Journal of the Operational Research Society

ISSN: 0160-5682 (Print) 1476-9360 (Online) Journal homepage: https://www.tandfonline.com/loi/tjor20

Use of individual and group seasonal indices in subaggregate demand forecasting

H Chen & J E Boylan

Journal of the Operational Research Society (2007) 58, 1660–1671 0160-5682/07 $30.00

www.palgrave-journals.com/jors

Use of individual and group seasonal indices in subaggregate demand forecasting

H Chen[1]* and JE Boylan[2]

[1]University of Salford, UK; and [2]Buckinghamshire Chilterns University College, Buckinghamshire, UK

An individual seasonal indices (ISI) method and two group seasonal indices (GSI) methods proposed in the literature are compared, based on two models. Rules have been established to choose between these methods and insights are gained on the conditions under which one method outperforms the others. Simulation findings confirm that using the rules improves forecasting accuracy against universal application of these methods.

Journal of the Operational Research Society (2007) **58,** 1660–1671. doi:10.1057/palgrave.jors.2602310
Published online 11 October 2006

Keywords: Seasonality; forecasting; grouping

Introduction

A substantial part of the forecasting literature has been devoted to models and methods for single time series, but most businesses operate multi-level systems. Therefore, the data available and the need for forecasts are hierarchical. The hierarchy can be product-oriented. For example, a company groups similar products in product families according to specifications, colours, sizes, etc. Alternatively, the hierarchy can be geographically oriented. A multi-echelon inventory system may have a main distribution centre, covering a whole country or many countries. Under the main distribution centre are regional distribution centres, which can be further branched out to local centres or outlets. The individual items from product groups and/or different locations are at the subaggregate level, while the groups are at the aggregate level.

The short-term forecasting task for sales and operations management is to address many items simultaneously. The conventional forecasting approach is to extrapolate each stock-keeping unit (SKU)'s data series individually. 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'. If each SKU within a group or across depot locations has the same underlying seasonality, then seasonality estimated at the aggregate level can be expected to be more accurate than that estimated individually.

Previous work on the conditions under which a grouped approach is better than individual extrapolations has not led to clear conclusions. Some research (Shlifer and Wolff, 1979; Schwarzkopf *et al*, 1988; Dangerfield and Morris, 1988, 1992) has been done without reference to seasonal demand. However, there has been confusion in the literature as to what factors affect the performance of aggregated and disaggregated approaches. Correlation has been identified as a very important factor, but it has not been clarified whether positive or negative correlation would benefit grouping. Duncan *et al* (1998) argued for positive correlation. They claimed that analogous series should correlate positively (co-vary) over time. However, Schwarzkopf *et al* (1988) supported negative correlation. 'If there is a strong positive correlation in demand for items in a group, the variance for the family is increased by the amount of the covariance term'. 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. This is formally established in our paper. However, the more consistent the model forms are, the more this favours the top-down (aggregated) approach; and consistency of model forms is associated with positive correlations between series, not negative correlations. 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.

Schwarzkopf *et al* (1988) have identified a problem of the top-down approach: although the forecast at the aggregate level may be accurate, this gain in accuracy may be lost at the subaggregate level when the total forecast is disaggregated. In seasonal demand forecasting, the aggregate level can be used to help subaggregate level forecasting without the potential problems caused by the disaggregation mechanism. If seasonality is multiplicative, for example, no disaggregation mechanism is needed, as seasonality is relative to the mean.

*Correspondence: *H Chen, Business School, University of Salford, The Crescent, Salford, M5 4WT, UK.*
E-mail: h.chen@salford.ac.uk, jboyla01@bcuc.ac.uk

Previous empirical studies (Bunn and Vassilopoulos, 1993, 1999, Dekker *et al*, 2004) have shown there is scope for improving forecasting accuracy by grouping. However, no theory has been developed to understand under what conditions the grouping approach would be better than the individual approach. Motivated by the lack of such theoretical development, this paper presents theoretical rules to decide when to use the grouping approach.

This paper is organized as follows: two group seasonal indices (GSI) methods suggested in the literature are described; then two models are introduced, one with multiplicative seasonality and the other with additive seasonality; rules are derived from the models, and the estimators of the parameters used in the rules are analysed; the following section quantifies the improvement in forecasting accuracy by using the rules; some guidelines are given on how to implement the rules; and finally, conclusions are summarized.

GSI methods

Two GSI methods have been proposed in the literature: one by Dalhart (1974) and the other by Withycombe (1989).

Under the assumption that 'whatever causes the seasonal fluctuation in demand operates *the same* on all products within the line' (Withycombe, 1989, author's own emphasis), Withycombe believed that estimating seasonal indices from the group was better than from the individual series. Dalhart (1974) made the same assumption that all subaggregate series had 'a consistent underlying seasonal behaviour' and argued that 'the seasonal component is never isolated from the noise component'.

Dalhart (1974) proposed a group seasonal estimation method by averaging the individual seasonal indices (ISI). Let $\mathbf{S}_i = [a_{i1}, a_{i2}, \ldots, a_{iq}]$, where a_{ij} is the seasonal index for item i at season j, $\mathbf{S}_i$ is the multiplicative seasonal index vector for item i, and q is the period of the seasonal cycle. $\mathbf{S}_{\mathrm{DGSI}} = 1/m\sum_{i=1}^{m}\mathbf{S}_i$, where $\mathbf{S}_{\mathrm{DGSI}}$ is the group seasonal vector of indices estimated by Dalhart's group seasonal index (DGSI) method and m is the number of series in the group. Therefore, Dalhart's method is a simple average of the ISI.

Withycombe (1989) proposed a different method to obtain GSI (Withycombe group seasonal indices (WGSI)). He totalled all the series in the group and then estimated 'combined seasonal indices' from this single time series. Therefore, Withycombe's method is a weighted average of the ISI.

Models and assumptions

Seasonality can be handled in two ways: the first is to seasonally adjust the data, construct a suitable forecast for the level and trend, and then incorporate the seasonal effects; the second is to build the seasonal effects directly into the forecasting procedure. The methods described by Dalhart (1974) and Withycombe (1989) are both seasonal adjustment methods; they are independent of level- and trend-forecasting methods. This paper also follows this approach.

Two models are proposed, as follows:

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

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

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; and $\varepsilon_{i,th}$ 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 cross-correlations ρ_{ij} between $\varepsilon_{i,th}$ and $\varepsilon_{j,th}$ at the same time period. Autocorrelations and cross-correlations at different time periods are assumed to be zero.

The models are stationary with either multiplicative or additive seasonality. Seasonality is assumed to be unchanging from year to year and the same across SKUs or locations. The assumption of common seasonality was imposed by both Dalhart (1974) and Withycombe (1989). 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. Especially for a short period of observations, it is difficult to distinguish between fixed seasonality and evolving seasonality. Seasonality is assumed to be fixed from year to year. For methods and models dealing with evolving seasonality, see Dekker *et al* (2004), and Ouwehand *et al* (2005). We also assume that there is no trend in the models, in order to concentrate on the seasonal component alone. Further research is planned to include trend as well as seasonality.

The cross-correlation ρ_{ij} between two items can be either positive or negative. An intuitive way to interpret a positive correlation coefficient is between complementary products, that is, demand for one product may stimulate demand for the complementary product as well. On the other hand, negative correlation of demands may arise between substitute products.

Model (1) has a multiplicative seasonal component and an additive error term, and is called the 'mixed model'. Both Dalhart (1974) and Withycombe (1989) assumed multiplicative seasonality. In this paper, the models are extended to consider additive seasonality. In model (2), both seasonality and the error term are additive; this is called the 'additive model'.

Rules to choose between ISI and GSI methods

The rules are derived based on the mean squared errors (MSE). Results are presented separately for the mixed and the additive models. Detailed derivations of the rules can be found in the appendices.

The mixed model

For the mixed model, the following ISI estimator of the seasonal component S_h is used:

$$\mathrm{ISI}_{i,h} = \frac{q\sum_{t=1}^{r} Y_{i,th}}{\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{i,th}} \tag{3}$$

This is classical decomposition without the trend component. It is the same as the 'seasonal index method' described by Gilchrist (1976, pp 124–129) for a global non-trending model. In the following section, it is shown that this is a maximum likelihood estimator.

DGSI and WGSI estimators, for the mixed model, follow directly from the definition of the ISI estimator. They are as follows:

$$\mathrm{DGSI}_h = \frac{1}{m}\sum_{i=1}^{m} \mathrm{ISI}_{i,h} = \frac{q}{m}\sum_{i=1}^{m} \frac{\sum_{t=1}^{r} Y_{i,th}}{\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{i,th}} \tag{4}$$

$$\mathrm{WGSI}_h = \frac{q\sum_{t=1}^{r} Y_{A,th}}{\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{A,th}} \tag{5}$$

where $Y_{A,th}$ is the aggregate demand.

In order for the derivation of the rules for the mixed model to proceed, it is assumed that:

$$\hat{\mu}_i = p_i \hat{\mu}_1 \quad \text{for } i = 1, 2, \ldots, m \quad \text{and}$$
$$\hat{\mu}_A = (p_1 + p_2 + \cdots + p_m)\hat{\mu}_1$$
$$\mu_i = p_i \mu_1 \quad \text{for } i = 1, 2, \ldots, m \quad \text{and} \quad p_1 = 1.$$

where $\hat{\mu}_i$ is the estimate for μ_i, $\hat{\mu}_A$ is the estimate for μ_A, $\hat{\mu}_1$ is the estimate for the smallest mean value, and p_i is the ratio between μ_i and μ_1 ($p_i = \mu_i/\mu_1$).

The rules for comparing the MSEs of seasonal adjustment methods will now be given:

$$\mathrm{MSEISI}_i > \mathrm{MSEDGSI}_i \text{ if and only if}$$

$$\frac{\sigma_i^2}{\mu_i^2} > \frac{1}{m^2}\left(\frac{\sigma_1^2}{\mu_1^2} + \frac{\sigma_2^2}{\mu_2^2} + \cdots + \frac{\sigma_m^2}{\mu_m^2} + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m} \frac{1}{\mu_j}\frac{1}{\mu_l}\rho_{jl}\sigma_j\sigma_l\right) \tag{6}$$

where MSEISI_i denotes the expected squared errors for the ith series using ISI, and $\mathrm{MSEDGSI}_i$ denotes the expected squared errors for the ith series using the DGSI method.

Equation (6) shows that the greater the positive correlation, the less beneficial it is to use DGSI.

$$\mathrm{MSEISI}_i > \mathrm{MSEWGSI}_i \text{ if and only if}$$

$$\frac{\sigma_i^2}{\mu_i^2} > \frac{\sigma_A^2}{\mu_A^2} \tag{7}$$

where $\mathrm{MSEWGSI}_i$ denotes the expected squared errors for the ith series using the WGSI method.

Using DGSI or WGSI would benefit noisy series. The measure of noisiness is the coefficient of variation of deseasonalized demand. Conceptually, the two GSI approaches are the same. They compare individual noisiness of series to the 'average' noisiness. The difference lies in the expression of 'average' noisiness. DGSI uses the average of individual noisiness, while WGSI uses the noisiness of the aggregate series. The different expressions of 'average' noisiness correspond to the difference in the two GSI methods themselves. The results also show that the GSI methods would benefit from negative cross-correlations. Our rules give an indication of how correlation can affect the performance of the ISI and GSI methods, clarifying confusion over this issue in the literature. Of course, the results hold only if all items follow the same model. The issue of varied models is not considered here.

The above analysis compares DGSI and WGSI to the ISI method. When we consider for an individual series whether the ISI method, the DGSI method, or the WGSI method should be applied, we should calculate σ_i^2/μ_i^2, $\frac{1}{m^2}(\sigma_1^2/\mu_1^2 + \sigma_2^2/\mu_2^2 + \cdots + \sigma_m^2/\mu_m^2 + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}(1/\mu_j)(1/\mu_l)\rho_{jl}\sigma_j\sigma_l)$ and σ_A^2/μ_A^2. Synthesizing the results presented in this subsection for the mixed model:

- If σ_i^2/μ_i^2 is minimal, then choose the ISI method.
- If $\frac{1}{m^2}(\sigma_1^2/\mu_1^2 + \sigma_2^2/\mu_2^2 + \cdots + \sigma_m^2/\mu_m^2 + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}(1/\mu_j)(1/\mu_l)\rho_{jl}\sigma_j\sigma_l)$ is minimal, then choose the DGSI method.
- If σ_A^2/μ_A^2 is minimal, then choose the WGSI method.

These results have interesting implications. Suppose that there are two series, 1 and 2, and DGSI is ignored. It is possible that the following inequality may arise:

$\sigma_1^2/\mu_1^2 < \sigma_A^2/\mu_A^2 < \sigma_2^2/\mu_2^2$. In this, case, ISI would be the preferred method for series 1, but WGSI would be the preferred method for series 2. This would mean that the seasonality of series 1 would contribute to the forecasting of series 2, even though it has been isolated for its own forecasting. While this may appear counterintuitive, it is a logical consequence of the rules. Series 1 would not benefit from inclusion of noisier data, from series 2, and would only 'borrow weakness'; on the other hand, series 2 'borrows strength' from the less noisy series 1. Although the two series are in the same pool for common seasonality, their asymmetric treatment reflects their differential levels of noisiness, as measured by the squared coefficient of variation.

The additive model

Both DGSI and WGSI were originally established for multiplicative seasonality. The possible benefit of grouping for additive seasonality has not been explored. If additive seasonality is the same across SKUs or locations, the GSI methods might be more accurate than the ISI method due to the reduction in randomness. However, in this case, it should be noted that WGSI will have to be disaggregated. DGSI is the simple average of ISIs, so there is no need for disaggregation.

For the additive model, the following ISI estimator of the seasonal component S_h is used:

$$\mathrm{ISI}_{i,h} = \frac{1}{r}\sum_{t=1}^{r} Y_{i,th} - \frac{1}{qr}\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{i,th} \tag{8}$$

Again, this is classical decomposition without trend. In the following section, it is shown that this is a maximum likelihood estimator for the additive model.

In the case of the additive model, the WGSI estimator follows directly from the definition of ISI. Since it is assumed that the seasonal effects are identical across items, an appropriate disaggregation mechanism is a simple average. Therefore,

$$\begin{aligned}\mathrm{WGSI}_h &= \frac{1}{m}\left[\frac{1}{r}\sum_{t=1}^{r} Y_{A,th} - \frac{1}{qr}\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{A,th}\right]\\ &= \frac{1}{mr}\sum_{i=1}^{m}\sum_{t=1}^{r} Y_{i,th} - \frac{1}{mqr}\sum_{i=1}^{m}\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{i,th} = \mathrm{DSGI}_h\end{aligned}$$

Since, for the additive model, the two methods yield the same result, it is no longer necessary to distinguish between DGSI and WGSI. Instead, the term GSI is used and

$$\mathrm{GSI}_h = \frac{1}{mr}\sum_{t=1}^{r} Y_{A,th} - \frac{1}{mqr}\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{A,th} \tag{9}$$

where m is the number of SKUs/locations, and $Y_{A,th}$ is the aggregate demand.

$$\mathrm{MSEISI}_i - \mathrm{MSEGSI}_i = \frac{(q-1)}{qr}\left(\sigma_i^2 - \frac{\sigma_A^2}{m^2}\right) \tag{10}$$

And it follows that:

$$\mathrm{MSEGSI}_i < \mathrm{MSEISI}_i \text{ if and only if}$$

$$\sigma_i^2 > \frac{\sigma_A^2}{m^2} \tag{11}$$

The inequality (11) shows that noisier series in the group benefit from the GSI method. This rule depends only on the variance of the deseasonalized demand of the target series, the variance of the deseasonalized aggregate demand, and the number of series in the group. The difference in MSE between ISI and GSI depends on two more factors: the number of years of data history and the length of the seasonal cycle (see Equation (10)). The longer the seasonal cycle and/or the shorter the years of observations, the greater is the difference in MSE. When negative correlations are present, more series will benefit from the GSI method.

Estimators for the rule and statistical properties of the ISI, DGSI, and WGSI methods

To apply the rules to decide on the best seasonal method for the mixed model (using inequalities (6) and (7)) and the additive model (using inequality (11)), the following parameters must be estimated: mean demands at subaggregate and aggregate levels, variance of demand at subaggregate and aggregate levels and the correlation between the noise terms.

For the mixed model, it is shown in Appendix C that the following are maximum likelihood estimators:

$$\hat{\mu}_i = \frac{1}{qr}\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{i,th} \tag{12}$$

$$\hat{S}_h = \mathrm{ISI}_{i,h} = \frac{q\sum_{t=1}^{r} Y_{i,th}}{\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{i,th}} \tag{13}$$

$$\hat{\sigma}_i^2 = \frac{1}{qr}\sum\sum\left(Y_{i.th} - \frac{1}{r}\sum_{t=1}^{r} Y_{i,th}\right)^2 \tag{14}$$

DGSI is a simple average of the ISIs. Given ISI (Equation (13)), DGSI is presented as follows:

$$\mathrm{DGSI}_h = \frac{1}{m}\sum_{i=1}^{m}\mathrm{ISI}_{i,h} = \frac{q}{m}\sum_{i=1}^{m}\frac{\sum_{t=1}^{r} Y_{i,th}}{\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{i,th}} \tag{15}$$

WGSI aggregates all the individual series first and then estimate seasonal indices from the aggregate series:

$$\mathrm{WGSI}_h = \frac{q\sum_{t=1}^{r} Y_{A,th}}{\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{A,th}} \tag{16}$$

Both the ISI and WGSI methods are maximum likelihood estimators (see Appendix C). However, this property does not hold for the DGSI method.

For the mixed model, the ISI, DGSI, and WGSI methods are all biased for two reasons. First, we assume the estimated ratios between underlying means equals the population values (p_i) to allow the mathematics to proceed. It is a reasonable assumption for stationary mean models. Second, for multiplicative seasonality, seasonal indices take the form of a ratio. Since $E(a/b) \neq E(a)/E(b)$, the expected values of seasonal estimates are approximated using Taylor series.

Table 1 shows the approximate expectations of ISI, DGSI, and WGSI and therefore, also shows the degree of bias (derivations are presented in Appendix D):

The approximations shown in Table 1 will be checked by simulation experiments and the magnitude of the biases will be quantified.

Table 1 Expectations of ISI, DGSI, and WGSI

	Expectation
ISI	$S_h + \dfrac{\sigma_i^2(S_h-1)}{qr\mu_i^2}$
DGSI	$S_h + \dfrac{1}{m}\sum_{i=1}^{m}\dfrac{\sigma_i^2(S_h-1)}{qr\mu_i^2}$
WGSI	$S_h + \dfrac{\sigma_A^2(S_h-1)}{qr\mu_A^2}$

Recalling the model rule for DGSI: DGSI is better than ISI if:

$$\frac{\sigma_i^2}{\mu_i^2} > \frac{1}{m^2}\left(\frac{\sigma_1^2}{\mu_1^2} + \frac{\sigma_2^2}{\mu_2^2} + \cdots + \frac{\sigma_m^2}{\mu_m^2} + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\frac{1}{\mu_j}\frac{1}{\mu_l}\rho_{jl}\sigma_j\sigma_l\right) \quad (17)$$

Replacing the population values by estimators

$$\frac{\hat{\sigma}_i^2}{\hat{\mu}_i^2} > \frac{1}{m^2}\left(\frac{\hat{\sigma}_1^2}{\hat{\mu}_1^2} + \frac{\hat{\sigma}_2^2}{\hat{\mu}_2^2} + \cdots + \frac{\hat{\sigma}_m^2}{\hat{\mu}_m^2} + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\frac{1}{\hat{\mu}_j}\frac{1}{\hat{\mu}_l}\hat{\rho}_{jl}\hat{\sigma}_j\hat{\sigma}_l\right) \quad (18)$$

If the assumption that $\hat{\varepsilon}_{i,th}$ and $\hat{\varepsilon}_{A,th}$ are approximately normally distributed holds, then $\hat{\sigma}_i^2$ and $\hat{\sigma}_A^2$ are maximum likelihood estimators. They are both biased. Unbiased corrections for them would be

$$\hat{\sigma}_i^2 = \frac{1}{qr-1}\sum_{t=1}^{r}\sum_{h=1}^{q}\left(Y_{i,th} - \frac{1}{r}\sum_{t=1}^{r}Y_{i,th}\right)^2 \quad (19)$$

$$\hat{\sigma}_A^2 = \frac{1}{qr-1}\sum_{t=1}^{r}\sum_{h=1}^{q}\left(Y_{A,th} - \frac{1}{r}\sum_{t=1}^{r}Y_{A,th}\right)^2 \quad (20)$$

The estimators (19) and (20) are unbiased but have greater variance than the maximum likelihood estimators. Therefore, in the decision rule, there is a trade-off between unbiasedness and variance. But because all of the $\hat{\sigma}_i^2$s and $\hat{\sigma}_j\hat{\sigma}_l$ have common denominators on both sides of the inequality, they are cancelled. Therefore, the biases caused by the estimation of the variance do not affect the rule.

The final parameter requiring estimation is the correlation between the error terms. According to Kendall *et al* (1998), the sample correlation coefficient, $\hat{\rho}_{jl}$, is slightly biased and:

$$E(\hat{\rho}_{jl}) = \rho_{jl} - \frac{\rho_{jl}(1-\rho_{jl}^2)}{2qr} + \rho_{jl}O[(qr)^{-2}] \quad (21)$$

The number of observations, qr, is usually quite large, especially when there is monthly seasonality, that is, $q = 12$. Therefore, $\rho_{jl}O[(qr)^{-2}]$ is small enough to be neglected.

The expression that appears in the rule is $\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}(1/\hat{\mu}_j)(1/\hat{\mu}_l)\hat{\rho}_{jl}\hat{\sigma}_j\hat{\sigma}_l$. We believe the bias associated with this expression is small for two reasons. First, the bias of $\hat{\rho}_{jl}$ is small because qr is far greater than $\rho_{jl}(1-\rho_{jl}^2)$. And second, some of the biases are cancelled by the presence of both positive and negative cross-correlation coefficients. Therefore, it is concluded that (18) is a reasonable representation of the model rule.

Recalling the model rule for WGSI, WGSI is better than ISI if

$$\frac{\sigma_i^2}{\mu_i^2} > \frac{\sigma_A^2}{\mu_A^2} \quad (22)$$

Replacing the population values by estimators

$$\frac{\hat{\sigma}_i^2}{\hat{\mu}_i^2} > \frac{\hat{\sigma}_A^2}{\hat{\mu}_A^2} \quad (23)$$

where

$$\hat{\sigma}_i^2 = \frac{1}{qr}\sum_{t=1}^{r}\sum_{h=1}^{q}\left(Y_{i,th} - \frac{1}{r}\sum_{t=1}^{r}Y_{i,th}\right)^2 \quad (24)$$

$$\hat{\sigma}_A^2 = \frac{1}{qr}\sum_{t=1}^{r}\sum_{h=1}^{q}\left(Y_{A,th} - \frac{1}{r}\sum_{t=1}^{r}Y_{A,th}\right)^2 \quad (25)$$

Again, although both (24) and (25) are biased, the bias is caused by the common denominator and thus is cancelled. Therefore, the estimated rule for WGSI is not affected by the biases in parameter estimation.

For the mixed model, to choose the best seasonal indices method from the ISI method, the DGSI method, and the WGSI method, the following should be calculated $\hat{\sigma}_i^2/\hat{\mu}_i^2$, $1/m^2(\hat{\sigma}_1^2/\hat{\mu}_1^2 + \hat{\sigma}_2^2/\hat{\mu}_2^2 + \cdots + \hat{\sigma}_m^2/\hat{\mu}_m^2 + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}(1/\hat{\mu}_j)(1/\hat{\mu}_l)\hat{\rho}_{jl}\hat{\sigma}_j\hat{\sigma}_l)$ and $\hat{\sigma}_A^2/\hat{\mu}_A^2$, and the minimum should be found. Although each term is biased, cancellation of the qr terms results in a rule that is unaffected by parameter bias. Therefore, they are reliable to decide the best seasonal indices method.

The additive model is more straightforward. The ISI and GSI methods are both maximum likelihood estimators. This may be shown using the same method of proof as for the mixed model (Appendix C). The methods are also unbiased, and so no complications arise in the replacement of population values by estimators in the rule to choose between the ISI and GSI methods.

Simulation findings

The aims of the simulation experiments are to quantify the benefit of using the rules in terms of the reduction in MSE, and to check whether the biases in some of the estimators affect the rules.

Quarterly seasonality was assumed in the simulations with four different seasonal profiles as Tables 2 and 3 show.

The aim is not to attain comprehensiveness of seasonal profiles, but to choose a few commonly occurring profile

Table 2 Seasonal profiles for the additive model

	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 3 Seasonal profiles for the mixed model

	*Q*1	*Q*2	*Q*3	*Q*4
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

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 (eg in the final quarter of the year, with higher demand before Christmas). LHLH represents alternative low and high seasons.

The group sizes are 2, 4, 8, 16, 32, and 64. When the group size is 2, the underlying mean values are decided arbitrarily. 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 50 000, representing a ratio of 1, 2, 4, 6, 8, 10, 100, or 1000. When the group size is more than 2, underlying means in each group are assumed to be lognormally distributed. The use of lognormal distributions to model demand patterns has been supported by Brown (1963b, 1967), Wharton (1975), and Johnston *et al* (1988). In the simulations, the mean of the logarithm was set to be 4 or 6, and the standard ratio was set to be 2, 6, 10, or 30. Brown (1963a) reported that the standard ratio for companies that have other companies as customers would typically be 10. Retail companies would have a standard ratio of about 3. A company based on a rapidly changing technology and stocking many parts still used for older products may have a standard ratio as high as 25 or 30.

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. Stevens (1974) reported that the range of the β parameter is between 1.4 and 1.8. Johnston *et al* (1988) reported that a typical β value is between 1.38 and 1.88. When group size is 2, we also simulated σ^2 with non-universal power laws and arbitrary values.

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.

$$\text{Series 1}: \quad \sigma_i^2 = 0.75 \times 0.5\mu_i^{1.5}$$
$$\text{Series 2}: \quad \sigma_i^2 = 1.25 \times 0.5\mu_i^{1.5}$$

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 4.

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 cross-correlation coefficient is set to be −0.9, −0.6, −0.3, 0, 0.3, 0.6, and 0.9 when group size is 2. This covers a wide range of correlation coefficients from highly negative to highly positive. When group size grows, the cross-correlations become matrices and, it is impossible to cover a comprehensive set of correlation matrices. For each group size, we generated 1000 different feasible matrices and chose five of them in the simulations. Detailed designs were described by Chen (2005).

The reduction in MSE by using the rules against universal applications of ISI and GSI methods will be reported separately for a group size of 2 and the rest due to the slight differences in simulation design.

Groups of two Items

The average reductions in MSE by applying the rule against universal applications of ISI and GSI across all parameters and all cases for the additive model are 11.63 and 3.49%, respectively. Following the rule yields the smallest MSE, while universal application of ISI gives the largest MSE.

The average reductions in MSE by following the rule against universal applications of ISI, DGSI, and WGSI

Table 4 Arbitrary variance values

Mean1		50	50	50	50	50	50	50	50
Mean2		50	100	200	300	400	500	5000	50 000
No law									
Low	V1	100	100	100	100	100	100	100	100
Low	V2	100	225	1600	2500	3600	4900	62 500	1 562 500
Low	V1	100	100	100	100	100	100	100	100
High	V2	400	900	4900	8100	10 000	22 500	490 000	49 000 000
High	V1	400	400	400	400	400	400	400	400
Low	V2	100	225	1600	2500	3600	4900	62 500	1 562 500
High	V1	400	400	400	400	400	400	400	400
High	V2	400	900	4900	8100	10 000	22 500	490 000	49 000 000

Table 5 Effect of data length on the percentage reduction in MSE (additive model)

Data length	8	16	24
Rule versus ISI	17.19	10.32	7.37
Rule versus GSI	5.32	3.04	2.12

Table 6 Effect of data length on the percentage reduction in MSE (mixed model)

Data length	8	16	24
Rules versus ISI	6.91	4.26	3.04
Rules versus DGSI	16.86	13.18	11.14
Rules versus WGSI	2.29	1.43	1.01

Table 7 Effect of data length on the percentage reduction in MSE (additive model)

Data length	8	16	24
Rule versus ISI	22.94	13.77	9.84
Rule versus GSI	0.00	0.00	0.00

Table 8 Effect of data length on the percentage reduction in MSE (mixed model)

Data length	8	16	24
Rule versus ISI	20.98	12.78	9.13
Rule versus DGSI	0.45	0.23	0.15
Rule versus WGSI	0.08	0.08	0.06

for the mixed model are 4.74, 13.73, and 1.58%. Again, following the rule is the best. Universal application of DGSI is the worst because it is not robust across the parameter range. It performs poorly when the series are not noisy and/or the ratio between underlying means is great.

Data length has the greatest effect on the reduction in MSE, as Tables 5 and 6 show, while other factors have little effect on the reduction in MSE.

The number of discrepancies between theory and simulation is small: 0.01% for the additive model and 0.26% for the mixed model.

Groups of more than two items

When averaged across all parameters, the reduction in MSE of using GSI on the additive model against universal application of ISI is 15.52%. The average reductions in MSE of using the rule on the mixed model against universal application of ISI, DGSI, and WGSI are 14.30, 0.28, and 0.07%, respectively. DGSI performs much better when group size is greater than two.

Again, data length has the greatest effect on the reduction in MSE.

Tables 7 and 8 show that when data length increases, the reduction in MSE of using the rules decreases. The decrease is greatest against ISI. This shows that the greatest benefit of grouping occurs when data history is short. Comparing these two tables with Tables 5 and 6, it was found that the reduction in MSE is greater when group size increases.

The percentage of the occurrences of discrepancy between theory and simulation is 5.49%, higher than the percentage in groups of two items. But all discrepancies occurred when DGSI and WGSI were very close and it was very difficult to distinguish between the two methods.

Simulation results show that using the rule is better than universal applications of ISI and GSI methods. The reduction in MSE is the greatest against ISI; it is less so against GSI methods. But across the parameter ranges specified in the simulations, following the rules yields the lowest MSEs. Again, data length has the greatest effect on the reduction in MSE. The shorter the data length, the more beneficial GSI methods become against ISI.

Implementation of the rules

A rule has been developed to choose between the ISI, DGSI, and WGSI methods. All the methods are easy to use and the implementation of the rule is straightforward with all the parameter estimators given in this paper.

In order to apply the rule, a minimum of two years' data is required to estimate σ_i^2. The rule was developed for fast-moving items; models and methods for slow-moving demand would be different. The rule was established for stationary underlying mean and stationary seasonality without a trend component. How it behaves for non-stationary data and/or trended data is not clear. The rule does not take forecasting horizon into account, as it does not matter for stationary data.

If the seasonality is additive, a simpler rule can be applied. Assuming the demand follows the model $Y_{i,th}=\mu_i+S_h+\varepsilon_{i,th}$, DGSI and WGSI are the same, and therefore, the rule is simplified to compare the ISI and GSI methods. The GSI method is better than the ISI method if and only if $\sigma_i^2>\sigma_A^2/m^2$. It shows that GSI is better than ISI if the individual series' variance is greater than the 'average' of the group. This additive rule is easy to understand and easier to calculate than the rule for the multiplicative seasonality. The ISI and GSI methods are not biased in the additive model.

The rules can certainly be applied on product families. They can also be applied across depot locations if visibility of data at different locations can be achieved. The rules can be applied for vendor managed inventory to improve forecasting accuracy for both suppliers and their distributors. Our simulation results suggest that when the MSE of the methods are close, it is difficult for the rules to distinguish the best method. However, in those situations, it is not important to distinguish them.

For new products, where there are no data or very few data observations, the rules cannot be applied directly. If similar products can be identified for this new product, the rules may be applied on those similar series to establish initial seasonal indices.

Conclusions

In this paper, we have explored the potential improvement in estimating unchanging seasonality by using GSI methods over the traditional ISI method. The analyses are based on two stationary models. One model has multiplicative seasonality and an additive random disturbance term, and the other is an additive model because it assumes additive seasonality and an additive random disturbance term. For each model, a decision rule was formulated to determine the conditions under which GSI is preferred to ISI.

These rules, though different in their forms, show that noisier series benefit from GSI methods. This makes sense because noisier series can ‘borrow strength’ from those series that are less noisy. Simulation results confirmed that following the rules proposed in this paper yielded the lowest MSE against universal applications of ISI and GSI methods.

Another important factor is the cross-correlation coefficient. It is widely considered from previous research but also caused many debates and confusions. For example, the question whether a positive or negative correlation would benefit grouping has not been resolved. The lack of rigorous modelling has contributed towards these confusions. Our rules indicate that GSI methods benefit from negative correlations. The results are established assuming a common model within a group; the issue of varied models is not considered.

The factor that has the greatest effect on the reduction in MSE is data length. The shorter the data length, the more beneficial the GSI methods are against the ISI method.

Under our modelling specifications, our theory shows under what conditions the GSI methods are better than the ISI method. Using the rules in simulations, we found that considerable improvement on forecasting accuracy can be achieved, especially against universal application of ISI.

References

Brown RG (1959). *Statistical Forecasting for Inventory Control*. McGraw-Hill: New York.

Brown RG (1963a). Estimating aggregate inventory standards. *Nav Res Logist Quart* **10**: 55–71.

Brown RG (1963b). *Smoothing, Forecasting and Prediction of Discrete Time Series*. Prentice-Hall: Englewood Cliffs, NJ.

Brown RG (1967). *Decision Rules for Inventory Management*. Holt, Rinehart and Winston: Austin, TX.

Bunn DW and Vassilopoulos AI (1993). Using group seasonal indices in multi-item short-term forecasting. *Int J Forecasting* **9**: 517–526.

Bunn DW and Vassilopoulos AI (1999). Comparison of seasonal estimation methods in multi-item short-term forecasting. *Int J Forecasting* **15**: 431–443.

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

Chen H (2005). *Comparing individual with group seasonal indices to forecast subaggregate demand*. PhD thesis, Buckinghamshire Chilterns University College, Brunel University, UK.

Dalhart G (1974). Class seasonality—a new approach. Published in Forecasting, 2nd edition. American Production and Inventory Control Society, Washington DC, pp 11–16.

Dangerfield B and Morris JS (1988). An empirical evaluation of top-down and bottom-up forecasting strategies. *Proceedings of the 1988 Meeting of The Western Decision Sciences Institute*, pp. 322–324.

Dangerfield BJ and Morris JS (1992). Top-down or bottom-up: Aggregate versus disaggregate extrapolations. *Int J Forecasting* **8**: 233–241.

Dekker M, van Donselaar K and Ouwehand P (2004). How to use aggregation and combined forecasting to improve seasonal demand forecasts. *Int J Prod Econom* **90**: 151–167.

Duncan G, Gorr W and Szczypula J (1993). Bayesian forecasting for seemingly unrelated time series: Application to local government revenue forecasting. *Mngt Sci* **39**: 275–293.

Duncan G, Gorr W and Szczypula J (1998). *Forecasting analogous time series*. Working Paper 1998-4, Heinz School, Carnegie Mellon University.

Gilchrist W (1976). *Statistical Forecasting*. Wiley: London.

Johnston FR, Taylor SJ and Oliveria RMMC (1988). Setting company stock levels. *J Opl Res Soc* **39**: 15–21.

Kendall MG, Stuart A and Ord JK and Arnold SF (eds) (1998). *Advanced Theory of Statistics: Classical Inference*, Vol. 2A. Charles Griffin & Company Limited: London.

Ouwehand P, van Donselaar KH and de Kok AG (2005). *The impact of the forecasting horizon when forecasting with group seasonal indices*. Working Paper, Department of Technology Management, Eindhoven University of Technology.

Schwarzkopf AB, Tersine RJ and Morris JS (1988). Top-down versus bottom-up forecasting strategies. *Int J Prod Res* **26**: 1833–1843.

Shlifer E and Wolff RW (1979). Aggregation and proration in forecasting. *Mngt Sci.* **25**: 594–603.

Stevens CF (1974). On the variability of demand for families of items. *Op Res Quart* **25**: 411–419.

Wharton F (1975). On estimating aggregate inventory characteristics. *Op Res Quart* **26**: 543–551.

Withycombe R (1989). Forecasting with combined seasonal indices. *Int J Forecasting* **5**: 547–552.

Appendix A. Rule for the mixed model

$$\mathrm{ISI}_{i,h} = \frac{Y_{i,1h} + Y_{i,2h} + \cdots + Y_{i,rh}}{r\hat{\mu}_i}$$

The forecast for the ith series, hth season, in year r+1 using ISI is

$$\begin{aligned}\mathrm{FISI}_{i,(r+1)h} &= \hat{\mu}_i \, \mathrm{ISI}_{i,h} \\ &= \frac{Y_{i,1h} + Y_{i,2h} + \cdots + Y_{i,rh}}{r}\end{aligned}$$

MSE is

$$\begin{aligned}\mathrm{MSEISI}_i &= E\left(Y_{i,(r+1)h} - \frac{Y_{i,1h} + Y_{i,2h} + \cdots + Y_{i,rh}}{r}\right)^2 \\ &= \sigma_i^2\left(1 + \frac{1}{r}\right)\end{aligned}$$

$$\mathrm{DGSI}_h = \frac{\mathrm{ISI}_{1,h} + \mathrm{ISI}_{2,h} + \cdots + \mathrm{ISI}_{m,h}}{m}$$

$$= \frac{\dfrac{Y_{1,1h} + Y_{1,2h} + \cdots + Y_{1,rh}}{r\hat{\mu}_1} + \cdots + \dfrac{Y_{m,1h} + Y_{m,2h} + \cdots + Y_{m,rh}}{r\hat{\mu}_m}}{m}$$

The same forecast using DGSI is

$$\mathrm{FDGSI}_{i,(r+1)h} = p_i\hat{\mu}_1 \times \mathrm{DGSI}_h = \frac{p_i(Y_{1,1h} + Y_{1,2h} + \cdots + Y_{1,rh})}{mr} + \frac{p_i(Y_{2,1h} + Y_{2,2h} + \cdots + Y_{2,rh})}{mrp_2} + \cdots + \frac{p_i(Y_{m,1h} + Y_{m,2h} + \cdots + Y_{m,rh})}{mrp_m}$$

MSE is

$$\mathrm{MSEDGSI}_i = E\left(Y_{i,(r+1)h} - \frac{p_i(Y_{1,1h} + Y_{1,2h} + \cdots + Y_{1,rh})}{mrp_1} - \cdots - \frac{p_i(Y_{m,1h} + Y_{m,2h} + \cdots + Y_{m,rh})}{mrp_m}\right)^2$$

$$= E\left[\varepsilon_{i,(r+1)h} - \frac{p_i(\varepsilon_{1,1h} + \varepsilon_{1,2h} + \cdots + \varepsilon_{1,rh})}{mr} - \cdots - \frac{p_i(\varepsilon_{m,1h} + \varepsilon_{m,2h} + \cdots + \varepsilon_{m,rh})}{mrp_m}\right]^2$$

$$= \sigma_i^2 + \frac{p_i^2\sigma_1^2}{m^2r} + \cdots + \frac{p_i^2\sigma_m^2}{m^2rp_m^2} + 2\frac{p_i^2}{m^2r}\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\frac{1}{p_j}\frac{1}{p_l}\rho_{jl}\sigma_j\sigma_l$$

$$\mathrm{MSEISI}_i - \mathrm{MSEDGSI}_i$$

$$= \sigma_i^2\left(1 + \frac{1}{r}\right) - \left[\sigma_i^2 + \frac{p_i^2\sigma_1^2}{m^2r} + \cdots + \frac{p_i^2\sigma_m^2}{m^2rp_m^2} + 2\frac{p_i^2}{m^2r}\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\frac{1}{p_j}\frac{1}{p_l}\rho_{jl}\sigma_j\sigma_l\right]$$

$$= \frac{1}{r}\left[\sigma_i^2 - \frac{\mu_i^2}{m^2}\left(\frac{\sigma_1^2}{\mu_1^2} + \frac{\sigma_2^2}{\mu_2^2} + \cdots + \frac{\sigma_m^2}{\mu_m^2} + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\frac{1}{\mu_j}\frac{1}{\mu_l}\rho_{jl}\sigma_j\sigma_l\right)\right]$$

$\mathrm{MSEISI}_i > \mathrm{MSEDGSI}_i$ if and only if

$$\sigma_i^2 - \frac{\mu_i^2}{m^2}\left(\frac{\sigma_1^2}{\mu_1^2} + \frac{\sigma_2^2}{\mu_2^2} + \cdots + \frac{\sigma_m^2}{\mu_m^2} + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\frac{1}{\mu_j}\frac{1}{\mu_l}\rho_{jl}\sigma_j\sigma_l\right) > 0$$

$$\frac{\sigma_i^2}{\mu_i^2} > \frac{1}{m^2}\left(\frac{\sigma_1^2}{\mu_1^2} + \frac{\sigma_2^2}{\mu_2^2} + \cdots + \frac{\sigma_m^2}{\mu_m^2} + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\frac{1}{\mu_j}\frac{1}{\mu_l}\rho_{jl}\sigma_j\sigma_l\right)$$

$$\mathrm{WGSI}_h = \frac{\dfrac{Y_{1,1h}+Y_{2,1h}+\cdots+Y_{m,1h}}{\hat{\mu}_A} + \dfrac{Y_{1,2h}+Y_{2,2h}+\cdots+Y_{m,2h}}{\hat{\mu}_A} + \cdots + \dfrac{Y_{1,rh} + Y_{2,rh} + \cdots + Y_{m,rh}}{\hat{\mu}_A}}{r}$$

$$= \frac{Y_{1,1h} + \cdots + Y_{1,rh} + \cdots + Y_{m,1h} + \cdots + Y_{m,rh}}{r(p_1 + p_2 + \cdots + p_m)\hat{\mu}_1}$$

The forecast of hth season in year $r+1$ for the ith item using WGSI is

$$\mathrm{FWGSI}_{i,(r+1)h} = p_i\hat{\mu}_1 \times \mathrm{WGSI}_h = \frac{p_i(Y_{1,1h}+Y_{1,2h}+\cdots+Y_{1,rh}+\cdots+Y_{m,1h}+Y_{m,2h}+\cdots+Y_{m,rh})}{r(p_1 + p_2 + \cdots + p_m)}$$

MSE is

$$\mathrm{MSEWGSI}_i = E\left(Y_{i,(r+1)h} - \frac{p_i(Y_{1,1h}+Y_{1,2h}+\cdots+Y_{1,rh}+\cdots+Y_{m,1h}+Y_{m,2h}+\cdots+Y_{m,rh})}{r(p_1 + p_2 + \cdots + p_m)}\right)^2$$

$$= E\left[\varepsilon_{i,(r+1)h} - \frac{p_i(\varepsilon_{1,1h} + \cdots + \varepsilon_{1,rh} + \cdots + \varepsilon_{m,1h} + \cdots + \varepsilon_{m,rh})}{r(p_1 + p_2 + \cdots + p_m)}\right]^2$$

$$= \sigma_i^2 + \frac{p_i^2}{r(p_1 + p_2 + \cdots + p_m)^2} \times \left[\sigma_1^2 + \cdots + \sigma_m^2 + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\rho_{jl}\sigma_j\sigma_l\right]$$

$$\mathrm{MSEISI}_i - \mathrm{MSEWGSI}_i$$

$$= \sigma_i^2\left(1 + \frac{1}{r}\right) - \left(\sigma_i^2 + \frac{p_i^2}{r(p_1 + p_2 + \cdots + p_m)^2} \times \left[\sigma_1^2 + \cdots + \sigma_m^2 + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\rho_{jl}\sigma_j\sigma_l\right]\right)$$

$$= \frac{1}{r}\left[\sigma_i^2 - \frac{p_i^2}{(p_1 + p_2 + \cdots + p_m)^2} \times \left(\sigma_1^2 + \cdots + \sigma_m^2 + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\rho_{jl}\sigma_j\sigma_l\right)\right]$$

$\mathrm{MSEISI}_i > \mathrm{MSEWGSI}_i$ if and only if

$$\sigma_i^2 - \frac{p_i^2}{(p_1+p_2+\cdots+p_m)^2} \times \left(\sigma_1^2+\cdots+\sigma_m^2+2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\rho_{jl}\sigma_j\sigma_l\right) > 0$$

$$\sigma_i > \frac{p_i}{(p_1+p_2+\cdots+p_m)} \times \sqrt{\sigma_1^2+\cdots+\sigma_m^2+2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\rho_{jl}\sigma_j\sigma_l}$$

$$\frac{\sigma_i}{\mu_i} > \frac{\sigma_A}{\mu_A}$$

Appendix B. Rule for the additive model

The additive model is specified as

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

Forecast for item i, the hth season in year $r+1$ using ISI is

$$\mathrm{FISI}_{i,(r+1)h} = \mu_i + \mathrm{ISI}_{i,h} = \frac{1}{r}\sum_{t=1}^{r} Y_{i,th}$$

MSE of the forecast is

$$\begin{aligned}
&\mathrm{MSEISI}_i \\
&= E\left(Y_{i,(r+1)h} - \frac{1}{r}\sum_{t=1}^{r}Y_{i,th}\right)^2 \\
&= E\Bigg[\mu_i + S_h + \varepsilon_{i,(r+1)h} - \mu_i - S_h \\
&\quad -\frac{1}{r}\left(\varepsilon_{i,1h}+\varepsilon_{i,2h}+\cdots+\varepsilon_{i,rh}\right)\Bigg]^2 \\
&= \sigma_i^2 + \frac{\sigma_i^2}{r}
\end{aligned}$$

The GSI estimator is given as

$$\mathrm{GSI}_h = \frac{1}{mr}\sum_{i=1}^{m}\sum_{t=1}^{r}Y_{i,th} - \frac{1}{mqr}\sum_{i=1}^{m}\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{i,th}$$

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

Therefore, the forecast of item i, the hth season in year $r+1$ is

$$\begin{aligned}
\mathrm{FGSI}_{i,(r+1)h} &= \hat{\mu}_i + \mathrm{GSI}_h \\
&= \frac{1}{qr}\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{i,th} + \frac{1}{mr}\sum_{i=1}^{m}\sum_{t=1}^{r}Y_{i,th} \\
&\quad - \frac{1}{mqr}\sum_{i=1}^{m}\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{i,th}
\end{aligned}$$

$$E(\mathrm{FGSI}_{i,(r+1)h}) = \mu_i + S_h$$

MSE of the forecast is as follows:

$$\begin{aligned}
\mathrm{MSEGSI}_i &= E\Bigg(Y_{i,(r+1)h} - \frac{1}{qr}\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{i,th} \\
&\quad -\frac{1}{mr}\sum_{i=1}^{m}\sum_{t=1}^{r}Y_{i,th} + \frac{1}{mqr}\sum_{i=1}^{m}\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{i,th}\Bigg)^2 \\
&= E\Bigg(\mu_i + S_h + \varepsilon_{i,(r+1)h} - \mu_i - S_h - \frac{1}{qr}\sum_{t=1}^{r}\sum_{h=1}^{q}\varepsilon_{i,th} \\
&\quad -\frac{1}{mr}\sum_{i=1}^{m}\sum_{t=1}^{r}\varepsilon_{i,th} + \frac{1}{mqr}\sum_{i=1}^{m}\sum_{t=1}^{r}\sum_{h=1}^{q}\varepsilon_{i,th}\Bigg)^2 \\
&= \sigma_i^2 + \frac{\sigma_i^2}{qr} + \frac{\sigma_A^2}{m^2 r} + \frac{\sigma_A^2}{m^2 qr} \\
&\quad + \frac{2r\left(\sigma_i^2 + \sum_{j\neq i}\rho_{ij}\sigma_i\sigma_j\right)}{mqr^2} \\
&\quad - \frac{2\left(qr\sigma_i^2 + qr\sum_{j\neq i}\rho_{ij}\sigma_i\sigma_j\right)}{mq^2r^2} \\
&\quad - \frac{2r\left(\sigma_1^2+\sigma_2^2+\cdots+\sigma_m^2+2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\rho_{jl}\sigma_j\sigma_l\right)}{m^2qr^2} \\
&= \sigma_i^2 + \frac{\sigma_i^2}{qr} + \frac{\sigma_A^2}{m^2 r} - \frac{\sigma_A^2}{m^2 qr} \\
&= \left(1+\frac{1}{qr}\right)\sigma_i^2 + \frac{(q-1)\sigma_A^2}{m^2 qr}
\end{aligned}$$

The difference in MSE for the forecasts using ISI and GSI is:

$$\mathrm{MSEISI}_i - \mathrm{MSEGSI}_i = \frac{(q-1)\sigma_i^2}{qr} - \frac{(q-1)\sigma_A^2}{m^2 qr}$$

$\mathrm{MSEGSI}_i < \mathrm{MSEISI}_i$ if and only if

$$\frac{(q-1)\sigma_i^2}{qr} > \frac{(q-1)\sigma_A^2}{m^2 qr}$$

$$\sigma_i^2 > \frac{\sigma_A^2}{m^2}$$

Appendix C. Maximum likelihood estimators for the mixed model

$$Y_{i,th} = \mu_i S_h + \varepsilon_{i,th} \qquad \text{for } h = 1,\ldots,q \text{ and for } t = 1,\ldots,r$$

where $\sum_{h=1}^{q} S_h = q$, and $\varepsilon_{i,th} \sim_{iid} \mathrm{N}(0, \sigma_i^2)$.

$$\begin{aligned}
&L(\mu_i, S_1, \ldots, S_q, \sigma_i^2 | Y_{i,th}, h = 1,\ldots,q; t = 1,\ldots,r) \\
&= \prod_{t=1}^{r}\prod_{h=1}^{q}\left\{\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left[-\frac{1}{2}\left(\frac{Y_{i,th}-\mu_i S_h}{\sigma_i}\right)^2\right]\right\}
\end{aligned}$$

$$\log L=\sum_{t=1}^{r}\sum_{h=1}^{q}\left\{\log\left(\frac{1}{\sqrt{2\pi}\sigma_i}\right)-\frac{1}{2}\left(\frac{Y_{i,th}-\mu_i S_h}{\sigma_i}\right)^2\right\}$$

$$\log L=-\frac{1}{2}qr\log(2\pi)-qr\log\sigma_i-\frac{1}{2\sigma_i^2}\sum_{t=1}^{r}\sum_{h=1}^{q}(Y_{i,th}-\mu_i S_h)^2$$

Hence, differentiating with respect to μ_i and $S_h (h=1,\ldots,q)$ and with respect to σ_i^2,

$$\frac{\partial\log L}{\partial\mu_i}=\frac{1}{\sigma_i^2}\left(\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{i,th}S_h-r\mu_i\sum_{h=1}^{q}S_h^2\right)=0$$

$$\frac{\partial\log L}{\partial S_h}=\frac{1}{\sigma_i^2}\left(\sum_{t=1}^{r}Y_{i,th}\mu_i-r\mu_i^2 S_h\right)=0$$

for $h=1,\ldots,q$

$$\frac{\partial\log L}{\partial\sigma_i}=-\frac{qr}{\sigma_i}+\frac{\sum_{t=1}^{r}\sum_{h=1}^{q}(Y_{i,th}-\mu_i S_h)^2}{\sigma_i^3}=0$$

If the q equations $\partial\log L/\partial S_h=0$ $(h=1,\ldots,q)$ are all satisfied, then:

$$\sum_{t=1}^{r}Y_{ith}S_h=r\mu_i S_h^2 \qquad \text{for } h=1,\ldots,q$$

And so, summing all of these equations:

$$\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{ith}S_h=r\mu_i\sum_{h=1}^{q}S_h^2$$

This gives the same condition as the first equation $\partial\log L/\partial\mu_i=0$

So, if all of the q equations are satisfied, the first equation will also be satisfied. Solving these q equations gives:

$$\sum_{t=1}^{r}Y_{ith}=r\mu_i S_h \qquad \text{for } h=1,\ldots,q$$

Summing, $\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{ith}=r\mu_i\sum_{h=1}^{q}S_h$

Since the mixed model assumes that the seasonal indices sum to q (the number of seasons):

$$\hat{\mu}_i=\frac{1}{qr}\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{ith}$$

And the estimators for $S_h (h=1,\ldots,q)$ are given by:

$$\hat{S}_h=\frac{q\sum_{t=1}^{r}Y_{ith}}{\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{ith}} \qquad \text{for } h=1,\ldots,q$$

And the final condition gives: $\hat{\sigma}_i^2=\frac{1}{qr}\sum_{t=1}^{r}\sum_{h=1}^{q}(Y_{ith}-\hat{\mu}_i\hat{S}_h)^2$

We checked that all the criteria for maximum likelihood estimators have been satisfied. The maximum likelihood estimators for the aggregate mixed model and for the additive model can be obtained by the same method.

Appendix D. Bias properties of ISI, DGSI, and WGSI

Expectation of ISI :

$$\mathrm{ISI}_h=\frac{q(Y_{i,1h}+Y_{i,2h}+\cdots+Y_{i,rh})}{\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{i,th}} \tag{D.1}$$

$$E(\mathrm{ISI}_h)=qE\left(\frac{Y_{i,1h}+Y_{i,2h}+\cdots+Y_{i,rh}}{\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{i,th}}\right) \tag{D.2}$$

Use Taylor series to approximate $E(Y_{i,1h}+Y_{i,2h}+\cdots+Y_{i,rh}/\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{i,th})$

Let $X_1=Y_{i,1h}+Y_{i,2h}+\cdots+Y_{i,rh}$, $X_2=\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{i,th}$

$$\mathbf{X}=(X_1,X_2)$$

$$g(\mathbf{X})=g(X_1,X_2)=\frac{X_1}{X_2}$$

$$\theta_1=E(X_1)=r\mu_i S_h$$

$$\theta_2=E(X_2)=qr\mu_i$$

$$\boldsymbol{\theta}=(\theta_1,\theta_2)$$

$$g(\boldsymbol{\theta})=g(\theta_1,\theta_2)=\frac{\theta_1}{\theta_2}=\frac{S_h}{q}$$

$$g(\mathbf{X})=g(\boldsymbol{\theta})+\left[\frac{\partial g}{\partial\theta_1}(X_1-\theta_1)+\frac{\partial g}{\partial\theta_2}(X_2-\theta_2)\right]+\frac{1}{2}\left[\frac{\partial^2 g}{\partial\theta_1^2}(X_1-\theta_1)^2+2\frac{\partial^2 g}{\partial\theta_1\partial\theta_2}(X_1-\theta_1)\times(X_2-\theta_2)+\frac{\partial^2 g}{\partial\theta_2^2}(X_2-\theta_2)^2\right]+\cdots$$

$$\begin{aligned}E\left(\frac{Y_{i,1h}+Y_{i,2h}+\cdots+Y_{i,rh}}{\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{i,th}}\right)&=E\left(\frac{X_1}{X_2}\right)\\&=\frac{\theta_1}{\theta_2}+E\left[\frac{1}{\theta_2}(X_1-\theta_1)-\frac{\theta_1}{\theta_2^2}(X_2-\theta_2)\right]\\&\quad+\frac{1}{2}E\left[-\frac{2}{\theta_2^2}(X_1-\theta_1)(X_2-\theta_2)+\frac{2\theta_1}{\theta_2^3}(X_2-\theta_2)^2\right]\\&\quad+\cdots\\&\approx\frac{S_h}{q}-\frac{1}{\theta_2^2}E(X_1-\theta_1)(X_2-\theta_2)+\frac{\theta_1}{\theta_2^3}E(X_2-\theta_2)^2\\&=\frac{S_h}{q}-\frac{E(X_1X_2)-\theta_1\theta_2}{\theta_2^2}+\frac{\theta_1}{\theta_2^3}[E(X_2^2)-\theta_2^2]\\&=\frac{S_h}{q}-\frac{E(X_1X_2)}{q^2r^2\mu_i^2}+\frac{S_hE(X_2^2)}{q^3r^2\mu_i^2}\end{aligned}$$

$$E(\mathrm{ISI}_h)=S_h-\frac{E(X_1X_2)}{qr^2\mu_i^2}+\frac{S_hE(X_2^2)}{q^2r^2\mu_i^2} \tag{D.3}$$

Without loss of generality, suppose $h = 1$,

$$
\begin{aligned}
E(X_1X_2) &= E[(Y_{i,11} + Y_{i,21} + \cdots + Y_{i,r1})(Y_{i,11} + Y_{i,12} + \cdots \\
&\quad + Y_{i,1q} + \cdots + Y_{i,r1} + Y_{i,r2} + \cdots + Y_{i,rq})] \\
&= E(Y_{i,11} + Y_{i,21} + \cdots + Y_{i,r1})[(Y_{i,11} + Y_{i,21} + \cdots \\
&\quad + Y_{i,r1}) + (Y_{i,12} + \cdots + Y_{i,1q} + \cdots + Y_{i,r2} \\
&\quad + \cdots + Y_{i,rq})] \\
&= E\begin{bmatrix}
Y_{i,11}(Y_{i,11} + Y_{i,21} + \cdots + Y_{i,r1}) + Y_{i,11}(Y_{i,12} \\
\quad + \cdots + Y_{i,1q} + \cdots + Y_{i,r2} + \cdots + Y_{i,rq}) \\
+Y_{i,21}(Y_{i,11} + Y_{i,21} + \cdots + Y_{i,r1}) + Y_{i,21}(Y_{i,12} \\
\quad + \cdots + Y_{i,1q} + \cdots + Y_{i,r2} + \cdots + Y_{i,rq}) \\
+\cdots \\
+Y_{i,r1}(Y_{i,11} + Y_{i,21} + \cdots + Y_{i,r1}) + Y_{i,r1}(Y_{i,12} \\
\quad + \cdots + Y_{i,1q} + \cdots + Y_{i,r2} + \cdots + Y_{i,rq})
\end{bmatrix} \\
&= E\begin{bmatrix}
(\mu_i S_1 + \varepsilon_{i,11})\left(r\mu_i S_1 + \sum_{t=1}^{r} \varepsilon_{i,t1}\right) \\
\quad +(\mu_i S_1 + \varepsilon_{i,11}) \\
\quad \times\left(r\mu_i(S_2 + S_3 + \cdots + S_q) + \sum_{t=1}^{r}\sum_{h=2}^{q} \varepsilon_{i,th}\right) \\
+(\mu_i S_1 + \varepsilon_{i,21})\left(r\mu_i S_1 + \sum_{t=1}^{r} \varepsilon_{i,t1}\right) \\
\quad +(\mu_i S_1 + \varepsilon_{i,21}) \\
\quad \times\left(r\mu_i(S_2 + S_3 + \cdots + S_q) + \sum_{t=1}^{r}\sum_{h=2}^{q} \varepsilon_{i,th}\right) \\
+\cdots \\
+(\mu_i S_1 + \varepsilon_{i,r1})\left(r\mu_i S_1 + \sum_{t=1}^{r} \varepsilon_{i,t1}\right) \\
\quad +(\mu_i S_1 + \varepsilon_{i,t1}) \\
\quad \times\left(r\mu_i(S_2 + S_3 + \cdots + S_q) + \sum_{t=1}^{r}\sum_{h=2}^{q} \varepsilon_{i,th}\right)
\end{bmatrix} \\
&= r^2\mu_i^2 S_1^2 + r\sigma_i^2 + r^2\mu_i^2(S_1S_2 + S_1S_3 + \cdots + S_1S_q)
\end{aligned}
$$

Then for any h,

$$
\begin{aligned}
E(X_1X_2) &= r^2\mu_i^2 S_h^2 + r\sigma_i^2 + r^2\mu_i^2 S_h(q - S_h) \\
&= r\sigma_i^2 + qr^2\mu_i^2 S_h
\end{aligned} \tag{D.4}
$$

$$
\begin{aligned}
E(X_2^2) &= E\left[\left(\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{i,th}\right)^2\right] \\
&= E\{[(Y_{i,11} + Y_{i,12} + \cdots + Y_{i,1q}) + \cdots \\
&\quad + (Y_{i,r1} + Y_{i,r2} + \cdots + Y_{i,rq})]^2\} \\
&= E[(qr\mu_i + \varepsilon_{i,11} + \varepsilon_{i,12} + \cdots + \varepsilon_{i,rq})^2] \\
&= q^2r^2\mu_i^2 + qr\sigma_i^2
\end{aligned} \tag{D.5}
$$

Therefore,

$$
\begin{aligned}
E(\mathrm{ISI}_{i,h}) &= S_h - \frac{E(X_1X_2)}{qr^2\mu_i^2} + \frac{S_h E(X_2^2)}{q^2r^2\mu_i^2} \\
&= S_h - \frac{r\sigma_i^2 + qr^2\mu_i^2 S_h}{qr^2\mu_i^2} + \frac{S_h(q^2r^2\mu_i^2 + qr\sigma_i^2)}{q^2r^2\mu_i^2} \\
&= S_h + \frac{(S_h - 1)\sigma_i^2}{qr\mu_i^2}
\end{aligned} \tag{D.6}
$$

Expectation of DGSI:

$$
\begin{aligned}
E(\mathrm{DGSI}_h) &= E\left(\frac{\mathrm{ISI}_{1,h} + \mathrm{ISI}_{2,h} + \cdots + \mathrm{ISI}_{m,h}}{m}\right) \\
&= S_h + \frac{1}{m}\sum_{i=1}^{m} \frac{\sigma_i^2(S_h - 1)}{qr\mu_i^2}
\end{aligned} \tag{D.7}
$$

Expectation of WGSI:

Both ISI and WGSI are maximum likelihood estimators, one at the individual level and one at the aggregate level. $E(\mathrm{ISI}_{i,h}) = S_h + (S_h - 1)\sigma_i^2/qr\mu_i^2$ and, therefore, $E(\mathrm{WGSI}_h) = S_h + \sigma_A^2(S_h - 1)/qr\mu_A^2$.

Received January 2006;
accepted June 2006 after two revisions