Generalised Estimators for Seasonal Forecasting by Combining Grouping with Shrinkage Approaches

Kui Zhang[*a], Huijing Chen[a], John Boylan[b], Philip Scarf[a]

[a]*Centre of OR and Applied Statistics, University of Salford, M5 4WT, UK*

[b]*Faculty of Enterprise and Innovation, Buckinghamshire New University, HP11 2JZ, UK*

ABSTRACT

In this paper, generalised estimators are proposed to estimate seasonal indices for certain forms of additive and mixed seasonality. The estimators combine one of two group seasonal indices methods, Dalhart's group method and Withycombe's group method, with a shrinkage method in different ways. Optimal shrinkage parameters are derived to maximise the performance of the estimators. Then, the generalised estimators, with the optimal shrinkage parameters, are evaluated based on forecasting accuracy. Moreover, the effects of three factors are examined, namely, the length of data history, variance of random components and the number of series. Finally, a simulation experiment is conducted to support the evaluation.

Keywords: Forecasting; Seasonality; Grouping; Shrinkage; Generalised estimator

1. Introduction

Forecasting is an integral part of supply chain management. Accuracy of forecasts influences business decision-making at many levels including, for example, strategic planning, budgeting, resource allocation, production and inventory control. Since the demand of many products exhibits a seasonal pattern, accurate seasonal forecasting at the Stock-Keeping Unit (SKU) level plays an important role in many organisations.

---

[*]Corresponding author. Tel: +44 161 2732539

E-mail address: kuizzz@hotmail.com (or k.zhang@salford.ac.uk)

The common approach for seasonal forecasting is to use the individual item’s data history to forecast seasonal demand. This is called the Individual Seasonal Indices (ISI) method or classical decomposition, i.e., 'deseasonalise' the data, forecast the deseasonalised data, and then reseasonalise the forecast. However, the ISI method is not always satisfactory if the data are noisy and the length of data is short. An alternative approach to estimating seasonality is from a product group or the same SKU across depot locations. Dalhart (1974) and Withycombe (1989) proposed two different group seasonality estimation methods. Dalhart's Group Seasonal Index (DGSI) is a simple average of ~~i~~Individual ~~s~~Seasonal Indices (Dalhart, 1974), while Withycombe's Seasonal Index (WGSI) is calculated by totalling all the series in the group and then estimating the seasonal indices from this single time series (Withycombe, 1989). Dalhart (1974) compared the performance of DGSI and ISI by simulating 100 series with 24 periods of monthly demand data. The results showed that the average absolute error of the 100 forecasts was lower by using the DGSI method. Withycombe (1989) applied the WGSI method to real data for 29 individual products, representing 6 product lines from 3 different companies. 17 out of 29 products showed a decrease in Mean Square Error (MSE) and the total MSE for each product line is lower for all 6 product lines. Bunn and Vassilopoulos (1993) provided an empirical comparison of DGSI, WGSI and ISI. They used 54 weekly series from 5 product groups with 42 observations in each series. Their investigation revealed that the DGSI and WGSI outperformed the ISI and that WGSI is better than DGSI. However, no reasons were given why this was the case and no theoretical analysis was presented to evaluate under what conditions one method was better than another. Chen and Boylan (2007) conducted a comparison between the individual seasonal indices method and two group seasonal indices (GSI) methods and ~~discovered the conditions under which one method outperforms the others~~ established theoretical rules to choose the best method.

Another approach to improving seasonal forecasting is to dampen or shrink seasonal indices. Bunn and Vassilpoulos (1999) applied shrinkage seasonal indices (SSI) estimator - James-Stein estimator to shrink the ISI towards DGSI or WGSI in multi-item short-term forecasting. Their empirical investigation indicated that the ~~application offered the highest improvement in forecast performance~~ SSI methods made a uniform improvement on forecasting accuracy over ISI and were generally better than the grouping methods (Bunn & Vassilopoulos, 1999). Also, Miller and Williams (2003) attempted to improve the accuracy of the ISI method through the shrinkage methods. The shrinkage methods adjusted the ISI towards one ~~in a~~ for multiplicative ~~model or zero in an additive model~~ seasonality. Their findings revealed that shrinkage methods are generally more accurate than individual seasonal estimation and their performance depends on characteristics of the time series (Miller & Williams, 2003). ~~Furthermore,~~ Miller and Williams (2004) investigated the potential of the shrinkage methods for improving X-12-ARIMA and concluded that forecasting accuracy improved when seasonal damping was used in the seasonal adjustment (Miller & Williams, 2004). This investigation inspired discussions on the topic of shrinking seasonal factor in a special issue of the International Journal of Forecasting (Amstrong, 2004; Findley, Wills, & Monsell, 2004; Koehler, 2004; Miller & Williams, 2004; Ord, 2004). ~~Following the discussions,~~ Chen and Boylan (2008) undertook an empirical comparison between the ISI, the GSI and the SSI. They found that both grouping methods and shrinkage methods improve forecasting accuracy over the ISI method, particularly when the data history is short and the data are noisy.

However, no previous studies have examined theoretically how forecasting accuracy can be further improved by bringing the grouping and shrinkage approaches together. Therefore, this paper proposes generalised estimators that combine one of the two grouping approaches with a generalised shrinkage approach aiming at further improvement~~s~~ in forecasting performance. The generalised estimators are presented in Section 2. Since the shrinkage parameter plays a key role in the performance of the generalised estimator, Section 3 focuses on discussion of the optimal shrinkage parameter which minimises the Mean-Square-Error (MSE) of the ~~corresponding estimator~~forecast. Given the optimal parameters, we compare different estimators theoretically in Section 4. Section 5 contains an analysis of the effect of three factors on the MSE, namely the length of data history, variance of random components and the number of series. In section 6, a simulation experiment is designed to compare the performance of the estimators and to examine the factors that have an important effect on forecasting accuracy. The final section contains a summary of the paper and an outlook on future work.

2. Generalised estimators

2.1 Models and assumptions

A forecasting method is a procedure for computing forecasts from present and past values (Chatfield, 2001), while a model is an equation or set of equations representing the stochastic structure of the time series (Meade, 2000). This paper uses two models previously analysed by Chen and Boylan (2007, 2008):

Mixed model: $$Y_{ith} = \mu_i S_h + \varepsilon_{ith} \quad (1)$$

Additive model: $$Y_{ith} = \mu_i + S_h + \varepsilon_{ith} \quad (2)$$

where suffix i represents the SKU and $i = 1,...,m$ where m is the number of series; suffix t represents the year and $t = 1, 2..., r$ where r is the number of years' data history; suffix h represents the seasonal period and $h = 1,...,q$ where q is the length of the seasonal cycle; Y represents demand; μ_i represents the underlying mean for the i th SKU; S_h represents a seasonal index at seasonal period h; and ε_{ith} represents a random disturbance term for the i th SKU at the t th year and h th period.

The two models are stationary. We assume that there is no trend in the models in order to concentrate on the seasonal component alone. The underlying mean is assumed to be constant over time but different for different SKUs. Also, we assume that seasonality is fixed from year to year and is the same for all SKUs. The sum of seasonal indices in the additive model is zero, i.e. $\sum_{h=1}^{q} S_h = 0$ and the average of the seasonal indices in the mixed model is one, i.e. $\frac{1}{q}\sum_{h=1}^{q} S_h = 1$. The random disturbance term is assumed to be normally distributed with mean zero and constant variance σ_i^2. There are no auto-correlations within individual series and no cross-correlations at different time periods. There are only non-zero cross-correlations ρ_{ij} between ε_{ith} and ε_{jth} at the same time period (same cycle, same season).

2.2 Generalised estimators for the additive model

In order to achieve further improvements in forecasting accuracy, we propose to combine one of the two group seasonal indices methods (DGSI and WGSI) with a shrinkage seasonal indices method (SSI) to form generalised estimators for the additive and mixed seasonal models.

Since the DGSI and WGSI estimators for the additive model yield the same result, we use GSI instead of DGSI and WGSI ~~for the additive model~~. Expressions of ISI and GSI estimators for the additive model were given by Chen and Boylan (2007):

$$\hat{S}_{iH}^{\mathrm{ISI}} = \frac{1}{r}\sum_{t=1}^{r} Y_{itH} - \frac{1}{qr}\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{ith} \tag{3}$$

$$\hat{S}_{H}^{\mathrm{GSI}} = \frac{\sum_{j=1}^{m} \hat{S}_{jH}^{\mathrm{ISI}}}{m} = \frac{1}{mr}\sum_{j=1}^{m}\sum_{t=1}^{r} Y_{jtH} - \frac{1}{mqr}\sum_{j=1}^{m}\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{jth} \tag{4}$$

where $\hat{S}_{iH}^{\mathrm{ISI}}$ represents the estimated seasonal indices of the *i*th series at the *H*th season period by using the ISI estimator; $\hat{S}_{H}^{\mathrm{GSI}}$ represents the estimated seasonal indices at the *H*th season period by using the GSI estimator.

Moreover, a shrinkage seasonal index (SSI) for the additive model is formed by adding a shrinkage parameter to the ISI:

$$\hat{S}_{iH}^{\mathrm{SSI}} = \lambda_i \hat{S}_{iH}^{\mathrm{ISI}} = \lambda_i \left(\frac{1}{r}\sum_{t=1}^{r} Y_{itH} - \frac{1}{qr}\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{ith} \right) \tag{5}$$

where λ_i is the shrinkage parameter of the SSI for the additive model.

Based on the above three estimators, two generalised estimators which include shrinkage and grouping methods for the additive model, can be produced. They are Shrinkage Group Seasonal Indices (SGSI) estimator produced by shrinking first and then grouping and Group Shrinkage Seasonal Indices (GSSI) estimator produced by grouping first and then shrinking.

Shrinkage Group Seasonal Indices (SGSI)

$$\hat{S}_{H}^{\mathrm{SGSI}} = \frac{1}{m}\sum_{j=1}^{m} \lambda_j \hat{S}_{jH}^{\mathrm{ISI}} = \frac{1}{m}\sum_{j=1}^{m} \lambda_j \left(\frac{1}{r}\sum_{t=1}^{r} Y_{jtH} - \frac{1}{qr}\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{jth} \right) \tag{6}$$

where $\lambda_1, \ldots, \lambda_m$ are the shrinkage parameters and $\lambda_j \geq 0$ for $j = 1, \ldots, m$.

Group Shrinkage Seasonal Indices (GSSI)

$$\hat{S}_H^{\mathrm{GSSI}} = \lambda \hat{S}_H^{\mathrm{GSI}} = \lambda\left(\frac{1}{mr}\sum_{j=1}^{m}\sum_{t=1}^{r} Y_{jtH} - \frac{1}{mqr}\sum_{j=1}^{m}\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{jth}\right) \tag{7}$$

where λ is the shrinkage parameter of the GSSI and $\lambda \geq 0$.

ISI, SSI, GSI and GSSI are special cases of SGSI:

- ISI: $m = 1$ and $\lambda_i = 1$;
- SSI: $m = 1$ and $\lambda_i \neq 1$;
- GSI: $m \geq 2$ and $\lambda_1 = ... = \lambda_m = 1$.
- GSSI: $m \geq 2$ and $\lambda_1 = ... = \lambda_m = \lambda$

ISI is the same as SGSI when there is only one sample and it does not shrink. SSI is the same as SGSI when there is only one sample but it shrinks. GSI is the same as SGSI when there is more than one series and it does not shrink. GSSI is the same as SGSI when there is more than one series and it shrinks with the same shrinkage parameters for all series.

2.3 Generalised estimators for the mixed model

Expressions of ISI, DGSI, WGSI and SSI estimators for the mixed model are described as follows:

$$\hat{S}_{iH}^{\mathrm{ISI}} = \frac{\sum_{t=1}^{r} Y_{itH}}{r\mu_i} = \frac{q\sum_{t=1}^{r} Y_{itH}}{\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{ith}} \tag{8}$$

$$\hat{S}_{iH}^{\mathrm{DGSI}} = \frac{\sum_{j=1}^{m} \hat{S}_{iH}^{\mathrm{ISI}}}{m} = \frac{q}{m}\sum_{j=1}^{m}\left(\frac{\sum_{t=1}^{r} Y_{jtH}}{\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{jth}}\right) \tag{9}$$

$$\hat{S}_{iH}^{\mathrm{WGSI}} = \frac{q\sum_{j=1}^{m}\sum_{t=1}^{r} Y_{jtH}}{\sum_{j=1}^{m}\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{jth}} \tag{10}$$

$$\hat{S}_{iH}^{\mathrm{SSI}} = \lambda_i \hat{S}_{iH}^{\mathrm{ISI}} = \lambda_i \frac{q\sum_{t=1}^{r} Y_{itH}}{\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{ith}} \tag{11}$$

where λ_i is the shrinkage parameter of the SSI for the mixed model.

Similar to the generalised estimators used for the additive model, three generalised estimators for the mixed seasonal model can be produced by shrinking first and then grouping or by grouping first and then shrinking. Since DGSI is different from WGSI ~~in the mixed model~~, two generalised estimators are formed based on DGSI and one based on WGSI. They are Shrinkage Dalhart Group Seasonal Indices (SDGSI) estimator produced by shrinking first and then grouping by using DGSI, Dalhart Group Shrinkage Seasonal Indices (DGSSI) estimator produced by grouping by using DGSI first and then shrinking, and Withycombe Group Shrinkage Seasonal Indices (WGSSI) estimator produced by grouping by using WGSI first and then shrinking. Here, it is worth noting that only the WGSSI estimator can be produced by grouping by using WGSI first and then shrinking. It is impossible to obtain an estimator by shrinking first then grouping by using WGSI. This is because WGSI is calculated by aggregating all series first before working out the seasonal indices. Therefore, shrinkage can only apply after grouping.

Shrinkage Dalhart Group Seasonal Indices (SDGSI)

$$\hat{S}_H^{\text{SDGSI}} = \frac{1}{m}\sum_{j=1}^{m}\lambda_j \hat{S}_{jH}^{\text{ISI}} = \frac{1}{m}\sum_{j=1}^{m}\lambda_j\left(\frac{q\sum_{t=1}^{r}Y_{jtH}}{\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{jth}}\right) \tag{12}$$

where $\lambda_1,\ldots,\lambda_m$ are the shrinkage parameters and $\lambda_i \geq 0$ for $i=1,\ldots,m$.

Dalhart Group Shrinkage Seasonal Indices (DGSSI)

$$\hat{S}_H^{\text{DGSSI}} = \lambda \hat{S}_H^{\text{DGSI}} = \lambda\frac{\sum_{j=1}^{m} ISI_{jH}}{m} = \lambda\frac{q}{m}\sum_{j=1}^{m}\left(\frac{\sum_{t=1}^{r}Y_{jtH}}{\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{jth}}\right) \tag{13}$$

where λ is the shrinkage parameter.

Withycombe Group Shrinkage Seasonal Indices (WGSSI)

$$\hat{S}_H^{\text{WGSSI}} = \lambda \hat{S}_H^{\text{WGSI}} = \lambda\frac{q\sum_{j=1}^{m}\sum_{t=1}^{r}Y_{jtH}}{\sum_{j=1}^{m}\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{jth}} \tag{14}$$

where λ is the shrinkage parameter.

As before, we can see that ISI, SSI, DGSI and DGSSI are special cases of SDGSI. The interpretation of the estimators is very similar to the additive model.

3. Optimal shrinkage parameters for generalised estimators

The shrinkage parameter plays an important role in the forecasting performance of the corresponding generalised (or shrinkage only) estimator. Therefore, optimal shrinkage parameters which minimise the MSE of the generalised (or shrinkage only) estimators are derived in this section.

Since ISI, SSI, GSI and GSSI are all special cases of SGSI in the additive model, the following only takes the SGSI as an example to show how the optimal shrinkage parameter is calculated. The optimal shrinkage parameters for other shrinkage only or generalised estimators can be obtained in a similar way. The MSE calculation of the shrinkage only or generalised estimators in the additive model and mixed model are described in Appendi~~x~~ces A and ~~Appendix~~ B respectively. The calculation of the optimal shrinkage parameters are presented in Appendi~~x~~ces C and ~~Appendix~~ D respectively.

The MSE of the SGSI can be calculated as (see Appendix A for details):

$$\mathrm{MSE}^{\mathrm{SGSI}} = \left(1+\frac{1}{qr}\right)\sigma_i^2 + \frac{(q-1)}{m^2 qr}\left[\sum_{j=1}^{m}\lambda_j^2\sigma_j^2 + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\lambda_j\lambda_l\rho_{jl}\sigma_i\sigma_j\right] + \left[\left(1-\frac{1}{m}\sum_{j=1}^{m}\lambda_j\right)S_H\right]^2 \tag{15}$$

where λ_j ($j=1,\ldots,m$) is a shrinkage parameter for the j th series and $\lambda_j \geq 0$.

A set of optimal shrinkage parameters can be obtained by differentiating the MSE of the SGSI with respect to any shrinkage parameter λ_k ($k=1,\ldots,m$):

$$\lambda_k = \frac{\left(m-\sum_{j=1,j\neq k}^{m}\lambda_j\right)S_H^2 - \left(\frac{q-1}{qr}\right)\sum_{j=1,j\neq k}^{m}\lambda_j\rho_{jk}\sigma_j\sigma_k}{S_H^2+\left(\frac{q-1}{qr}\right)\sigma_k^2} \tag{16}$$

Equation (16) is a set of m simultaneous equations in m unknowns. The shrinkage parameters λ_1, λ_2,..., λ_m can be obtained by solving the simultaneous equations.

4. Comparisons of the estimators

Chen and Boylan (2007) carried out a comparison between ~~individual seasonal indices method (ISI) and group seasonal indices method (GSI)~~ ISI and GSI. In this section, we will extend the comparison to other estimators which include the shrinkage-only method (SSI), the grouping and then shrinking method (GSSI) and the shrinking and then grouping method (SGSI). Although the following will only take the estimators for the additive model as an example, similar conclusions can be reached for the mixed model. In addition, the additive model does not differentiate ~~the~~ DGSSI from ~~the~~ WGSSI, but they are different in the mixed model. Thus, a comparison between ~~the~~ DGSSI and ~~the~~ WGSSI for the mixed model is added in the last subsection.

4.1 ISI and SSI

This subsection will compare ~~the~~ SSI to ~~the~~ ISI, aiming to examine if the performance of the original individual estimator can be improved through shrinking.

The MSE expressions of ISI for the additive model was given in (Chen & Boylan, 2007).

$$\mathrm{MSE}^{\mathrm{ISI}} = \left(1+\frac{1}{r}\right)\sigma_i^2 \tag{17}$$

The MSE of the optimal SSI for the additive model can be calculated by inserting the corresponding optimal shrinkage parameter $\lambda_i = \dfrac{S_H^2}{S_H^2 + \left(\dfrac{q-1}{qr}\right)\sigma_i^2}$ shown in Appendix C in its MSE expressions shown in Appendix A.

$$\mathrm{MSE}_{\min}^{\mathrm{SSI}} = \left(1+\frac{1}{qr}\right)\sigma_i^2 + \frac{\left(\dfrac{q-1}{qr}\right)S_H^2\sigma_i^2}{S_H^2 + \left(\dfrac{q-1}{qr}\right)\sigma_i^2} \tag{18}$$

The difference in MSE between ~~the~~ ISI and the optimal SSI can thus be calculated as:

$$\mathrm{MSE}^{\mathrm{ISI}}-\mathrm{MSE}_{\min}^{\mathrm{SSI}}=\frac{\left(\frac{q-1}{qr}\right)^{2}\sigma_i^4}{S_H^2+\left(\frac{q-1}{qr}\right)\sigma_i^2}>0 \tag{19}$$

Equation (19) shows that in the additive model ~~the~~ SSI is less than ~~the~~ ISI in MSE when ~~the~~ SSI is optimal, i.e., when its MSE is minimum. Actually, the optimal estimator produced by shrinking is better than the basic individual estimator whether in the additive model or in the mixed model.

4.2 GSI and GSSI

Similar to the comparison between ~~the~~ ISI and ~~the~~ SSI, this subsection devotes to examining if the performance of a group estimator will be improved through further shrinking.

Also, the MSE of the optimal GSSI can be calculated by inserting the corresponding optimal shrinkage parameter:

$$\mathrm{MSE}_{\min}^{\mathrm{GSSI}}=\left(1+\frac{1}{qr}\right)\sigma_i^2+\frac{(q-1)}{qr}\frac{S_H^2}{S_H^2+\left(\frac{q-1}{qr}\right)\frac{\sigma_A^2}{m^2}}\frac{\sigma_A^2}{m^2} \tag{20}$$

where $\sigma_A^2=\sum_{j=1}^{m}\sigma_j^2+2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\rho_{jl}\sigma_j\sigma_l$ is the variance of the deseasonalised aggregate demand.

The difference in MSE between ~~the~~ GSI and the optimal GSSI is:

$$\mathrm{MSE}^{\mathrm{GSI}}-\mathrm{MSE}_{\min}^{\mathrm{GSSI}}=\left(\frac{q-1}{qr}\right)\left(\frac{\frac{(q-1)}{qr}\frac{\sigma_A^2}{m^2}}{S_H^2+\frac{(q-1)}{qr}\frac{\sigma_A^2}{m^2}}\right)\frac{\sigma_A^2}{m^2}>0 \tag{21}$$

Equation (21) shows that ~~the~~ GSSI is less than ~~the~~ GSI in MSE when ~~the~~ GSSI is optimal, i.e., the group shrinkage method is better than group-only method. With the result of the previous subsection, it is concluded that the performance of estimators (whether individual one or group-only one) can be improved through further shrinking for both models. The empirical results in Bunn and Vassilopoulos (1999) are consistent with this finding. Moreover, the difference between GSI and GSSI will approach zero when the number of series (m) goes infinity, i.e., the group-only estimator and the group shrinkage estimator will be the same in terms of forecasting accuracy when the number of ~~serious~~ series are large enough.

4.3 SSI and GSI

This subsection will compare the shrinkage-only method to group-only method, aiming to discover which method is better or under what conditions a method is better than the other.

The difference in MSE between the optimal SSI and ~~the~~ GSI can be calculated as:

$$\mathrm{MSE}_{\min}^{\mathrm{SSI}} - \mathrm{MSE}^{\mathrm{GSI}} = \left(\frac{q-1}{qr}\right)\left(\frac{S_H^2}{S_H^2 + \frac{(q-1)}{qr}\sigma_i^2}\sigma_i^2 - \frac{\sigma_A^2}{m^2}\right) = \left(\frac{q-1}{qr}\right)\left(\lambda_i\sigma_i^2 - \frac{\sigma_A^2}{m^2}\right) \quad (22)$$

where $\lambda_i = \dfrac{S_H^2}{S_H^2 + \left(\frac{q-1}{qr}\right)\sigma_i^2}$ is the optimal shrinkage parameter.

Thus, $\mathrm{MSE}^{\mathrm{SSI}} > \mathrm{MSE}^{\mathrm{GSI}}$ if and only if

$$\lambda_i\sigma_i^2 > \frac{\sigma_A^2}{m^2} \quad (23)$$

That means that in the additive model ~~the~~ GSI is better than the optimal SSI if and only if the 'shrunk' individual series' variance ($\lambda_i\sigma_i^2$) is greater than the 'average' variance of the group ($\frac{\sigma_A^2}{m^2}$), otherwise, the optimal SSI is better.

Similarly, in the mixed model ~~the~~ DGSI is better than the optimal SSI if and only if $\lambda_i\left(\frac{\sigma_i^2}{\mu_i^2}\right) > \frac{\sigma_{\mu A}^2}{m^2}$ (where $\sigma_{\mu A}^2 = \sum_{j=1}^{m}\frac{\sigma_j^2}{\mu_j^2} + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\left(\frac{1}{\mu_j\mu_l}\right)\rho_{jl}\sigma_j\sigma_l$) and ~~the~~ WGSI is better if $\lambda_i\left(\frac{\sigma_i^2}{\mu_i^2}\right) > \frac{\sigma_A^2}{\mu_A^2}$ (where $\sigma_A^2 = \sum_{i=1}^{m}\sigma_i^2 + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\rho_{jl}\sigma_j\sigma_l$ and $\mu_A = \mu_1 + \mu_2 + \ldots + \mu_m$). Conceptually, the three conditions are the same. The difference lies in the expression of noisiness of individual series and the 'average' of the group. In the additive model, noisiness is measured by variance (σ_i^2). In the mixed model, it is measured by the square of coefficient of variation ($\frac{\sigma_i^2}{\mu_i^2}$). The GSI and the DGSI use the average of individual noisiness ($\frac{\sigma_A^2}{m^2}$ or $\frac{\sigma_{\mu A}^2}{m^2}$), while the WGSI uses the noisiness of aggregate series ($\frac{\sigma_A^2}{\mu_A^2}$). These findings agree with Chen and Boylan (2008) that the grouping and shrinkage approaches are competitive with each other. Here the theoretical conditions are established to understand when one is better than the other.

4.4 GSSI and SGSI

Having known that both grouping and shrinking have effects on the performance of the estimator, we now wish to discover which is better way to combine them together, shrinking first and then grouping or grouping first and then shrinking. Since the optimal shrinkage parameters of SGSI have to be calculated by solving simultaneous equations (16), an MSE expression of the optimal SGSI cannot be obtained directly in a similar way to the optimal SSI in equation (18) or the optimal GSSI in equation (20). Therefore, a direct comparison between the optimal SGSI and other estimators is intractable. Here, a general discussion on the SGSI and the GSSI is given, which can be applied to the comparison between any estimators.

The difference between ~~the~~ GSSI and ~~the~~ SGSI can be calculated by using their MSE expressions presented in Appendix A:

$$\mathrm{MSE}^{\mathrm{GSSI}} - \mathrm{MSE}^{\mathrm{SGSI}} = \frac{(q-1)}{qr}\left(\frac{\lambda^2\sigma_A^2}{m^2} - \frac{\sigma_{\lambda A}^2}{m^2}\right) - \left[\left(S_H - \left(\frac{1}{m}\sum_{j=1}^{m}\lambda_j\right)S_H\right)^2 - \left(S_H - \lambda S_H\right)^2\right] \tag{24}$$

where λ_j ($j = 1,\ldots,m$) is the shrinkage parameters of the SGSI and λ is the shrinkage parameter of the GSSI; $\sigma_{\lambda A}^2 = \sum_{j=1}^{m}\lambda_j^2\sigma_j^2 + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\lambda_j\lambda_l\rho_{jl}\sigma_i\sigma_j$ is the aggregate of shrunk variances and $\sigma_A^2 = \sum_{j=1}^{m}\sigma_j^2 + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\rho_{jl}\sigma_j\sigma_l$ is the aggregate variance.

Thus, $\mathrm{MSE}^{\mathrm{GSSI}} > \mathrm{MSE}^{\mathrm{SGSI}}$ if and only if

$$\frac{(q-1)}{qr}\left(\frac{\lambda^2\sigma_A^2}{m^2}-\frac{\sigma_{\lambda A}^2}{m^2}\right)>\left[\left(S_H-\left(\frac{1}{m}\sum_{j=1}^{m}\lambda_j\right)S_H\right)^2-(S_H-\lambda S_H)^2\right] \tag{25}$$

From equation (25), it is found that $\frac{\lambda^2\sigma_A^2}{m^2}$ is the 'shrunk average of aggregate variance' with ~~the~~ GSSI, $\frac{\sigma_{\lambda A}^2}{m^2}$ is the 'average of shrunk aggregate variances' with ~~the~~ SGSI, $(S_H-\lambda S_H)$ is the bias of shrunk seasonal index with ~~the~~ GSSI and $\left(S_H-\left(\frac{1}{m}\sum_{j=1}^{m}\lambda_j\right)S_H\right)$ is the bias of average shrunk seasonal index with ~~the~~ SGSI. The comparison of the two estimators is actually a comparison between the difference of 'two generalised average variances' and the difference of 'the squared bias of the two generalised seasonal indices'. When the difference in the 'variances' multiplied by a coefficient $\frac{(q-1)}{qr}$ is greater than the difference of 'the squared bias of the two generalised seasonal indices', shrinking before grouping is better than grouping first, i.e., applying multiple shrinkage parameters is better than a universal one.

4.5 DGSSI and WGSSI

Although DGSI and WGSI or DGSSI and WGSSI are the same in the additive model, they are different in the mixed model. Thus, a comparison between ~~the~~ DGSSI and WGSSI for the mixed model ~~are~~ is added here.

The difference between the optimal DGSSI and the optimal WGSSI is:

$$\mathrm{MSE}_{\min}^{\mathrm{DGSSI}}-\mathrm{MSE}_{\min}^{\mathrm{WGSSI}}=\frac{\mu_i^2}{r}\left[\frac{S_H^4\left(\frac{\sigma_{\mu A}^2}{m^2}-\frac{\sigma_A^2}{\mu_A^2}\right)}{\left(S_H^2+\frac{\sigma_{\mu A}^2}{rm^2}\right)\left(S_H^2+\frac{\sigma_A^2}{r\mu_A^2}\right)}\right] \tag{26}$$

Thus, $\mathrm{MSE}_{\min}^{\mathrm{DGSSI}}>\mathrm{MSE}_{\min}^{\mathrm{WGSSI}}$ if and only if

$$\frac{\sigma_{\mu A}^2}{m^2}>\frac{\sigma_A^2}{\mu_A^2} \tag{27}$$

where $\sigma_{\mu A}^2 = \sum_{j=1}^{m} \frac{\sigma_j^2}{\mu_j^2} + 2\sum_{j=1}^{m-1} \sum_{l=j+1}^{m} \left(\frac{1}{\mu_j \mu_l} \right) \rho_{jl} \sigma_j \sigma_l$ is the aggregate of squared coefficient of variation.

The condition (equation (27)) is the same as the rule of comparing ~~the~~ DGSI and WGSI which was derived by Chen and Boylan (2007). That means that the relationship in forecasting performance between ~~the~~ DGSSI and ~~the~~ WGSSI is similar to ~~the~~ DGSI and ~~the~~ WGSI.

In summary, the optimal shrinkage estimator is better than individual seasonal indices method and the group shrinkage method is better than the corresponding group only method. The comparison between the group method and shrinkage method, or between the group shrinkage method and shrinkage group method, or between two group shrinkage methods depends on conditions that have been specified in this section ~~of the paper~~.

5. Effect of three factors on MSE

The differences between different estimators were discussed in the ~~last~~ previous section. This section will be devoted to an evaluation of the effect of three factors which are included in the MSE expressions. They are: the length of data history used for estimation of seasonal indices, the variances of random components and the number of series. From the evaluation, the similarities of different estimators can be detected. The following will take ~~the~~ SGSI as an example to discuss the effect of the three factors on MSE. There are similar effects on other estimators for both the additive model and the mixed model.

5.1 Length of data history

Assume that all the series are independent, i.e., $\rho_{jl} = 0$, differentiate the MSE of SGSI calculated in equation (15) with respect to the length of data history r:

$$\frac{\partial\left(\mathrm{MSE}^{\mathrm{SGSI}}\right)}{\partial r} = \left(\frac{-1}{qr^2}\right)\sigma_i^2 + \frac{-(q-1)}{m^2 qr^2}\left(\sum_{j=1}^{m} \lambda_j^2 \sigma_j^2\right) < 0 \tag{28}$$

The differentiation of the MSE is less than zero, which means that the MSE of SGSI will decrease as the length of data history increases if all the series are independent. The conclusion can be applied to all estimators: the longer the data history, the higher the estimation accuracy.

5.2 Variances

Assume that all the series are independent, i.e., $\rho_{jl} = 0$, differentiate the MSE of SGSI calculated in equation (15) with respect to the variances σ_k^2 ($k = 1,...,m$):

$$\frac{\partial\left(\mathrm{MSE}^{\mathrm{SGSI}}\right)}{\partial\sigma_k^2} = \begin{cases} \left(1+\dfrac{1}{qr}\right)+\dfrac{(q-1)}{m^2qr}\left[\lambda_i^2+\sum_{j\neq i}^{m}\lambda_j^2\sigma_j^2\right]>0 & (k=i) \\ \dfrac{(q-1)}{m^2qr}\left[\lambda_k^2+\sum_{j\neq i}^{m}\lambda_j^2\sigma_j^2\right]>0 & (k\neq i) \end{cases} \tag{29}$$

The differentiation of the MSE is greater than zero whether differentiating with respect to the variance of the estimated item ($k = i$) or not ($k \neq i$), which shows that the MSE of SGSI will increase as the variance increases if all the series are independent. The conclusion can be applied to all estimators: the higher noisiness will lead to the lower estimation accuracy.

5.3 Number of series

Assume that all the series are independent, i.e., $\rho_{jl} = 0$, and the number of series m goes to infinity, MSE of the SGSI can be calculated by inserting the optimal shrinkage parameters shown in equation (16):

$$\begin{aligned} &\lim_{m\to\infty} \mathrm{MSE}^{\mathrm{SGSI}} \\ &= \lim_{m\to\infty}\left[\left(1+\frac{1}{qr}\right)\sigma_i^2+\frac{(q-1)}{m^2qr}\left(\sum_{j=1}^{m}\lambda_j^2\sigma_j^2\right)+\left[\left(1-\frac{1}{m}\sum_{j=1}^{m}\lambda_j\right)S_H\right]^2\right] \\ &= \left(1+\frac{1}{qr}\right)\sigma_i^2 \end{aligned} \tag{30}$$

That means that if all the series are independent, i.e., $\rho_{jl} = 0$, then the MSE of SGSI will approach a constant $\left(1+\dfrac{1}{qr}\right)\sigma_i^2$ when m goes to infinity. Also, the MSEs of GSI and GSSI for the additive model will approach a constant $\left(1+\frac{1}{qr}\right)\sigma_i^2$ under the same condition. A similar calculation can be performed on the MSEs of all grouping-related methods (DGSI, WGSI, SDGSI, DGSSI and WGSSI) for the mixed model. It is found that all the MSEs will approach a constant σ_i^2 when m goes to infinity.

Furthermore, by comparing equation (18) with equation (30), it is found that the MSE of the optimal SSI is greater than all the grouping-related methods (GSI, SGSI and GSSI) for the additive model when all the series are independent and the number of series goes to infinity. Also, the optimal SSI for the mixed model is worse than all grouping-related methods (DGSI, WGSI, SDGSI, DGSSI and WGSSI) in the same situation.

6. Simulation experiment

6.1 Design of simulation experiment

The previous two sections presented the theoretical evaluations of different estimators. However, the shrinkage parameters of the SGSI are not easily determined theoretically due to the difficulty in solving simultaneous equations, which prevents the comparison between the estimator and others. In this study, a simulation experiment is designed to compare the performance of the estimators and examine the crucial factors which affect forecasting accuracy. The examination focuses on the effect of three factors mentioned in Section 5. Here, the optimal shrinkage parameters which minimise the MSE of corresponding generalised estimators are applied to the corresponding estimators. Since this study is a further exploration of previous research, the reasons for the choice of the parameters in the experiment can be found in (Chen & Boylan, 2007). These parameters include:

- Seasonal profiles

Three seasonal profiles are used for quarterly seasonality forecasting in each of models, which include: WS (weak seasonality), LLLH (low values for three quarters and high for one) and LHLH (low values for two quarters and high for the other two). The details of parameters are given in Table 2 and Table 3 of Chen & Boylan (2007).

- Number of series

Six groups of data are used with the following number of series: 4, 8, 16, 32, 64, and 128. In this experiment, the number of series made little difference on MSE when the number of series was increased to 128. Therefore, ~~a further examination of higher volume of series were~~a further examination of higher volume of series was not conducted.

- Underlying mean

Underlying mean value is generated by assuming its distribution is lognormal distribution with mean 4 and standard ratio 0.69.

- Variance

Variance of the random noise is generated by using universal power law of the form $\sigma_i^2 = \alpha\mu_i^\beta$ where σ_i^2 is the variance, μ_i is the underlying mean and α and β are constants (Brown, 1959). Here we choose α to be 0.5, β to be 1.2, 1.4, 1.6 or 1.8.

- Cross correlation coefficient

The correlation coefficients are assumed to be zero ($\rho_{ij} = 0$) in this experiment, i.e., there is no cross correlation between different series at the same time period. In the theoretical analysis in Section 5, the correlation coefficients are assumed to be zero in order to focus on the discussion of effect of three main factors. Since the experiment aims at a further examination of the theoretical research, it is designed to follow the same condition.

- Data history

3, 4 or 5 years' data are generated and the last year's observations are used as the holdout sample. So the data history used for estimation are 2, 3 or 4 years.

- Replications

For each parameter setting, 1000 replication were run to reduce sampling errors.

6.2 Experiment on length of data history

In order to present the effect of data history, the MSE results shown in Fig. 1 were obtained based on the average of data from three kinds of seasonal patterns (WS, LLLH and LHLH), 4 kinds of variances ($\beta = 1.2, 1.4, 1.6, 1.8$) and 6 groups of series ($m = 4, 8, 16, 32, 64, 128$). It was found that the performance of all estimators was closely related to the length of data history. The MSE decreased as the length of data history increased from 2 years to 4 years. The effect was obvious especially when using ISI and SSI. However, the MSE of grouping-related estimators decreased less. ~~This effect was caused by the grouping methods involved in the grouping related estimators.~~ The grouping methods helped the noisier data to borrow strength from less noisy data in the group, which reduced the sensitivity of data to the length of data history.

Although the effects of data history on the estimation accuracy were similar for all estimators, the difference between different estimators apparently existed. With the same data history, the ISI delivered the worst forecasting accuracy. The SSI is better than the ISI but worse than all grouping-related methods. Since the MSE results of different grouping-related methods are very close, it is difficult to detect their differences by visually examining the two figures. However, the differences can be found in Table 1: the group shrinkage method (DGSSI and WGSSI) and shrinkage group method (SDGSI) are better than the group-only method. Moreover, the SDGSI is ~~a bit~~ slightly better than DGSSI, and WGSSI is ~~a bit~~ slightly better than DGSSI. Another notable finding from the two figures is that the difference in MSE between different estimators decreased as the length of data increased, as expected from the theoretical analysis.

6.3 Experiment on variances

The MSE results regarding the effect of variance (See Fig. 2) were obtained based on the average of data from three kinds of seasonal patterns (WS, LLLH and LHLH), 3 kinds of history data ($r = 2, 3, 4$) and 6 groups of series ($m = 4, 8, 16, 32, 64, 128$). Two figures showed the effect of variances which were generated by using universal power law ~~of the form~~ $\sigma_i^2 = 0.5\mu_i^\beta$. ~~It is found the variances caused significant increase in MSEs of all estimators.~~ As the β constant in the ~~universal~~ power law increased from 1.2 to 1.8, the MSEs increased by around 15 times or more for all methods in both mixed and additive models.

The difference between different estimators was evident, especially when the variance was big, for example, when $\beta = 1.8$ (See Table 2). The results are very similar to the ~~last~~ previous subsection: the ISI is the worst; the SSI is better than the ISI but worse than group method; the group shrinkage method is better than the corresponding group method but worse than shrinkage group method. However, the comparison between two different group methods (DGSI and WGSI) or group shrinkage methods (DGSSI and WGSSI) in the mixed model depends greatly on the variance. The WGSI is better than the DGSI when the variance is small, for example, when $\beta = 1.2$, while the DGSI is better than the WGSI when the variance is big, for example, when $\beta = 1.8$. These are consistent with the findings in Chen and Boylan (2007).

6.4 Experiment on number of series

The MSE results regarding number of series were obtained based on the average of data from three kinds of seasonal patterns (WS, LLLH and LHLH), 4 kinds of variances ($\beta = 1.2, 1.4, 1.6, 1.8$) and 3 kinds of history data ($r = 2, 3, 4$). As shown in Fig. 3, the number of series had no great effect on MSE. However, a small change happened on the estimators which involve grouping methods. For example, MSE of the GSI for the additive model reduced from 0.8515 to 0.8018 as the number of series increased from 4 to 128 (see Table 3).. Although the change is small, it showed a benefit from the number of series. The more the number of series were, the better these grouping-related methods performed. Moreover, the MSE ratios of the grouping-related methods to the ISI approached ~~to~~ 0.802 in the additive model or 0.818 in the mixed model when the number of series was increased to 128. That means that MSEs of all grouping-related methods are nearly the same when the number of series are big enough. The theoretical analyses ~~in last section~~ showed that MSE of all grouping-related methods will approach to a constant $\left(1+\frac{1}{qr}\right)\sigma_i^2$ in the additive model or σ_i^2 in the mixed model when m goes to infinity and $\rho_{jl} = 0$. Hence, the experimental results proved theoretical predictions.

Another notable fact is that MSE of the ISI is greater than the SSI and the SSI is greater than the grouping-related methods whatever the number of series is. The results are also consistent with the theory.

7. Conclusions

This paper presented several generalised estimators used to estimate seasonal indices for both additive and mixed models. The estimators were formed by combining one of two group seasonal indices methods - Dalhart Group Seasonal Indices estimator (DGSI) or Withycombe Group Seasonal Indices estimator (WGSI) with a shrinkage approach in different ways. Including the basic ISI, GSI and SSI methods, a total of seven estimators are obtained in the mixed model and five estimators in the additive model. With the help of the optimal shrinkage parameters, the emphasis of this paper is placed on theoretical comparison of all estimators. The comparisons are then enhanced through a further discussion about the effect of three factors on forecasting accuracy. Finally, a simulation experiment is designed to support these theoretical analyses. It is found that the simulation results are consistent with the theoretical analyses.

Through the theoretical and ~~empirical~~ simulation comparisons, the following conclusions are made:

- The optimal SSI is better than ~~the~~ ISI and the optimal group shrinkage method (DGSSI or WGSSI) is better than the corresponding group-only method (DGSI or WGSI).

- The group-only method is better than the optimal shrinkage method if and only if the 'shrunk variance' with the optimal shrinkage method is greater the 'average' variance with the group-only method. ~~The~~ GSI is better than the optimal SSI only if $\lambda_i \sigma_i^2 > \frac{\sigma_A^2}{m^2}$ in the additive model. ~~The~~ DGSI is better than the optimal SSI only if $\lambda_i \left(\frac{\sigma_i^2}{\mu_i^2} \right) > \frac{\sigma_{\mu A}^2}{m^2}$ and ~~the~~ WGSI is better only if $\lambda_i \left(\frac{\sigma_i^2}{\mu_i^2} \right) > \frac{\sigma_A^2}{\mu_A^2}$ in the mixed model.

- The shrinkage group method (SGSI for the additive model or SDGSI for the mixed model) is better than the optimal shrinkage-only method (SSI) when all the series are independent and the number of series goes to infinity.
- The optimal SSI is worse than all grouping-related methods (GSI, SGSI and GSSI for additive model; DGSI, WGSI, SDGSI, DGSSI and WGSSI for the mixed model) when m goes to infinity.
- The theoretical comparisons between the optimal shrinkage group method and the optimal group shrinkage method is not easy to be articulated. Only general conclusions can be made: which estimator is better depends on the difference of 'generalised variances' and the difference of 'squared bias of seasonal indices' between two compared estimators. However, the simulation experiment contributed to a fact that the optimal shrinkage group method ~~is better than~~ made a small improvement over the optimal group shrinkage method.
- Two different group methods (DGSI and WGSI) or group shrinkage methods (DGSSI and WGSSI) in the mixed model depend greatly on the 'variance'. The WGSI is better than the DGSI or the WGSSI is better than the DGSSI when the variance is small, and vice versa.
- The length of data history has an effect on the relative forecasting performance of the estimators. The longer data history, the higher the forecasting accuracy. Moreover, the difference in MSE between different estimators decreased as the length of data increased.
- ~~The effect of variance is obvious for all the estimators. The greater variance caused an obvious reduction in forecast accuracy.~~
- The number of series has no significant effect on forecasting performance. Especially when the number of series is large enough, the MSE of all grouping-related methods approached ~~to~~ a constant if all the series are independent.

For future work, we intend to establish the link between these generalised estimators and James-Stein type of shrinkage estimator. The latter was applied by Bunn and Vassilpoulos (1999) and Miller and Williams (2003), which showed promising results. Once such theoretical understanding is achieved, we will test these estimators on large scale real data.

Acknowledgements

We are deeply grateful to the editor for all his helpful advice and great encouragement. This research ~~was supported~~ is funded by the UK Engineering and Physical Science Research Council (EPSRC) ~~funds~~ (EP/G003858/1).

References

Amstrong, J. S. (2004). Damped Seasonality Factors: Introduction. *International Journal of Forecasting, 20*(4), 525-527.

Brown, R. G. (1959). *Statistical forecasting for inventory control*. New York: McGraw-Hill.

Bunn, D. W., & Vassilopoulos, A. I. (1993). Using Group Seasonal Indices in Multi-item Short-term Forecasting. *International Journal of Forecasting, 9*, 517-526.

Bunn, D. W., & Vassilopoulos, A. I. (1999). Comparison of seasonal estimation methods in multi-item short-term forecasting. *International Journal of Forecasting, 15*(4), 431-443.

Chatfield, C. (2001). *Time-series Forecasting*. Boca Raton: Chapman & Hall/CRC Press.

Chen, H., & Boylan, J. E. (2007). Use of Individual and Group Seasonal Indices in Subaggregate Demand Forecasting. *Journal of the Operational Research Society, 58*, 1660-1671.

Chen, H., & Boylan, J. E. (2008). Empirical Evidence on Individual, Group and Shrinkage Seasonal Indices. *International Journal of Forecasting, 24*, 525-534.

Dalhart, G. (1974). *Class Seasonality - A New Approach.* Paper presented at the American Production and Inventory Control Society Conference Proceedings.

Findley, D. F., Wills, K. C., & Monsell, B. C. (2004). Seasonal adjustment perspectives on "Damping seasonal factors: shrinkage estimators for the X-12-ARIMA program". *International Journal of Forecasting, 20*(4), 551-556.

Koehler, A. B. (2004). Comments on Damped Seasonal Factors and Decisions by Potential Users. *International Journal of Forecasting, 20*(4), 565-566.

Meade, N. (2000). Evidence for the selection of forecasting methods. *Journal of Forecasting, 19*(6), 515 - 535.

Miller, D. M., & Williams, D. (2003). Shrinkage estimators of time series seasonal factors and their effect on forecasting accuracy. *International Journal of Forecasting*, *19*, 669-684.

Miller, D. M., & Williams, D. (2004). Damping seasonal factors: Shrinkage estimators for the X-12-ARIMA program. *International Journal of Forecasting*, *20*(4), 529-549.

Ord, J. K. (2004). Shrinking: When and How? *International Journal of Forecasting*, *20*(4), 567-568.

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

Appendix A. MSEs of estimators - SGSI, SSI and GSSI in the additive model

MSE of SGSI

$Y_{ith} = \mu_i + S_h + \varepsilon_{ith}$ and

$$\hat{S}_H^{\text{SGSI}} = \frac{1}{m}\sum_{j=1}^{m}\lambda_j \hat{S}_{jH}^{\text{ISI}} = \frac{1}{m}\sum_{j=1}^{m}\lambda_j\left(\frac{1}{r}\sum_{t=1}^{r}Y_{jtH} - \frac{1}{qr}\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{jth}\right)$$

The forecast for the i th series, H th season, in year $r+1$ using the SGSI is:

$$\hat{Y}_{i(r+1)H}^{\text{SGSI}} = \hat{\mu}_i + \hat{S}_H^{\text{SGSI}}$$

$$= \frac{1}{qr}\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{ith} + \frac{1}{m}\sum_{j=1}^{m}\lambda_j\left(\frac{1}{r}\sum_{t=1}^{r}Y_{jtH} - \frac{1}{qr}\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{jth}\right)$$

Thus, MSE of the SGSI can be calculated:

$$\text{MSE}^{\text{SGSI}} = \text{E}\left(Y_{i(r+1)H} - \hat{Y}_{i(r+1)H}^{\text{SGSI}}\right)^2$$

$$= \text{E}\left[Y_{i(r+1)H} - \frac{1}{qr}\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{ith} - \frac{1}{m}\sum_{j=1}^{m}\lambda_j\left(\frac{1}{r}\sum_{t=1}^{r}Y_{jtH} - \frac{1}{qr}\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{jth}\right)\right]^2$$

$$= \left(1+\frac{1}{qr}\right)\sigma_i^2 + \frac{(q-1)}{m^2 qr}\left[\sum_{j=1}^{m}\lambda_j^2\sigma_j^2 + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\lambda_j\lambda_l\rho_{jl}\sigma_i\sigma_j\right] + \left[\left(1-\frac{1}{m}\sum_{j=1}^{m}\lambda_j\right)S_H\right]^2$$

MSE of SSI

Since the SSI for the additive model is a special case of the SGSI when $m=1$ and $\lambda_i \neq 1$, the MSE is:

$$\text{MSE}^{\text{SSI}} = \left(1+\frac{1}{qr}\right)\sigma_i^2 + \left(\frac{q-1}{qr}\right)\left(\lambda_i^2\sigma_i^2\right) + \left(1-\lambda_i\right)^2 S_H^2$$

MSE of GSSI

Since the GSSI for the additive model is a special case of the SGSI when $m \geq 2$ and $\lambda_1 = \ldots = \lambda_m = \lambda$, the MSE is:

$$\mathrm{MSE}^{\mathrm{GSSI}} = \left(1 + \frac{1}{qr}\right)\sigma_i^2 + \frac{(q-1)\lambda^2}{m^2 qr}\left[\sum_{j=1}^{m}\sigma_j^2 + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\rho_{jl}\sigma_i\sigma_j\right] + \left[(1-\lambda)S_H\right]^2$$

Appendix B. MSEs of estimators - SDGSI, SSI, DGSSI and WGSSI in the mixed model

MSE of SDGSI

$Y_{ith} = \mu_i S_h + \varepsilon_{ith}$ and

$$\hat{S}_H^{\mathrm{SDGSI}} = \frac{1}{m}\sum_{j=1}^{m}\lambda_j \hat{S}_{jH}^{\mathrm{ISI}} = \frac{1}{m}\sum_{j=1}^{m}\lambda_j \left(\frac{q\sum_{t=1}^{r}Y_{jtH}}{\sum_{t=1}^{r}\sum_{h=1}^{q}Y_{jth}}\right) = \frac{1}{m}\sum_{j=1}^{m}\lambda_j \frac{\sum_{t=1}^{r}Y_{jtH}}{r\mu_j}$$

The forecast for the i th series, H th season, in year $r+1$ using the SDGSI is:

$$\hat{Y}_{i(r+1)H}^{\mathrm{SDGSI}} = \mu_i \times \hat{S}_H^{\mathrm{SDGSI}} = \mu_i \times \frac{1}{m}\sum_{j=1}^{m}\lambda_j \frac{\sum_{t=1}^{r}Y_{jtH}}{r\mu_j}$$

$$= \frac{\mu_i\lambda_1\left(Y_{11H} + Y_{12H} + \ldots + Y_{1rH}\right)}{mr\mu_1} + \frac{\mu_i\lambda_2\left(Y_{21H} + Y_{22H} + \ldots + Y_{2rH}\right)}{mr\mu_2} + \ldots + \frac{\mu_i\lambda_m\left(Y_{m1H} + Y_{m2H} + \ldots + Y_{mrH}\right)}{mr\mu_m}$$

Thus, MSE of SDGSI can be calculated:

$$\mathrm{MSE}^{\mathrm{SDGSI}} = \mathrm{E}\left(Y_{i(r+1)H} - \hat{Y}_{i(r+1)H}^{\mathrm{SDGSI}}\right)^2$$

$$= \mathrm{E}\left[Y_{i(r+1)H} - \frac{\mu_i\lambda_1\left(Y_{11H} + Y_{12H} + \ldots + Y_{1rH}\right)}{mr\mu_1} + \frac{\mu_i\lambda_2\left(Y_{21H} + Y_{22H} + \ldots + Y_{2rH}\right)}{mr\mu_2} + \ldots + \frac{\mu_i\lambda_m\left(Y_{m1H} + Y_{m2H} + \ldots + Y_{mrH}\right)}{mr\mu_m}\right]^2$$

$$= \left(1 - \frac{\sum_{j=1}^{m}\lambda_j}{m}\right)^2 \mu_i^2 S_H^2 + \sigma_i^2 + \frac{\mu_i^2}{m^2 r}\left[\sum_{j=1}^{m}\frac{\lambda_j^2}{\mu_j^2}\sigma_j^2 + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\left(\frac{\lambda_j\lambda_l}{\mu_j\mu_l}\right)\rho_{jl}\sigma_j\sigma_l\right]$$

MSE of SSI

Since the SSI for the mixed model is a special case of the SDGSI when $m = 1$ and $\lambda_i \neq 1$, the MSE is:

$$\mathrm{MSE}^{\mathrm{SSI}} = \left(1-\lambda_i\right)^2 \mu_i^2 S_H^2 + \sigma_i^2 + \frac{\lambda_i^2}{r}\sigma_i^2$$

MSE of DGSSI

Since the DGSSI for the mixed model is a special case of the SDGSI when $m \geq 2$ and $\lambda_1 = ... = \lambda_m = \lambda$, the MSE is:

$$\mathrm{MSE}^{\mathrm{DGSSI}} = \left(1-\lambda\right)^2 \mu_i^2 S_H^2 + \sigma_i^2 + \frac{\mu_i^2\lambda^2}{m^2 r}\left[\sum_{j=1}^{m}\frac{\sigma_j^2}{\mu_j^2} + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\left(\frac{1}{\mu_j\mu_l}\right)\rho_{jl}\sigma_j\sigma_l\right]$$

MSE of WGSSI

Similar to the SDGSI, MSE of WGSSI can be calculated as follows:

$$\mathrm{MSE}^{\mathrm{WGSSI}} = \left(1-\lambda\right)^2 \mu_i^2 S_H^2 + \sigma_i^2 + \frac{\mu_i^2\lambda^2}{r\mu_A^2}\sigma_A^2$$

where $\sigma_A^2 = \sum_{i=1}^{m}\sigma_i^2 + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\rho_{jl}\sigma_j\sigma_l$ is the variance of deseasonlised aggregate demand and $\mu_A = \mu_1 + \mu_2 + ... + \mu_m$ is the aggregate mean.

Appendix C. Optimal shrinkage parameters of SGSI, SSI and GSSI in the additive model

Optimal shrinkage parameters of SGSI

The MSE of the SGSI is (see Appendix A):

$$\mathrm{MSE}^{\mathrm{SGSI}} = \left(1+\frac{1}{qr}\right)\sigma_i^2 + \frac{\left(q-1\right)}{m^2 qr}\left[\sum_{j=1}^{m}\lambda_j^2\sigma_j^2 + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\lambda_j\lambda_l\rho_{jl}\sigma_i\sigma_j\right] + \left[\left(1-\frac{1}{m}\sum_{j=1}^{m}\lambda_j\right)S_H\right]^2$$

where λ_j ($j = 1,...,m$) is a shrinkage parameter for the j th series and $\lambda_j \geq 0$.

A set of optimal shrinkage parameters can be obtained by differentiating the MSE of the SGSI with respect to any shrinkage parameter λ_k ($k = 1,...,m$):

$$\lambda_k = \frac{\left(m - \sum_{j=1,\, j\neq k}^{m}\lambda_j\right)S_H^2 - \left(\frac{q-1}{qr}\right)\sum_{j=1,\, j\neq k}^{m}\lambda_j\rho_{jk}\sigma_j\sigma_k}{S_H^2 + \left(\frac{q-1}{qr}\right)\sigma_k^2}$$

Optimal shrinkage parameters of SSI

$$\lambda_i = \frac{S_H^2}{S_H^2 + \left(\frac{q-1}{qr}\right)\sigma_i^2}$$

Optimal shrinkage parameters of GSSI

$$\lambda = \frac{S_H^2}{S_H^2 + \left(\frac{q-1}{qr}\right)\frac{\sigma_A^2}{m^2}}$$

Appendix D. Optimal shrinkage parameters of SDGSI, SSI, DGSSI and WGSSI in the mixed model

Optimal shrinkage parameters of SDGSI

The MSE of the SDGSI is (see Appendix B for details):

$$\mathrm{MSE}^{\mathrm{SDGSI}} = \left(1 - \frac{\sum_{j=1}^{m}\lambda_j}{m}\right)^2 \mu_i^2 S_H^2 + \sigma_i^2 + \frac{\mu_i^2}{m^2 r}\left[\sum_{j=1}^{m}\frac{\lambda_j^2}{\mu_j^2}\sigma_j^2 + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\left(\frac{\lambda_j\lambda_l}{\mu_j\mu_l}\right)\rho_{jl}\sigma_j\sigma_l\right]$$

where λ_j ($j = 1,...,m$) is a shrinkage parameter for the j th series and $\lambda_j \geq 0$.

A set of optimal shrinkage parameters can be found out by differentiating the MSE of the SDGSI with respect to any shrinkage parameter λ_k ($k = 1,...,m$):

$$\lambda_k = \frac{S_H^2\left(m - \sum_{j=1, j\neq k}^{m}\lambda_j\right) - \frac{1}{r}\left[\sum_{j=1, j\neq k}^{m}\left(\frac{\lambda_j}{\mu_j\mu_k}\right)\rho_{jk}\sigma_j\sigma_k\right]}{\left(S_H^2\right) + \frac{\sigma_k^2}{r\mu_k^2}}$$

Optimal shrinkage parameters of SSI

$$\lambda_i = \frac{S_H^2}{S_H^2 + \frac{\sigma_i^2}{r\mu_i^2}}$$

Optimal shrinkage parameters of DGSSI

$$\lambda = \frac{S_H^2}{S_H^2 + \frac{1}{m^2 r}\left[\sum_{j=1}^{m}\frac{\sigma_j^2}{\mu_j^2} + 2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\left(\frac{1}{\mu_j\mu_l}\right)\rho_{jl}\sigma_j\sigma_l\right]}$$

Optimal shrinkage parameter in WGSSI

Similar to the SDGSI, the optimal shrinkage parameters can be found out by differentiating the MSE of the WGSSI with respect to λ:

$$\lambda = \frac{\mu_i^2 S_H^2}{\mu_i^2 S_H^2 + \dfrac{\mu_i^2 \sigma_A^2}{r\mu_A^2}} = \frac{S_H^2}{S_H^2 + \dfrac{\sigma_A^2}{r\mu_A^2}}$$

where $\sigma_A^2 = \sum_{i=1}^{m} \sigma_i^2 + 2\sum_{j=1}^{m-1} \sum_{l=j+1}^{m} \rho_{jl}\sigma_j\sigma_l$ is the variance of deseasonlised aggregate demand and $\mu_A = \mu_1 + \mu_2 + ... + \mu_m$ is the aggregate mean.

Table 1

The ratios of MSE of all estimators to MSE of the ISI when the length of data history is 2, 3, 4.

	Years	ISI	DGSI	WGSI	SSI	SDGSI	DGSSI	WGSSI
Additive model	2	1	0.7834***		0.8419**	0.7622***	0.7718***	
	3	1	0.8387***		0.8940**	0.8240***	0.8320***	
	4	1	0.8705***		0.9206*	0.8585***	0.8652***	
Mixed model	2	1	0.8107***	0.8093***	0.9673	0.8047***	0.8078***	0.8055***
	3	1	0.8571***	0.8569***	0.9823	0.8534***	0.8556***	0.8549***
	4	1	0.8863***	0.8861***	0.9897	0.8838***	0.8856***	0.8851***

Asterisks indicate p-value for one-tailed paired t-tests over 128 series: *p<0.1;**p<0.05; ***p<0.01.

Table 2

The ratios of MSEs of all estimators to MSE of the ISI when $\beta = 1.2$ and $\beta = 1.8$.

	β	ISI	DGSI	WGSI	SSI	SDGSI	DGSSI	WGSSI
Additive model	1.2	1	0.8297***		0.9624*	0.8206***	0.8278***	
	1.8	1	0.8278***		0.8647**	0.8102***	0.8182***	
Mixed model	1.2	1	0.8603***	0.8491***	0.9973	0.8480***	0.8600***	0.8488***
	1.8	1	0.8468***	0.8479***	0.9723	0.8434***	0.8445***	0.8448***

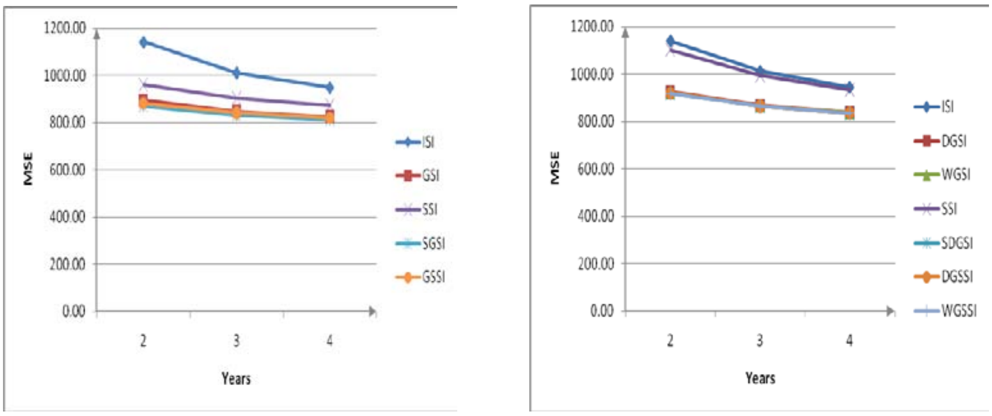
Asterisks indicate p-value for one-tailed paired t-tests over 128 series: *p<0.1;**p<0.05; ***p<0.01.

Table 3

The ratios of MSEs of all estimators to MSE of the ISI when $m = 4$ and $m = 128$.

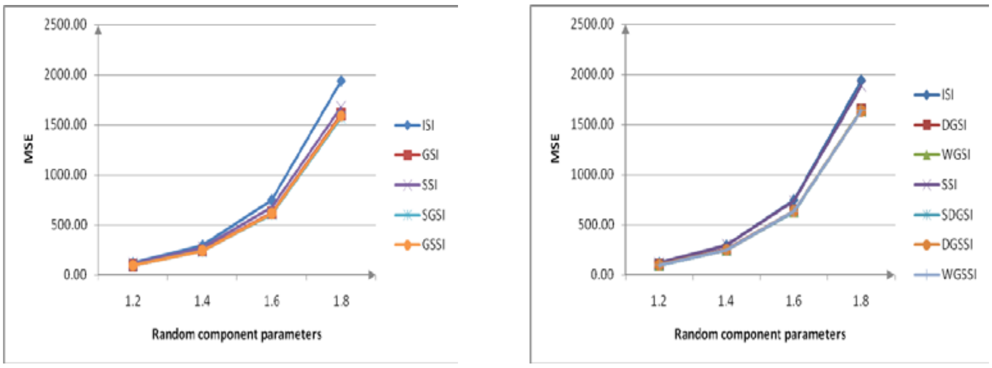
	m	ISI	DGSI	WGSI	SSI	SDGSI	DGSSI	WGSSI
Additive model	4	1	0.8515		0.8818	0.8193	0.8359	
	128	1	0.8018***		0.8817**	0.8017***	0.8018***	
Mixed model	4	1	0.8787	0.8786	0.9789	0.8688	0.8755	0.8739
	128	1	0.8185***	0.8178***	0.9790	0.8182***	0.8185***	0.8177***

Asterisks indicate p-value for one-tailed paired t-tests over 128 series: *p<0.1;**p<0.05; ***p<0.01.



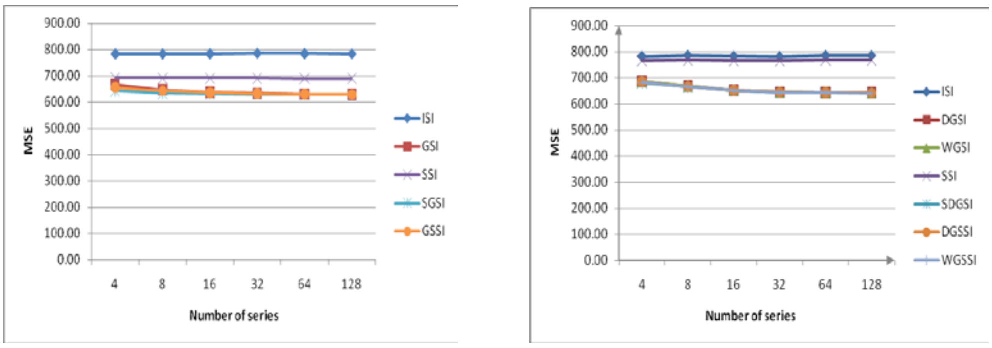
(a) additive model (b)mixed model

Fig. 1. Effect of data history on the reduction in MSE



(a) additive model (b)mixed model

Fig. 2. Effect of variance on the increase in MSE



(a) additive model (b)mixed model

Fig. 3. Effect of number of series on MSE