an identical value that is not necessarily optimal. Based on empirical evidence and recommended practice, managers would normally only prefer to set the smoothing constant to be a small value in the range $0.1 \leq \alpha \leq 0.3$. So we will specifically consider only values of $\alpha = 0.1, 0.2$ and 0.3 . To facilitate the analysis, we will assume that the standard deviation of the error terms for all the items is of the same order or close to each other, i.e. $\sigma_i \approx \sigma$ for all $i = 1, \dots, N$. This eliminates σ_i from the expression in (25). Further analysis of expressions (25) and (24) reveals that the highest value of the ratio V_{TD}/V_{BU} is obtained when $\hat{\theta} = 1$, $\rho_{ij} \approx 1$, for $i = 1, \dots, N$, $j = 1, \dots, N$, and $\sum_i p_i^2$ is at its maximum. $\sum_i p_i^2$ achieves its maximum when one item accounts for a significant proportion of the aggregate demand. Clearly, when one or a few items account for a significant proportion of the aggregate family, it is best to separate them from the family and consider only items with a certain degree of homogeneity in their properties including demand values in the aggregate family.

THEOREM 2 If (i) the time series of all the item demands follow an MA(1) process with $\theta_1 = \theta_2 = \dots = \theta_N$, (ii) the smoothing constants used for forecasting the item demands under the BU strategy and the aggregate demand under the TD strategy are set equal to 0.2 and (iii) demand proportion of any item is less than twice the average proportion (i.e. $p_i \leq 2(1/N)$, $i = 1, \dots, N$), then $0.8182 \leq V_{TD}/V_{BU} \leq 1.2$.

Proof. See Appendix. $\square$

COROLLARY 2.1 When the smoothing constants are set equal to 0.3, in Theorem 2 above, then $0.7727 \leq V_{TD}/V_{BU} \leq 1.3$. When the smoothing constants are set equal to 0.1, in Theorem 2 above, then $0.8636 \leq V_{TD}/V_{BU} \leq 1.16$.

Proof. See Appendix. $\square$

COROLLARY 2.2 When the correlation coefficient $\rho_{ij} \leq 0$, for all i, j , then irrespective of the smoothing constants values used and the value of $\hat{\theta}$, TD forecast is better, i.e. $V_{TD}/V_{BU} \leq 1$.

Proof. See Appendix. $\square$

Theorem 1 shows that when the optimal smoothing constants are used, there is no difference between using TD or BU strategy for forecasting the items demands in a family. These findings contradict some of the earlier studies which usually supported either TD or BU forecasting strategy (Schwarzkopf *et al.*, 1988; Dangerfield & Morris, 1992). Theorem 2 and its corollaries show that even when the smoothing constants for determining the forecasts are not set to the optimal, the difference between TD and BU forecasting is not very significant. When the error terms of item demands are negatively correlated, clearly TD forecast would do better. But when the error terms of the item demands are positively correlated, the relative performance also depends on the value of $\hat{\theta}$. Therefore, when a specific value of the smoothing constant is used, BU forecast is generally preferred unless the error terms of the item demand are highly positively correlated. The numerical evaluation technique used for proving Theorem 2

and its corollaries also revealed that when a few items had a relatively high proportion of the demand, the demand correlations were highly positive and a fixed, non-optimal smoothing constant was used, TD dominated BU. This finding agrees with that of Dangerfield & Morris (1992) though theirs was an empirical study based on the M-competition data. Even though Dangerfield and Morris attempt to determine the optimal smoothing constant in their study, their data series is limited in size for each data set and hence perhaps the smoothing constant was not fully optimized.

4. Simulation study

We have shown in Section 3 that when the item demands follow a MA(1) process, the coefficients of the serial correlation terms are identical for all the item demands processes and the optimal smoothing constant is used, the performance of TD and BU forecasting strategies for forecasting the item demands are identical. In this section, we consider the situation when the coefficients of the serial correlation term for item demand processes are not identical. Unfortunately, in this case, analytical evaluation of the relative effectiveness of the two forecasting strategies becomes difficult. We therefore resort to a simulation study to evaluate the relative performance of TD and BU strategies.

In order to get meaningful insights, the simulation study was restricted to a family with two items. With more items in the family, it is difficult to meaningfully analyse experimental parameters such as the correlation coefficient ρ_{12} . Moreover, most of the earlier papers using the simulation approach have also restricted the number of items to two (Dangerfield & Morris, 1992; Fliedner, 1999). The simulation study was carried out for a number of different values of the various statistical parameters as shown in Table 1.

A particular combination of the value of the problem parameters constituted a single experiment. In each experiment, the simulation was conducted for a total of 884 periods. The first 104 periods was used to obtaining the initial value of F_{t-1} . The second part containing 260 periods was used to determine the optimal smoothing constant [by minimizing the forecast error (root mean squared error, RMSE)] for this part]. The optimal smoothing constant was determined by searching for all possible values of the smoothing constant between 0 and 1 in step sizes of 0.01. Note that for the case of BU forecasting, the smoothing constants were optimized for each item individually. This part containing 260 periods was also used to estimate the demand proportions p_i as explained later. Finally, the last 520 periods were used to compile the performance statistics for the two forecasting strategies. To evaluate the performance of the two forecasting strategies, we calculated the value of the root mean squared error for each strategy. The RMSE has a close relationship to the variance of forecast error that is used in the analytical part of this study. Using these two error measures allows us to directly relate the analytical results to the simulation results. For each experiment, the RMSE is calculated as $\sqrt{(\sum_{t=365}^{884} (d_{i,t} - F_{i,t})^2)/520}$. The relative benefit of one forecasting strategy over the other was measured by $\nabla = \text{RMSE(TD)}/\text{RMSE(BU)}$.

TABLE 1 *Problem parameters used in the experiment*

μ	$\sigma_1^2 (= \sigma_2^2)$	$p_1^\dagger$	$\theta_1 (= \theta_2)$	ρ_{12}
200 units	400 units ²	0.1 to 0.9 (in increments of 0.1)	-0.9 to 0.9 (in increments of 0.2)	-0.9 to 0.9 (in increments of 0.1)

[†]The relative proportion of the demand for item 2, $p_2 = 1 - p_1$.

FIG. 1. Average performance ratio between TD and BU forecasting strategies under different scenarios of θ_1 and θ_2 .

Clearly, the value of $\nabla < 1$ implies that the TD strategy is superior to the BU strategy, whereas the value of $\nabla > 1$ implies otherwise.

For the TD strategy, determining the item forecast for each item, i , requires an estimate of the demand proportion p_i . Note that in practice, the inventory manager would not be aware of the true value of demand proportion, p_i . However, the inventory manager would have access to past demand data and can therefore estimate the value of p_i for all items i . Even though we know the theoretical value of the p_i before the commencement of the simulation experiment, to replicate practice, we actually use an estimate of the demand proportion p_i that is calculated during the 260 periods from period 105 to 364.

The result of the simulation study is illustrated in Fig. 1. It plots the impact of varying the coefficient of correlation between the error terms of the item demand processes (ρ_{12}) on the average performance ratio of TD and BU forecasting strategies (∇) under different scenarios of θ_1 and θ_2 . Each data point plotted in the graph was obtained by computing the average value of ∇ across all possible combinations of p_1 , θ_1 and θ_2 as shown in Table 1.

In general, it can be seen that the difference in performance between the two strategies was remarkably small, with the value of ∇ very close to 1. This is in agreement to our analytical finding which suggested that the performance difference between TD and BU strategies was insignificant. This phenomenon also held true for the case with $\theta_1 \neq \theta_2$. In this case, the performance of the two forecasting strategies was even closer with each other as compared to the other cases when $\theta_1 = \theta_2$.

We also carried out tests with the smoothing constants fixed at a value of either 0.2 or 0.3 instead of being optimized. Table 1 is also indicative of the results we obtained in these cases.

5. Conclusions

We have analytically investigated the relative performance of TD and BU strategies for forecasting the demand of an item belonging to a product family under a production-planning environment. All the item demands, which might be correlated with each other, were assumed to follow a first-order univariate moving average [MA(1)] process. SES which is commonly adopted in a production-planning framework was used as the forecasting technique under both forecasting strategies. In contradiction to most of the earlier studies, we show that the difference in the mean and variance of forecast error between TD and BU strategies is relatively insignificant.

The analytical result in the paper was obtained only for the case where the coefficient of the serial correlation term of all the item demand processes is identical. Future studies could attempt to remove this weakness. The study was carried out only for item demands that follow MA(1) process. Item demands that follow AR(1) process have been treated in a companion paper by Widiarta *et al.* (2007), but the study could possibly be extended in the future to other demand processes and also to demand processes that exhibit seasonality.

REFERENCES

- BALLOU, R. H. (1999) *Business Logistics Management*, 4th edn. Upper Saddle River, NJ: Prentice-Hall.
- BARNEA, A. & LAKONISHOK, J. (1980) An analysis of the usefulness of disaggregated accounting data for forecasts of corporate performance. *Decis. Sci.*, **11**, 17–26.
- DANGERFIELD, B. J. & MORRIS, J. S. (1992) Top-down or bottom-up: aggregate versus disaggregate extrapolations. *Int. J. Forecast.*, **8**, 233–241.
- DELURGIO, S. A. (1998) *Forecasting Principles and Applications*, 1st edn. Boston, MA: Irwin McGraw Hill.
- DIEBOLD, F. X. (1998) *Elements of Forecasting*. Cincinnati, OH: Southwestern College Pub.
- DUNN, D. M., WILLIAMS, W. H. & SPIVEY, W. A. (1971) Analysis and prediction of telephone demand in local geographic areas. *Bell J. Econ. Manage. Sci.*, **2**, 561–576.
- FLIEDNER, G. (1999) An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation. *Comput. Oper. Res.*, **26**, 1133–1149.
- FOGARTY, D. W., HOFFMAN, T. R. & BLACKSTONE JR, J. H. (1991) *Production and Inventory Management*. Cincinnati, OH: South-Western Publishing.
- GORDON, T. P., MORRIS, J. S. & DANGERFIELD, B. J. (1997) Top-down or bottom-up: which is the best approach to forecasting? *J. Bus. Forecast. Methods Syst.*, **16**, 13–16.
- GRANGER, C. W. J. & MORRIS, M. J. (1976) Time series modelling and interpretation. *J. R. Stat. Soc. Ser. A*, **139**, 246–257.
- GROSS, C. W. & SOHL, J. (1990) Disaggregation methods to expedite product line forecasting. *J. Forecast.*, **9**, 233–254.
- GRUNFELD, Y. & GRILICHES, Z. (1960) Is aggregation necessarily bad? *Rev. Econ. Stat.*, **42**, 1–13.
- KAHN, K. B. (1998) Revisiting top-down versus bottom-up forecasting. *J. Bus. Forecast. Methods Syst.*, **17**, 14–19.
- LUTKEPOHL, H. (1984) Forecasting contemporaneously aggregated vector ARMA processes. *J. Bus. Econ. Stat.*, **2**, 201–214.
- MILLER, J. G., BERRY, W. & LAI, C.-Y. F. (1976) A comparison of alternative forecasting strategies for multi-stage production inventory systems. *Decis. Sci.*, **7**, 714–724.
- ORCUTT, G., WATTS, H. W. & EDWARDS, J. B. (1968) Data aggregation and information loss. *Am. Econ. Rev.*, **58**, 773–787.
- SCHWARZKOPF, A. B., TERSINE, R. J. & MORRIS, J. S. (1988) Top-down versus bottom-up forecasting strategies. *Int. J. Prod. Res.*, **26**, 1833–1843.
- SHLIFER, E. & WOLFF, R. W. (1979) Aggregation and proration in forecasting. *Manage. Sci.*, **25**, 594–603.
- WEATHERFORD, L. R., KIMES, S. E. & SCOTT, D. A. (2001) Forecasting for hotel revenue management—testing aggregation against disaggregation. *Cornell Hotel Restaur. Adm. Q.*, **42**, 53–64.
- WEI, W. W. S. (1993) *Time Series Analysis: Univariate and Multivariate Methods*. Redwood City, CA: Addison-Wesley Publishing.
- WIDIARTA, H., VISWANATHAN, S. & PIPLANI, R. (2007) On the effectiveness of top-down approach for forecasting autoregressive demands. *Nav. Res. Logist.*, **54**, 176–188.
- ZELLNER, A. & TOBIAS, J. (2000) A note on aggregation, disaggregation and forecasting performance. *J. Forecast.*, **19**, 457–469.

ZOTTERI, G. & KALCHSCHMIDT, M. (2007) A model for selecting the appropriate level of aggregation in forecasting processes. *Int. J. Prod. Econ.*, **108**, 74–83.

Appendix

Proof of Theorem 1. As discussed in the main body of the paper, the optimal value of the smoothing constant is to have it as close to zero as possible. Therefore, we need to show that $\lim_{\alpha^* \rightarrow 0} \{V_{TD}/V_{BU}\} = 1$. As

$$\begin{aligned} \varpi &= 2(2 - \alpha^*)\alpha^*\hat{\theta} \sum_{i=1}^N \left(p_i \sigma_i \sum_{j=1}^N \rho_{ij} \sigma_j \right) + \alpha^*(1 + \hat{\theta}^2 - 2\hat{\theta} + 2\hat{\theta}\alpha^*) \\ &\quad \times \left(\sum_{i=1}^N p_i^2 \right) \left(\sum_{l=1}^N \sigma_l^2 + 2 \sum_{l=1}^{N-1} \sum_{j=l+1}^N \rho_{lj} \sigma_l \sigma_j \right), \quad \lim_{\alpha^* \rightarrow 0} \varpi = 0. \end{aligned}$$

Therefore, $\lim_{\alpha^* \rightarrow 0} \{V_{TD}/V_{BU}\} = \lim_{\alpha^* \rightarrow 0} \frac{(1+\hat{\theta}^2)(2-\alpha^*)+\varpi/\sum_{i=1}^N \sigma_i^2}{2(\hat{\theta}^2+\alpha^*\hat{\theta}+1)} = \frac{(1+\hat{\theta}^2)2}{(1+\hat{\theta}^2)2} = 1$. $\square$

Proof of Theorem 2 and corollaries. For the case when the smoothing constant is fixed to a specific value that is not necessarily optimal, we assume that $\sigma_i \approx \sigma$. Then, the expression for V_{TD}/V_{BU} can be simplified as

$$\begin{aligned} &(1 + \hat{\theta}^2)(2 - \alpha) + 2(2 - \alpha)\alpha\hat{\theta} \sum_{j=1}^N \rho_{ij} + \alpha(1 + \hat{\theta}^2 - 2\hat{\theta} + 2\hat{\theta}\alpha) \\ \frac{V_{TD}}{V_{BU}} &= \frac{\times (\sum_{i=1}^N p_i^2) (1 + (2/N) \sum_{l=1}^{N-1} \sum_{j=l+1}^N \rho_{lj})}{2(\hat{\theta}^2 + \alpha\hat{\theta} + 1)}. \end{aligned} \quad (\text{A.1})$$

For a fixed value of α , the expression (A.1) above can be enumerated for all possible values of $\hat{\theta}$ for different extremes values of ρ_{ij} (+1, 0 and -1) and $\sum p_i^2$ (1/N to 2/N) and for different values of N ranging from $N = 2$ to very large values of N . Note that it is not possible to have $\rho_{ij} = -1, \forall i, j$. Therefore, the lower bound on V_{TD}/V_{BU} was generated using only the extreme values of $\rho_{ij} = 0, \forall i, j$. For the upper bound, $\rho_{ij} = +1, \forall i, j$, generated a larger value than $\rho_{ij} = -1, \forall i, j$. Expression (A.1) was enumerated for all possible extreme values of ρ_{ij} , $\sum p_i^2$ and N using excel, and the lower and upper bounds on V_{TD}/V_{BU} were numerically evaluated.

This gave the required results for Theorem 2 and its corollaries. $\square$