The effect of correlation between demands on hierarchical forecasting

Chen, H and Boylan, JE

http://dx.doi.org/10.1108/S1477-4070(2009)0000006011

Title	The effect of correlation between demands on hierarchical forecasting
Authors	Chen, H and Boylan, JE
Publication title	Advances in Business and Management Forecasting
Publisher	Emerald
Type	Book Section
USIR URL	This version is available at: http://usir.salford.ac.uk/id/eprint/19162/
Published Date	2009

USIR is a digital collection of the research output of the University of Salford. Where copyright permits, full text material held in the repository is made freely available online and can be read, downloaded and copied for non-commercial private study or research purposes. Please check the manuscript for any further copyright restrictions.

For more information, including our policy and submission procedure, please contact the Repository Team at: library-research@salford.ac.uk.

The effect of correlation between demands on hierarchical forecasting

Huijing Chen[*], and John E. Boylan[**]

Abstract

The forecasting needs for inventory control purposes are hierarchical. For SKUs in a product family or a SKU stored across different depot locations, forecasts can be made from the individual series' history or derived top-down. Many discussions have been found in the literature, but it is not clear under what conditions one approach is better than the other. Correlation between demands has been identified as a very important factor to affect the performance of the two approaches, but there has been much confusion on whether positive or negative correlation. This paper summarises the conflicting discussions in the literature, argues that it is negative correlation that benefits the top-down or grouping approach, and quantifies the effect of correlation through simulation experiments.

Introduction

Many organisations operate in a multi-item, multi-level environment. In general, they have to "*cope with well over 100 time series with numbers over 10,000 being quite common*" (Fildes and Beard, 1992). These time-series are often related. For example, a company may group similar products in product families according to specifications, colours, sizes etc. Alternatively, in a multi-echelon inventory system, a stock-keeping unit's sales may be recorded in many different locations at varying levels of aggregation. Therefore, in such cases, the data available and the need for forecasts are hierarchical.

A substantial part of the forecasting literature has been devoted to models and

---

[*] Corresponding author. Address: Room 616e, Salford Business School, University of Salford, Salford, M5 4WT, UK. Email: h.chen@salford.ac.uk
[**] Buckinghamshire New University, UK. Email: john.boylan@bucks.ac.uk

methods for single time series. However, as indicated above, the short-term forecasting need for production and inventory control purposes is to address a large amount of series simultaneously. Duncan *et al.* (1993) argued that, "*forecasting for a particular observational unit should be more accurate if effective use is made of information, not only from a time series on that observational unit, but also from time series on similar observational units*".

There have been many discussions on group forecasting in the literature. However, no clear conclusions have been reached on the conditions under which the grouping approach is better than the individual approach. Correlation between demands has been identified as a very important factor, but there has been much confusion about whether positive or negative correlation would benefit grouping. This paper is presented as follows: we will discuss contrasting arguments in the next section; then present our findings on the role of correlation from simulation experiments; and, finally, summarise our conclusions. The overall purpose of this paper is to dispel some of the confusion in the literature on how correlation affects the grouping approach.

Debates and confusion in the literature

It is well recognised that in order to obtain better forecasts, one should make better use of available forecasting series. Some practitioners such as Muir (1983), McLeavy and Narasimhan (1985) and Fogarty and Hoffmann (1983) have argued that forecasting an aggregate and then allocating it to items is more accurate than generating individual forecasts. Their argument was that the top-down approach resulted in more accurate predictions since aggregate data were more stable.

Schwarzkopf *et al.,* (1988) pointed out two problems of using the top-down approach: model incompleteness and positive correlation. They argued that the aggregate model may not completely describe the processes in the individual series, i.e. there were model differences among the series. When the total forecast was disaggregated

2

back to the item level, correlated errors were produced. They commented that "*this modelling error can be quite large and may override the more precise potential of top-down forecasts*" (Schwarzkopf *et al.,* 1988). The same point was also made by Shlifer and Wolff (1979). The second problem was that if there was a strong positive correlation in demand for items in a group, the variance for the aggregate was increased by the amount of the covariance term. Schwarzkopf *et al.,* (1988) advanced our understanding of some of the reasons why the top-down approach does not always lead to a more accurate subaggregate forecast.

Top-down forecasts have to be treated with caution. If the individual series follow different demand generating processes, then the aggregate model does not reflect any of those individual processes. Although the aggregate data are less noisy, it does not always result in more accurate subaggregate forecasts. Even when the modelling difference is not an issue, there is an additional problem of the disaggregation mechanism to be applied.

One way to get around these problems is to group seasonal homogeneous series. From a classical decomposition point of view, demand consists of level, trend, seasonality and noise. In a group of items, levels can be varying. Trends can be upwards or downwards and can have various degrees. However, seasonality is often more stable as it is affected by weather and customs. Chatfield (2004) pointed out that seasonal indices are usually assumed to change slowly through time so that $S_t \approx S_{t-q}$, where q is the seasonal cycle. It makes more sense to use the grouping approach to estimate seasonality than to estimate level and trend as there is an issue of modelling difference. The problem of an appropriate disaggregation mechanism can also be avoided. For multiplicative seasonality, no disaggregation mechanism is needed as seasonality is relative to the mean. For an additive model with common seasonal components across the group, a simple average can be used as the disaggregation method. Although it is difficult to apply the grouping approach in

general, we found it helpful in seasonal demand forecasting, i.e. estimating level and/or trend individually but seasonality from the group.

Correlation has been identified as a very important factor to affect the grouping and individual approaches, but there has been some confusion about whether positive or negative correlation benefits grouping. Duncan *et al.* (1998) argued for positive correlation. They claimed that analogous series should correlate positively (co-vary) over time. Then the covariation would be able to "*add precision to model estimates and to adapt quickly to time-series pattern changes*". However, Schwarzkopf *et al.* (1988) supported negative correlation as the covariance term was increased by positive correlation. The confusion lies in the distinction between a common model and varied models. Given the same model, it is negative correlation between series that reduces variability of the total and favours the top-down approach. However, the more consistent the model forms are, the more this favours the grouping approach; and consistency of model forms is associated with positive correlations between series, not negative correlations. Duncan *et al.* (1998) also identified the association between consistency of model forms and positive correlations. However, positive correlations should not be used to identify whether different series follow the same model, as sometimes the positive correlations may be incurred by a trend component rather than the model form. Therefore, checks should be made on the consistency of models using other diagnostics, before employing correlation analysis to establish whether a grouped or individual approach is preferable.

Simulation experiments

We used simulation experiments to examine the effect of correlation between demands on forecasting performance of the grouping and individual approaches. We have argued that it is negative correlation that will benefit the grouping approach when a common model is assumed; the main purpose of the simulation experiments is to quantify the effect.

4

Two simple models are assumed to generate demand:

$$Y_{i,th} = \mu_i S_h + \varepsilon_{i,th} \tag{1}$$

$$Y_{i,th} = \mu_i + S_h + \varepsilon_{i,th} \tag{2}$$

where i is a suffix representing the SKU or the location

suffix t represents the year and $t=1,2,\ldots r$ (where r is the number of years' data history)

suffix h represents the seasonal period and $h=1,2,\ldots q$ (where q is the length of the seasonal cycle)

Y represents demand

μ_i represents the underlying mean for the ith SKU or location and is assumed to be constant over time but different for different SKUs or locations

S_h represents a seasonal index at seasonal period h; it is unchanging from year to year and the same for all SKUs or locations under consideration

$\varepsilon_{i,th}$ is a random disturbance term for the ith SKU / location at the tth year and hth period; it is assumed to be normally distributed with mean zero and constant variance σ_i^2. There are correlations ρ_{ij} between $\varepsilon_{i,th}$ and $\varepsilon_{j,th}$ at the same time period. Auto-correlations and correlations at different time periods are assumed to be zero.

Model (1) has multiplicative seasonality and model (2) has additive seasonality. It is assumed there is no trend so that we may focus on the seasonality. The underlying mean is assumed to be stationary. Seasonality is also assumed to be stationary and the same within the group.

Trend components are not considered in the current models to avoid the complexity of different degrees of trend when aggregating, and thus focus mainly to gain insights

into the effect of correlation from simple models.

The estimator for the underlying mean is

$$\hat{\mu}_i = \frac{1}{qr} \sum_{t=1}^{r} \sum_{h=1}^{q} Y_{i,th} \tag{3}$$

The individual seasonal indices (ISI) estimator for the mixed model (Equation (1)) is:

$$\hat{S}_h = ISI_{i,h} = \frac{q \sum_{t=1}^{r} Y_{i,th}}{\sum_{t=1}^{r} \sum_{h=1}^{q} Y_{i,th}} \tag{4}$$

The individual seasonal indices (ISI) estimator for the additive model (Equation (2)) is:

$$\hat{S}_h = ISI_{i,h} = \frac{1}{r} \sum_{t=1}^{r} Y_{i,th} - \frac{1}{qr} \sum_{t=1}^{r} \sum_{h=1}^{q} Y_{i,th} \tag{5}$$

Two group seasonal indices (GSI) methods have been proposed from the literature. Dalhart (1974) proposed a method which was a simple average of the ISI.

$$DGSI_h = \frac{1}{m} \sum_{i=1}^{m} ISI_{i,h} = \frac{q}{m} \sum_{i=1}^{m} \frac{\sum_{t=1}^{r} Y_{i,th}}{\sum_{t=1}^{r} \sum_{h=1}^{q} Y_{i,th}} \tag{6}$$

Withycombe (1989) suggested aggregating all the individual series first and then estimating seasonal indices from the aggregate series:

$$WGSI_h = \frac{q \sum_{t=1}^{r} Y_{A,th}}{\sum_{t=1}^{r} \sum_{h=1}^{q} Y_{A,th}} \tag{7}$$

Both DGSI and WGSI were proposed to multiplicative seasonality. When

seasonality is additive, the two methods are the same and we call it GSI.

$$GSI_h \frac{1}{mr} \sum_{t=1}^{r} Y_{A,th} - \frac{1}{mqr} \sum_{t=1}^{r} \sum_{h=1}^{q} Y_{A,th}$$ (8)

We have developed rules to choose the best method between the ISI and GSI methods. Interested readers can refer to Chen and Boylan (2007).

Simulation findings

The simulation results quantify the effect of correlation on the forecasting performance of the individual and grouping approaches. We use mean square error (MSE) as the error measure and report the percentage best (PB) results.

Results for the additive model are presented first, followed by results for the mixed model. Detailed simulation designs are presented in Appendix 1.

Correlation	-0.9	-0.6	-0.3	0	0.3	0.6	0.9
ISI	40.00%	41.43%	43.84%	45.63%	48.57%	50.00%	50.00%
GSI	60.00%	58.57%	56.16%	54.38%	51.43%	50.00%	50.00%

Table 1: Effect of correlation on the percentage of series

for which ISI or GSI is the best (additive model)

Table 1 shows that negative correlation favours GSI. As the correlation coefficient changes from highly negative to highly positive, the number of series for which GSI is the best decreases. This is consistent with the theory that as correlation changes from highly negative to highly positive, ISI will be the best for more series. When the correlation coefficient is 0.6 or 0.9, ISI and GSI are equally good.

Correlation	-0.9	-0.6	-0.3	0	0.3	0.6	0.9
ISI	0.00%	0.00%	0.00%	1.13%	13.96%	29.55%	48.78%
DGSI	64.11%	52.92%	46.67%	40.65%	27.35%	11.16%	5.71%
WGSI	35.89%	47.08%	53.33%	58.21%	58.69%	59.29%	45.51%

Table 2: Effect of correlation on the percentage of series
for which ISI or GSI is the best (mixed model)

The above table shows that for the mixed model, ISI is never the best when correlation is negative. DGSI is the best when correlation is highly negative (between -0.9 and –0.6), and the number of series for which DGSI is the best decreases when correlation increases. The number of series for which WGSI is the best increases as correlation increases. But for a very high positive correlation coefficient of 0.9, ISI becomes the best method. Simulation results clearly show that GSI is better than ISI for the majority of possible correlations within the range.

The case of two series is simplistic, although it provides useful insight into the effect of correlation. In reality, the number of items in a group can be as large as hundreds or even thousands. To cover more realistic situations, we now proceed to simulate groups of more than two series (detailed design can be found in the Appendix 1).

The group size has to be determined somewhat arbitrarily. In this simulation, we define the group size to be 2^n (n=1, 2, 3, …,6). So the group sizes are 2, 4, 8, 16, 32 and 64. The group size increases with an unequal and growing increment because when the group size is smaller, we want to examine the effect at a finer level. When the size is larger, it is increasingly difficult to do so. The maximum group size is 64 because of the time and complexity of computing.

Correlation coefficients cannot be decided arbitrarily as in the case of a group of two series, nor can the correlation matrix be generated randomly. A feasible correlation matrix must be positive semi-definite, i.e. all the eigenvalues must be non-negative

(see, for example, Xu and Evers (2003)). Therefore, we followed the algorithm suggested by Lin and Bendel (1985) to generate feasible matrices with specified eigenvalues.

Ideally we would like to cover a comprehensive set of correlation matrices, but the number of possible combinations of feasible matrices makes this impossible. Instead, we will look at a number of feasible correlation matrices covering a range as large as possible.

Correlation does not affect the ISI method. For DGSI, it is $\sum_{j=1}^{m-1} \sum_{l=j+1}^{m} \frac{1}{\mu_j} \frac{1}{\mu_l} \rho_{jl} \sigma_j \sigma_l$

that matters. For WGSI, it is σ_A^2, which equals $\sigma_1^2 + \sigma_2^2 + ... + \sigma_m^2 + 2\sum_{j=1}^{m-1} \sum_{l=j+1}^{m} \rho_{jl} \sigma_j \sigma_l$.

It is the term $\sum_{j=1}^{m-1} \sum_{l=j+1}^{m} \rho_{jl} \sigma_j \sigma_l$ that involves the correlation coefficients (Chen and

Boylan, 2007). However, for both DGSI and WGSI, it is not straightforward from the theoretical expressions how correlation affects the rules. The standard deviations (coefficients of variation in DGSI) are interacting with the correlation coefficients and cannot be separated. What we want to see is what structure of correlation matrix affects the rules and we will do this by calculating the lower and upper bounds of the cross terms (details can be found in Appendix 2).

Let $P^+ = \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \rho_{ij}^+$ and $P^- = \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \rho_{ij}^-$; for simulation purposes, we can experiment

with different values of P^+ and P^- to evaluate the effect of correlation. With the bounds, the σ_i terms are separated from the correlation coefficients. However, in reality the cross term is not a simple function of the correlation coefficients but the interaction of correlation coefficients and the standard deviation terms. For given σ_i terms, the cancellation depends not only on the values of P^+ and P^- but also

9

on the positions of the positive and negative coefficients. From a simulation perspective, it is difficult to experiment with both sign and position of each correlation coefficient. Therefore, we bring the problem down to the two dimensions of P^+ and P^-.

We will generate 1000 different feasible correlation matrices for each group size n. It is a very small proportion of all possible combinations of feasible correlation matrices. We cannot use all of these feasible correlation matrices in our simulation to examine the effect of correlation along with other parameters. Just as we vary all the other parameters that affect the rules, we will vary P^+ and P^- too. Out of the 1000 feasible correlation matrices we generate, we will calculate $\left| \dfrac{P^+}{P^-} \right|$ and then choose the minimum, the first quartile, the second quartile (median), the third quartile and the maximum. This covers the whole range of the correlation matrices we generated. Then these 5 matrices are used in the simulations and their interactions with other parameters can be assessed. The following table shows the range of $\left| \dfrac{P^+}{P^-} \right|$ for each group size.

Group size	4	8	16	32	64
Minimum	0.0000	0.0847	0.2313	0.4436	0.6497
Lower Quartile	0.1674	0.4142	0.5719	0.6851	0.7592
Median	0.3500	0.5715	0.6919	0.7637	0.8260
Upper Quartile	0.6267	0.7639	0.8287	0.8678	0.9031
Maximum	2.7847	1.4352	1.0817	1.0217	0.9975

Table 3:Range of ratios of positive and negative correlation coefficients

When the additive model is assumed, GSI outperformed ISI universally. Therefore, we cannot analyse the effect of the different correlation matrices on ISI and GSI. However, the effect is analysed for the mixed model in the following table.

10

Correlation matrix	1	2	3	4	5
DGSI	76.05%	52.90%	50.99%	45.48%	41.97%
WGSI	23.95%	47.10%	49.01%	54.52%	58.03%

Table 4:Effect of correlation matrix on the percentage of series

for which DGSI or WGSI is the best

For each group size, five different correlation matrices are chosen in our simulation experiments according to different ratios of $\left|\frac{P^+}{P^-}\right|$. Matrix 1 has the lowest $\left|\frac{P^+}{P^-}\right|$ and matrix 5 has the highest $\left|\frac{P^+}{P^-}\right|$. ISI is never the best. When $\left|\frac{P^+}{P^-}\right|$ increases, the percentage of series for which DGSI is the best decreases and the percentage of series for which WGSI is the best increases. This is what we expected. Simulation results from group of two series show that DGSI was the best when correlation was between –0.9 to –0.6, WGSI was the best when correlation was between –0.3 to 0.6 and beyond that ISI became the best. Therefore, the greater the sum of all negative correlation coefficients, the more series for which DGSI would be expected to be the best.

Previous research on the issue of grouping has consistently suggested correlation as the most important factor to decide whether a direct forecast or a derived (top-down) forecast should be used. However, there have been arguments on whether series with positive or negative correlation favours the derived approach. Our simulation results reveal that for a wide range of positive correlation values GSI methods are still better than the ISI method but the gain of using the GSI methods is greater when series are negatively correlated.

Our simulation of two series in a group is much more specific than previous research: it does not only show the range of correlation that a GSI method outperforms the ISI method, but also it shows the range of correlation for which one GSI method

outperforms the other.   Within the former range (-0.9 to 0.6 in our simulation), DGSI outperforms WGSI when correlation is between –0.9 and –0.6 and WGSI is better when correlation is between –0.3 to 0.6.   It is not until correlation is almost as high as 0.9 that ISI becomes the best performing method.

When there are more than two series in the group, it is more difficult to find a clear cut how correlation affects the individual and grouping approaches.   Our simulations of up to 64 items in a group and five different correlation matrices show that ISI is never better than the grouping approach.   Moreover, we found that DGSI is better for lower $\left|\dfrac{P^+}{P^-}\right|$ and WGSI is better for higher $\left|\dfrac{P^+}{P^-}\right|$.   This is consistent with the findings in the case of two series.

Extension to models with trend

The current models we assume are simple ones without a trend component.   A key finding is that correlation between demands is induced only by correlation between the error terms in the model.

Take the additive model $Y_{i,th} = \mu_i + S_h + \varepsilon_{i,th}$

The deseasonalised demand is: $Y_{i,th}^* = Y_{i,th} - \hat{S}_h$

Since $\hat{S}_h$ (Equation (5)) is an unbiased estimator, $E(\hat{S}_h) = S_h$

$$
\begin{aligned}
\text{cov}(Y_{i,th}^* Y_{j,th}^*) &= E\{[(Y_{i,th} - \hat{S}_h - E(Y_{i,th}^*))][(Y_{j,th} - \hat{S}_h - E(Y_{j,th}^*))]\} \\
&= E[(\mu_i + S_h + \varepsilon_{i,th} - \mu_i - S_h)(\mu_j + S_h + \varepsilon_{j,th} - \mu_j - S_h) \qquad (9) \\
&= E(\varepsilon_{i,th}\varepsilon_{j,th})
\end{aligned}
$$

Therefore, the only source of correlation between demands is from correlation between the random error terms.

We can extend the analysis to an additive trend and seasonal model.

Assume $Y_{i,th} = \mu_i + [(t-1)q + h]\beta_i + S_h + \varepsilon_{i,th}$ (10)

where $[(t-1)q + h]\beta_i$ is the trend term and β_i is the growth rate.

Suppose we can find an estimator $\hat{\beta}_i$ for β_i, then to detrend the model we have:

$Y_{i,th} - [(t-1)q + h]\hat{\beta}_i = \mu_i + [(t-1)q + h](\beta_i - \hat{\beta}_i) + S_h + \varepsilon_{i,th}$ (11)

The detrended and deseasonalised demand is:

$Y_{i,th}^{*} = Y_{i,th} - [(t-1)q + h]\hat{\beta}_i - S_h$ (12)

Therefore, assuming $\beta_i - \hat{\beta}_i$ is independent of $\beta_j - \hat{\beta}_j$ and $\beta_i - \hat{\beta}_i$ is independent

of $\varepsilon_{j,th}$,

$$\begin{aligned} \mathrm{cov}(Y_{i,th}^{*}Y_{j,th}^{*}) &= E[(Y_{i,th}^{*} - \mu_i)(Y_{j,th}^{*} - \mu_j)] \\ &= E\{([(t-1)q + h](\beta_i - \hat{\beta}_i) + \varepsilon_{i,th})([(t-1)q + h](\beta_j - \hat{\beta}_j) + \varepsilon_{j,th})\} \\ &= E(\varepsilon_{i,th}\varepsilon_{j,th}) \end{aligned} \quad (13)$$

This result shows that correlation between the demands is induced only by correlation between the error terms in the model.   This is the same as equation (9); so the same result carries through from a non-trended model to a trended model.

This same approach does not apply for the mixed model though.   It will require a different approach to investigate the effect of correlation assuming a multiplicative trend and seasonality model.

Further research can also extend beyond stationary seasonality and consider time-varying Winters' type models.   This line of research is undertaken by another group of researchers (Dekker et al, 2004; Ouwehand, 2006).   They derived a model

13

to underlie the multivariate version of the Holt-Winters' method, i.e. estimating level and trend individually and seasonal indices from the group. However, the effect of correlation has yet to be addressed.

Conclusions

This paper clarifies some of the confusion in the literature regarding how top-down forecasts might improve on individual forecasts, especially the effect of correlation on the top-down approach. In the literature there were arguments about whether positive or negative correlation would benefit the top-down approach. We conducted simulation experiments, assuming series share a common model and common seasonality within a group, to quantify the effect of correlation on the individual and grouping approaches in terms of forecasting accuracy. Our simulation results reveal that, when there are two items in the group, the individual approach outperforms the grouping approach only when the correlation is very strongly positive. The grouping approach is better than the individual approach most of the time, with the benefit greater when correlation is negative. When there are more than two items in the group, the individual approach never outperforms the grouping approach in our simulations. DGSI is better for lower $\left| \dfrac{P^+}{P^-} \right|$ and WGSI is better for higher $\left| \dfrac{P^+}{P^-} \right|$.

Our current models do not take into account trend components. However, we have demonstrated that, for the additive model, the correlation between demands comes from the random error terms, with or without trend.

The conclusions from this paper are general. Further research can build on the results and insights offered by this paper and investigate the effect of correlation between demands by examining different models and assumptions.

References:

Brown, R.G. (1959) *Statistical Forecasting for Inventory Control*, McGraw-Hill, New York.

Chatfield, C. (2004) *The Analysis of Time Series*, 6th edition, Chapman & Hall/CRC, London.

Chen, H. and Boylan, J.E. (2007) Use of Individual and Group Seasonal Indices in Subaggregate Demand Forecasting, *Journal of the Operational Research Society*, d.o.i: 10.1057/palgrave.jors.2602310.

Dalhart, G. (1974) Class Seasonality—A New Approach, *American Production and Inventory Control Society Conference Proceedings*.

Dekker, M., van Donselaar, K., and Ouwehand, P. (2004) How to Use Aggregation and Combined Forecasting to Improve Seasonal Demand Forecasts, *International Journal of Production Economics*, 90, 151-167.

Duncan, G., Gorr, W. and Szczypula. J. (1993) Bayesian Forecasting for Seemingly Unrelated Time Series: Application to Local Government Revenue Forecasting, *Management Science*, 39, 275-293.

Duncan, G., Gorr, W. and Szczypula. J. (1998) Forecasting Analogous Time Series, Working Paper 1998-4, Heinz School, Carnegie Mellon University.

Fildes, R. and Beard C. (1992) Forecasting Systems for Production and Inventory Control, *International Journal of Operations and Production Management*, 12 (5), 4-27.

Fogarty, D.W., and Hoffmann, T.R. (1983) *Production and Inventory Management*, South-Western Publishing Co., Cincinnati.

Lin, S.P. and Bendel, R.B. (1985) Algorithm AS213: Generation of Population Correlation Matrices with Specified Eigenvalues, *Applied Statistics*, 34, 193-198.

McLeavey, D.W. and Narasimhan, S.L. (1985) *Production Planning and Inventory Control*, Allyn and Bacon, Inc., Boston, MA.

Muir, J. W. (1983) Problems in Sales Forecasting Needing Pragmatic Solutions, *APICS Conference Proceedings*, 4-7.

Ouwehand, P. (2006) *Forecasting with Group Seasonality*, Unpublished PhD thesis, Technische Universiteit Eindhoven, the Netherlands.

Schwarzkopf, A.B., Tersine, R.J. and Morris, J.S., (1988) Top-down versus Bottom-up Forecasting Strategies, *International Journal of Production Research*, 26, 1833-1843.

Shlifer E. and Wolff, R.W., (1979) Aggregation and Proration in Forecasting, *Management Science*, 25, 594-603.

Withycombe, R. (1989) Forecasting With Combined Seasonal Indices, *International Journal of Forecasting*, 5, 547-552.

Xu, K. and Evers, P.T., (2003) Managing Single Echelon Inventories Through Demand Aggregation and the Feasibility of a Correlation Matrix, *Computers and Operations Research*, 30, 297-308.

Appendix 1: Simulation designs

Two series

Quarterly seasonality was assumed in the simulations with four different seasonal profiles as the following two tables show.

	Q1	Q2	Q3	Q4
No Seasonality (NS)	0	0	0	0
Weak Seasonality (WS)	-5	-10	5	10
Low,Low,Low,High (LLLH)	-20	-15	-15	50
Low,High,Low,High (LHLH)	-25	25	-25	25

Table 1: Seasonal Profiles for the Additive Model

	Q1	Q2	Q3	Q4
No Seasonality (NS)	1	1	1	1
Weak Seasonality (WS)	0.9	0.8	1.1	1.2
Low,Low,Low,High (LLLH)	0.6	0.7	0.7	2
Low,High,Low,High (LHLH)	0.5	1.5	0.5	1.5

Table 2: Seasonal Profiles for the Mixed Model

The aim is not to attain comprehensiveness of seasonal profiles, but to choose a few commonly occurring profile shapes to check whether they affect the rules. WS represents a weak seasonality. LLLH represents a situation where there is a single very high season (e.g. in the final quarter of the year, with higher demand before Christmas). LHLH represents alternative low and high seasons.

The underlying mean for one item is fixed to be 50, and the mean of the other item in the group varies. It can take a value of 50, 100, 200, 300, 400, 500, 5000 or 50000, representing a ratio of 1, 2, 4, 6, 8, 10, 100, or 1000.

Variances of the random error terms in the models are generated using power laws of the form $\sigma^2 = \alpha\mu^\beta$, where μ is the underlying mean, and α and β are constants (Brown, 1959). Our preliminary results agreed with Shlifer and Wolff (1979) that the α parameter does not affect the rules because it appears on both sides of the rule

and can be cancelled out. Therefore, only the β parameter is allowed to vary in these power laws. We choose α to be 0.5 and β to be 1.2, 1.4, 1.6 or 1.8.

Variances of series of a group may follow power laws, but different series in a group may not follow the same power law. Therefore, we also simulate situations in which non-universal power laws are applied on a group. Series 1 in the group follows one law and series 2 follows the other law.

Series 1: $\sigma_i^2 = 0.75 \times 0.5 \mu_i^{1.5}$

Series 2: $\sigma_i^2 = 1.25 \times 0.5 \mu_i^{1.5}$

Alternatively, it may be assumed that the series follow no power laws. In this case, various combinations of mean and variance values have been identified, somewhat arbitrarily, for experimentation, as shown in Table 3:

Mean1		50	50	50	50	50	50	50	50
Mean2		50	100	200	300	400	500	5000	50000
No law									
low	V1	100	100	100	100	100	100	100	100
low	V2	100	225	1600	2500	3600	4900	62500	1562500
low	V1	100	100	100	100	100	100	100	100
high	V2	400	900	4900	8100	10000	22500	490000	49000000
high	V1	400	400	400	400	400	400	400	400
low	V2	100	225	1600	2500	3600	4900	62500	1562500
high	V1	400	400	400	400	400	400	400	400
high	V2	400	900	4900	8100	10000	22500	490000	49000000

Table 3: Arbitrary Variance Values

Data history is set to be 3, 5 or 7 years with the last year's observations used as the holdout sample. So the estimation periods are 2, 4 or 6 years.

The correlation coefficient is set to be –0.9, -0.6, -0.3, 0, 0.3, 0.6 and 0.9. This covers a wide range of correlation coefficients from highly negative to highly positive. These are correlations between the random variables in the model; they are also correlations between deseasonalised demands.

More than two series

We assume that the underlying mean values in a group follow a lognormal distribution. The details can be found in the following table:

Standard ratio	standard deviation	mean of the logarithm 4	6
2	0.69	69	513
6	1.79	272	2009
10	2.30	774	5716
30	3.40	17749	131147

Table 4: Mean values of the lognormal distribution

Each combination (2 means $\times$ 4 standard ratios) is replicated 50 times. MSE values are averaged over the 50 replications and then the results are compared. The purpose of replicating the lognormal distributions is to reduce randomness, especially when the group size is small (e.g. 4 items in the group) as the lognormal distribution may not be apparent. Such replication of distributions can also reduce the risk of some unusual values distorting the simulation results. For each replication of the lognormal distributions, 500 replications of the simulation are run. So, for each parameter setting, a total of 25,000 replications are run: 50 to replicate the lognormal distribution and 500 to replicate the estimation and forecasting process to reduce randomness (for each of the 50 distribution replications).

Variances are generated using only the universal power laws. The β parameter takes the values of 1.2, 1.4, 1.6 and 1.8. Non-universal power laws or arbitrary variance values are not examined in this chapter, owing to the greatly increased complexity of specifying the values.

Appendix 2: Simulating correlation matrices

Let $s = \sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\rho_{ij}\sigma_i\sigma_j = \sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\rho_{ij}^{+}\sigma_i\sigma_j - \sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\rho_{ij}^{-}\sigma_i\sigma_j$ \hfill (1)

where $\rho_{ij}^{+} = \rho_{ij}$ if $\rho_{ij} > 0$ and $\rho_{ij}^{+} = 0$ otherwise. $\rho_{ij}^{-} = -\rho_{ij}$ if $\rho_{ij} < 0$ and $\rho_{ij}^{-} = 0$ otherwise.

$$s \leq \sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\rho_{ij}^{+}\sigma_{max}^2 - \sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\rho_{ij}^{-}\sigma_{min}^2 = \sigma_{max}^2\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\rho_{ij}^{+} - \sigma_{min}^2\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\rho_{ij}^{-} \qquad (2)$$

By a similar argument, $s \geq \sigma_{min}^2\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\rho_{ij}^{+} - \sigma_{max}^2\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\rho_{ij}^{-}$, where σ_{min}^2 is the minimum variance and σ_{max}^2 is the maximum variance.

Given all the σ values, it is clear that the sum of the positive correlation coefficients and the negative coefficients can be used to determine bounds on the cross-term corresponding to WGSI.

The same argument applies for DGSI. Let $s' = \sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\rho_{ij}\dfrac{\sigma_i}{\mu_i}\dfrac{\sigma_j}{\mu_j}$,

$$CV_{min}^2\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\rho_{ij}^{+} - CV_{max}^2\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\rho_{ij}^{-} \leq s' \leq CV_{max}^2\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\rho_{ij}^{+} - CV_{min}^2\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}\rho_{ij}^{-} \qquad (3)$$

where CV_{min}^2 is the minimum coefficient of variation squared and CV_{max}^2 is the maximum coefficient of variation squared.