Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 24 (2008) 525 – 534
www.elsevier.com/locate/ijforecast

Empirical evidence on individual, group and shrinkage seasonal indices

Huijing Chen [a,*], John E. Boylan [b,1]

[a] Maxwell 616e, Centre for Operational Research and Applied Statistics (CORAS), the University of Salford, Salford , M5 4WT, UK
[b] Buckinghamshire Business School, Buckinghamshire New University, Gorelands Lane, Chalfont St. Giles, Buckinghamshire, HP8 4AD, UK

Abstract

This paper provides empirical evidence on forecasting seasonal demand using both individual and group seasonal indices methods. The findings show that the group seasonal indices methods outperform the individual seasonal indices method. This paper also offers empirical results from comparing two shrinkage methods with the group seasonal indices methods. The theoretical rules developed by the authors for choosing between group seasonal indices and individual seasonal indices produce more accurate forecasts than do published rules for choosing between shrinkage methods, when measured by the MSE, and are competitive when measured by the symmetric MAPE.

Keywords: Forecasting; Seasonality; Grouping; Shrinkage

1. Introduction

Short-term forecasting for sales and operations management often concerns many items, perhaps thousands, simultaneously. The conventional forecasting approach is to extrapolate the data series for each stock-keeping unit (SKU) individually. However, most businesses have natural groupings of SKUs, for example by product or geographical families. Duncan, Gorr, and Szczypula (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 product group or across locations has approximately the same seasonality, then the argument of Duncan et al. (1993) may apply, and seasonality estimated at the aggregate level would be more accurate than that estimated individually. We examine the validity of this argument in this paper.

Another approach to improving seasonal forecasting is to dampen (or shrink) individually-estimated

* Corresponding author. Tel.: +44 161 2955414.
E-mail addresses: huijingchen@hotmail.com (H. Chen), john.boylan@bcuc.ac.uk (J.E. Boylan).
[1] Tel.: +44 1494 605130; fax: +44 1494 874230.

0169-2070/$ - see front matter © 2008 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2008.02.005

multiplicative seasonal indices toward 1. Miller and Williams (2003) suggested that individual seasonal indices may not be accurate when uncertainty is large. The authors argued that "…classical decomposition tends to exaggerate seasonal variation, especially under conditions for which simple methods are often preferred (relatively short series, noisy data)". Under those conditions, shrinkage methods may improve forecast accuracy.

This paper is the first to offer an empirical comparison between the grouping and shrinkage approaches. The evaluation confirms that they both improve the accuracy relative to individual seasonal estimation, but that they are comparable to each other. This result provides a stimulus for future work to examine whether further improvements may result from using both methods together.

This paper is organised into the following sections: Section 2 provides a literature survey; Section 3 describes the empirical data set; Section 4 looks at the universal application of a range of methods, including shrinkage methods; Section 5 is a restatement of the model and the rules developed by the authors; Section 6 contains a comparison between the new rules and the guidelines for shrinkage methods; and Section 7 concludes.

2. Previous research

Dalhart (1974) and Withycombe (1989) proposed different forms of group seasonal indices (GSI) methods. Dalhart's Group Seasonal Index (DGSI) is a simple average of individual seasonal indices (ISI), while Withycombe's Group Seasonal Index (WGSI) applies the ISI method to the total demand in the group.

Suppose that data are available for all series in the group for the last n complete seasonal cycles, where the seasonal cycle is q periods, and $T = nq$. Then, if

$$ISI_i = f(X_{i1}, \ldots, X_{iT}) \tag{1}$$

for the ith series, Dalhart's and Withycombe's Group Seasonal Indices are defined respectively by:

$$DGSI = \frac{1}{m}\sum_{i=1}^{m} f(X_{i1}, \ldots, X_{iT}) = \frac{1}{m}\sum_{i=1}^{m} ISI_i \tag{2}$$

and

$$WGSI = f\left(\sum_{i=1}^{m} X_{i1}, \ldots, \sum_{i=1}^{m} X_{iT}\right) \tag{3}$$

for a group of m series.

Dalhart (1974) compared the performances of DGSI and ISI using simulated data only; no tests were conducted on real data. Unfortunately, the details of the methods used for data generation were not fully reported. Also, the forecast for the second year was found by applying seasonal indices to the average monthly demand in the second year, therefore not providing *ex ante* results.

Withycombe (1989) examined the performance of WGSI on 29 products, representing six product groups from three different companies. He found WGSI to be more accurate than the traditional seasonal estimation method for 17 series and for all six product groups, when mean squared error (MSE) results are totalled. One drawback of this analysis is that, for three product groups, one product's MSE was so large that it dominated the results.

Bunn and Vassilopoulos (1993) compared the performance of ISI, DGSI and WGSI empirically. They used 54 weekly series in 5 product groups, with 42 observations in each series. Their research addressed the complementary issue of the formation of groups of SKUs. They regrouped all of the data series into 12 groups using cluster analysis, to obtain homogeneous seasonal profiles for each group. According to their results, WGSI had the lowest MSE for 1, 2 and 3 step-ahead forecasts for 45%, 51% and 50% of the series, respectively. The percentages for DGSI were 33%, 30% and 24%. Similar results were found using the Mean Absolute Error (MAE). The authors concluded that the GSI methods outperformed the ISI method, and overall, WGSI was better than DGSI. However, there were no theoretical comparisons of these methods to understand what conditions each one is the best under.

Gorr, Olligschlaeger, and Thompson (2003) applied ISI and WGSI to crime forecasting, using the city of Pittsburgh and its six police precincts for their study. Forecasting errors were measured by MAE, MSE and MAPE. Because the results were very similar, only the MAPE results were reported. The overall results clearly showed that forecasts using WGSI (pooled seasonality) were more accurate than forecasts using ISI. The authors recommended using WGSI to estimate seasonality at the city-wide level for use in forecasting at the spatially-disaggregated precinct level.

Dekker, van Donselaar, and Ouwehand (2004) proposed a modified version of the Winters' method without the trend component. They estimated the seasonality from the group, and levels from individual items, similar to Gorr et al. (2003). They tested the new method on 67 SKUs in four groups. The results of the one-step-ahead forecasts for the four groups measured by the MAD, MSE and symmetric MAPE show that the modified Winters' method dominates Winters' method.

Miller and Williams (2003) examined "...the possibility of improving forecasting accuracy through better estimation of seasonal factors in situations where relatively simple methods are preferred". The authors proposed two shrinkage estimation methods for examining the potential improvement in forecasting accuracy over the ISI method. Both methods adjust the ISI towards one.

Because the James-Stein shrinkage estimator is based on the restrictive assumption of normally distributed seasonals, the authors suggested another method, the Lemon-Krutchkoff (L-K) estimator. This shrinkage estimator is nonparametric with regard to the prior distribution of the seasonal factors.

To test their shrinkage methods and guidelines (Table 1) empirically, the authors analysed the monthly series from the subset of 111 series used in the M-competition. After discarding unsuitable data, 55 series remained. It was found that both the James-Stein method and the Lemon-Krutchkoff method were more accurate than classical decomposition, and that the selection rules outperformed both individual methods.

In this paper, we make an empirical comparison between the ISI, grouping (DGSI and WGSI methods), and shrinkage approach (J-S and L-K methods). No previous studies have compared the latter schools of thought with the ISI approaches. Furthermore, we compare a set of theoretical rules (Chen & Boylan, 2007) that allow one to choose the best method out of ISI, DGSI and WGSI with the guidelines suggested by Miller and Williams (2003). A later paper by Miller and Williams (2004) developed shrinkage estimators for X-12-ARIMA. We do not discuss this paper here because the methods we consider are based on seasonal decomposition.

Table 1
Guidelines for Using Shrinkage Estimators (Miller & Williams, 2003)

$W^{J\text{–}S}$	Symmetric seasonals	Skewed seasonals
<0.2	Either CD or J-S	L-K
0.2 to 0.5	J-S	L-K
>0.5	J-S	J-S

3. Empirical data set

We obtained a data set from one of the leading manufacturers of light bulbs in the UK. The manufacturer grouped the products into article groups; after removing items with occurrences of zero sales, the final sample has seven groups with 218 items in total. The group sizes, from lowest to highest, are 2, 5, 12, 27, 45, 61 and 66.

In total, we examined 218 series. These are monthly sales data with five years of history. We used the last year's data (October 2002 to September 2003) for forecast evaluation. To assess the effect of data history on the forecasting performance of different seasonal indices methods, and the rules that choose between them, we used four years (October 1998 to September 2002), three years (October 1999 to September 2002), or two years (October 2000 to September 2002) as the estimation period. We use a minimum of two years of estimation data to validate the rules examined in this paper. We applied rolling origins and used the same forecasting horizons to examine the effect of forecasting horizons on the rules and forecasting performances of different methods.

We used the company's definition of groups to examine the robustness of methods to departures from strict seasonal homogeneity. By visual inspection, reasonably consistent seasonal patterns are observed: in general, demands in winter months are higher than in other seasons.

The company treated all of the data series as non-trending with seasonality. However, to check whether they are trended or not, we applied regression to deseasonalised series. Detailed results are presented in the following section.

The ISI method used was $ISI_{i,h} = \frac{q\sum_{t=1}^{r} Y_{ith}}{\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{ith}}$, where Y_{ith} is the demand for item i at the tth year and hth season, q is the number of seasons and r is the number of years. DGSI is a simple average of all the ISIs, while WGSI calculates the seasonal indices from the aggregate series. The level is estimated using $\hat{\mu}_i = \frac{\sum_{t=1}^{r}\sum_{h=1}^{q} Y_{ith}}{qr}$, and the forecast for a particular month is the estimated level multiplied by the estimated seasonal index for that month. The underlying level is assumed to be stationary, to correspond to the model we assumed to derive the rules. The rules were validated by simulated data assuming stationary means (Chen, 2005). In this empirical analysis, we wanted to test how well the rules work even when that assumption might be violated.

4. Findings on the universal application of methods

The error measures we choose are the mean square error (MSE) and symmetric mean absolute percentage error (sMAPE). The MSE is the measure we used to derive the rules to choose the best method between ISI, DGSI and WGSI, as it gives a heavier penalty to larger errors, while sMAPE is a unit-free error measure. Miller and Williams (2003) used MAPE. Both the MAPE and the symmetric MAPE are scale independent. We use the symmetric MAPE because it is less sensitive to low demand than MAPE.

Table 2
Pair-wise percentage better between ISI and grouping methods with 4 years' history

	MSE			sMAPE		
	DGSI< ISI	WGSI< ISI	DGSI< WGSI	DGSI< ISI	WGSI< ISI	DGSI< WGSI
Point forecasts						
1	76.1***	70.2***	64.7***	56.9**	54.1	56.4**
3	65.1***	63.3***	62.8***	54.1	50.5	61.0***
6	56.9**	52.8	65.1***	42.7	44.5	63.8***
9	54.1	46.8	64.7***	44.5	39.4	65.6***
Cumulative forecasts						
3	66.1***	62.8***	60.6***	57.3**	52.2	64.7***
6	61.0***	56.9**	52.8	50.0	47.2	48.6
9	53.2	51.8	45.4*	53.2	51.4	41.3***

p-values for sign tests: *$p<0.10$; **$p<0.05$; ***$p<0.01$. All tests are one-tailed.

Table 3
Pair-wise percentage better between ISI and shrinkage methods with 4 years' history

	MSE			sMAPE		
	SJS< ISI	SLK< ISI	SJS< SLK	SJS< ISI	SLK< ISI	SJS< SLK
Point forecasts						
1	86.7***	78.4***	78.0***	55.5*	57.8**	45.0
3	80.7***	73.4***	64.7***	51.8	53.7	41.3**
6	63.8***	67.4***	50.9	43.1	48.6	36.7***
9	51.8	55.5*	46.8	39.9	43.1	37.6***
Cumulative forecasts						
3	68.8***	68.8***	56.9**	50.9	54.1	44.5
6	59.2***	65.1***	50.0	45.9	57.8**	42.7**
9	48.6	54.1	48.2	45.4	56.0**	44.0*

p-values for sign tests: *$p<0.10$; **$p<0.05$; ***$p<0.01$. All tests are one-tailed.

In the following tables, we report overall pair-wise Percentage Better (PB) results for the 218 series in this data set under MSE and sMAPE for applications of ISI, DGSI, and WGSI. PB reports how often one method performs better than another. It is an intuitive nonparametric procedure (Makridakis & Hibon, 1995). We then conducted sign tests to assess whether the difference between the two methods in a pair is statistically significant. The same tests were used in Miller and Williams (2003). We only present the results with four years' history, because the results with two and three years' history were very similar.

We made point forecasts of 1, 3, 6 and 9 steps ahead and cumulative forecasts of 3, 6 and 9 steps ahead (see Table 2).

Overall, both GSI methods have higher PB results than ISI. DGSI is the best universal application. The results are consistent under the MSE and symmetric MAPE, but the percentages are higher for the GSI methods when the errors are measured by the MSE. For this data set, ISI performs better for longer forecasting horizons. This finding is consistent with Ouwehand, van Donselaar, and de Kok (2005), but the reasons for it require further investigation.

Similarly, we use pair-wise sign tests to assess whether there are significant differences between ISI and the shrinkage methods. SJS stands for the James-Stein shrinkage method and SLK stands for the Lemon-Krutchkoff shrinkage method. The results are shown in Table 3.

Table 4
Pair-wise percentage better between the grouping and shrinkage methods with 4 years' history

	MSE				sMAPE			
	DGSI < SJS	DGSI < SLK	WGSI < SJS	WGSI < SLK	DGSI < SJS	DGSI < SLK	WGSI < SJS	WGSI < SLK
Point forecasts								
1	61.5***	64.7***	54.6*	59.2***	54.6*	56.4**	47.2	48.6
3	56.9**	58.3***	51.8	56.9**	52.8	52.8	48.2	48.6
6	50.9	52.3	50.5	50.0	50.5	46.3	45.4*	43.1**
9	53.7	50.5	47.2	46.3	48.6	42.2**	49.1	41.7***
Cumulative forecasts								
3	57.8**	59.2***	50.5	53.2	60.6***	55.0*	46.8	47.2
6	61.0***	57.3**	51.8	51.8	56.4**	50.0	51.8	50.0
9	60.1***	57.8**	54.1	54.1	58.7***	55.5*	56.4**	54.6*

p-values for sign tests: *$p<0.10$; **$p<0.05$; ***$p<0.01$. All tests are one-tailed.

Our results are consistent with the Miller and Williams (2003) finding that both shrinkage methods improve in forecasting accuracy over ISI. The shrinkage methods have higher PB results when MSE is used. The percentage decreases when the forecasting horizon increases. It is not clear which of the two shrinkage methods is better. In general, SJS is better more often under MSE; SLK is better more often under symmetric MAPE.

Table 4 presents the pairwise comparisons between the grouping approach and the shrinkage approach and shows that the methods are comparable.

Under the MSE, DGSI is slightly better than SJS or SLK, but the differences are less significant under the symmetric MAPE. WGSI appears to be better than SLK under MSE for most forecasting horizons, but the percentage differences are small.

Bunn and Vassilopoulos (1993) directly compared ISI, DGSI and WGSI for 1, 2, and 3 step lead-time forecasts with rolling origins under MAD and MSE. They found that WGSI had the highest PB (from 40% to 51%). This does not agree with our results; however, the seasonal patterns in their data were very homogeneous within the groups selected by cluster analysis. Earlier simulation findings (Chen, 2005) have shown that, when assuming common seasonality, WGSI performs better than DGSI. Although in an empirical analysis one cannot assume common seasonality, Bunn and Vassilopoulos (1993) defined groups with highly homogeneous seasonal patterns, and WGSI was

Table 5
Relative MSE, based on ISI, for the grouping and shrinkage methods

	Methods	Point forecasts				Cumulative forecasts		
		1	3	6	9	3	6	9
4 years	DGSI	**0.84**	**0.85**	**0.92**	**0.89**	**0.88**	**0.93**	**0.94**
	WGSI	0.87	0.88	0.95	0.98	0.93	0.98	0.96
	SJS	0.87	0.91	0.96	0.99	0.96	1.03	1.00
	SLK	0.91	0.91	0.93	0.92	0.96	1.01	1.00
3 years	DGSI	**0.81**	**0.84**	**0.88**	**0.91**	**0.87**	**0.96**	**0.93**
	WGSI	0.86	0.87	0.90	1.00	0.93	1.01	0.94
	SJS	0.84	0.88	0.93	0.99	0.97	1.07	1.01
	SLK	0.88	0.90	0.91	0.96	0.96	1.02	0.99
2 years	DGSI	**0.74**	**0.74**	**0.79**	0.88	**0.83**	0.93	0.96
	WGSI	0.78	0.77	0.83	0.97	0.91	0.98	0.96
	SJS	0.75	0.78	0.81	**0.87**	0.84	**0.92**	0.97
	SLK	0.83	0.84	0.82	0.89	0.89	0.93	**0.95**

the best method. We use groups defined by the company, and cannot be sure that the seasonal patterns are from the same population.

Bunn and Vassilopoulos (1993) also compared relative percentage differences between ISI and GSIs. They calculated (DGSI − ISI)/ISI and (WGSI − ISI)/ISI for each series and averaged over each lead time. Overall, WGSI was 6% better than ISI under both MAD and MSE. DGSI was 3% better than ISI under MAD but 58% worse under MSE. No reasons were given for this large percentage difference. Relative percentage differences can avoid the scale dependency problem of MAD and MSE when summarised across series. Using the arithmetic mean to average the results may be misleading because the overall results can be dominated by a few extremely large percentages. In our empirical analysis, when we summarise the overall findings, we use the geometric mean rather than the arithmetic mean to overcome this problem.

We report the relative MSEs of DGSI, WGSI, SJS and SLK based on ISI in Table 5. The MSEs are geometric mean values for different forecasting horizons across all series. The best result for each forecasting horizon and data length is in bold font.

Under MSE, DGSI is never worse than ISI for any forecasting horizon or any of the three different data lengths. WGSI is better than ISI for most of the horizons, but there are a few exceptions. Our empirical results also agree with the simulation experiments of Miller and Williams (2003), that both shrinkage methods are better than ISI overall in terms of forecasting accuracy. Under MSE, DGSI is the overall best universal application for all of the *n*-step-ahead and cumulative forecasts. The benefit of using the grouping or shrinkage estimators is greater when the data history is shorter.

We also report the sMAPE results in Table 6.

Both of the grouping and shrinkage approaches are better than ISI when judged by sMAPE. Overall, DGSI performed better than WGSI and SLK performed better than SJS. The overall best method under sMAPE depends on the data length and forecasting horizon.

To determine whether the series are trended or not, we used an estimation period of four years for each series, and applied a classical decomposition. We also tested the significance of the trend lines. Those with p values greater than 0.05 were treated as non-trended. We then divided the series into two subsets: trended series and non-trended series. For the trended series, we deseasonalised the series with each set of seasonal indices, estimated the trend using regression, and reseasonalised the forecasts. For the non-trended series, the forecasts are the overall mean multiplied by each set of seasonal indices. The smallest group originally had two series, and the test shows that one is

Table 6
sMAPEs of the universal applications

	Methods	Point forecasts				Cumulative forecasts		
		1	3	6	9	3	6	9
2 years	ISI	49.24	49.31	50.51	49.35	33.42	27.99	25.24
	DGSI	45.84	46.24	49.03	49.05	32.09	27.36	25.01
	WGSI	46.59	46.78	49.72	50.05	33.32	27.73	**24.92**
	SJS	**45.34**	46.06	47.91	46.55	**31.97**	27.60	25.19
	SLK	45.67	**45.86**	**47.01**	**46.18**	32.11	**27.20**	24.98
3 years	ISI	46.57	47.01	48.54	47.32	33.87	29.27	27.00
	DGSI	45.29	45.96	48.66	47.41	**32.94**	**29.20**	**26.60**
	WGSI	46.08	46.56	49.43	48.67	33.77	29.46	26.55
	SJS	45.05	46.05	48.85	47.91	33.81	29.73	26.97
	SLK	**44.78**	**45.52**	**47.56**	**46.74**	33.27	29.21	26.78
4 years	ISI	47.15	47.27	48.67	48.25	35.00	30.89	28.44
	DGSI	**45.62**	**45.94**	48.92	47.68	**34.24**	**30.61**	**27.94**
	WGSI	46.17	46.49	49.53	48.78	34.99	31.12	28.14
	SJS	46.02	46.65	49.30	48.77	34.89	31.04	28.46
	SLK	46.09	46.26	**48.28**	**47.66**	34.61	30.95	28.43

Table 7
Average sMAPE results

Methods	Point forecasts				Cumulative forecasts		
	1	3	6	9	3	6	9
ISI	80.04	85.89	97.48	107.50	149.40	99.77	191.53
DGSI	77.83	82.68	91.17	**71.44**	78.16	**49.99**	**44.01**
WGSI	**77.31**	**76.69**	**87.48**	87.46	**61.69**	57.77	53.00
SJS	80.34	86.25	100.61	77.56	132.21	52.19	56.75
SLK	81.79	88.36	89.62	92.51	69.48	54.44	64.10

trended and the other is not. After they are divided into the two subsets, the group no longer exists. These two series are therefore removed. Table 7 shows the overall sMAPE results, averaged over 177 trended series and 39 non-trended series.

Comparing the sMAPE results in the above table with those in Table 6, where all series were considered as non-trended, we observe that using the trending method does not improve the forecasting accuracy. The simple method of treating the data series as non-trended proved to be robust. Even when some of them are trended according to a statistical test, to estimate the trend accurately with a short data history is not an easy task. In our case, assuming the series to be non-trended proved to be beneficial.

We also compared all the seasonal indices methods to a non-seasonal method; that is, using the estimate of the level as the forecast, we (a) confirmed that the data are indeed seasonal, even though for some groups the seasonal patterns seem weak when inspected visually; and (b) quantified the benefit by using a seasonal method. Table 8 shows the relative MSE based on the non-seasonal method.

These results confirmed that the data were seasonal, and therefore, that it was better to use a seasonal forecasting method. The benefit is greater when data lengths increase. It is worth noting that sometimes (e.g, when there are only two years' data available to make 1 step ahead forecasts) it is worse to use ISI than a non-seasonal method even when the data are seasonal. A poor estimation of seasonal indices could be worse than not estimating the seasonal effect at all. This is also true for the trend component.

Table 9
Relative MSE for groups for which DGSI and WGSI were recommended

Methods	Point forecasts				Cumulative forecasts		
	1	3	6	9	3	6	9
DGSI best: 6 groups (173 series)							
DGSI	**0.79**	**0.80**	**0.84**	**0.87**	**0.84**	**0.90**	**0.93**
WGSI	0.83	0.83	0.88	0.97	0.92	0.98	0.95
WGSI best: 1 group (45 series)							
DGSI	0.88	0.92	0.97	**0.95**	1.02	1.12	1.02
WGSI	**0.86**	**0.89**	**0.95**	1.09	**0.95**	**1.02**	**0.96**

Table 8
Relative MSEs based on the non-seasonal method

	Methods	Point forecasts				Cumulative forecasts		
		1	3	6	9	3	6	9
2 years	ISI	1.38	1.28	1.05	0.92	0.96	0.81	0.93
	DGSI	**1.02**	**0.94**	**0.82**	0.80	**0.79**	**0.75**	0.90
	WGSI	1.08	0.99	0.87	0.89	0.87	0.79	0.90
	SJS	1.04	1.00	0.85	**0.79**	0.81	**0.75**	0.91
	SLK	1.15	1.08	0.86	0.81	0.85	**0.75**	**0.89**
3 years	ISI	1.15	1.06	0.89	0.78	0.87	0.74	0.89
	DGSI	**0.93**	**0.89**	**0.79**	**0.71**	**0.76**	**0.72**	**0.83**
	WGSI	0.98	0.92	0.81	0.79	0.80	0.75	**0.83**
	SJS	0.96	0.93	0.83	0.78	0.84	0.80	0.90
	SLK	1.01	0.96	0.82	0.76	0.84	0.76	0.88
4 years	ISI	1.08	0.96	0.81	0.75	0.85	0.76	0.89
	DGSI	**0.91**	**0.82**	**0.74**	**0.67**	**0.75**	**0.71**	**0.83**
	WGSI	0.94	0.85	0.77	0.73	0.79	0.75	0.85
	SJS	0.94	0.88	0.78	0.74	0.81	0.78	0.89
	SLK	0.99	0.88	0.75	0.69	0.82	0.77	0.89

Table 10
Ratio of MSE for Miller and Williams' (2003) rules to Chen and Boylan's (2007) rules

	Point forecasts				Cumulative forecasts		
	1	3	6	9	3	6	9
2 years	1.01	1.06	1.09	1.02	1.04	1.05	1.03
3 years	1.04	1.05	1.06	1.07	1.12	1.13	1.10
4 years	1.03	1.06	1.04	1.08	1.08	1.10	1.07

5. Theoretical rules for choosing the best method

Theoretical rules determine under what conditions the ISI, DGSI and WGSI methods minimize the MSE (see the Appendix A). Detailed derivations can be found in Chen and Boylan (2007).

With these rules, we compare each series' coefficient of variation to two different versions of the "group" coefficient of variation, corresponding to the ways DGSI and WGSI are calculated. If the individual series is less noisy than the group average, ISI should be used. If the "group" coefficient of variation is less noisy than the individual series, then it is better to use one of the GSI methods, because the noisy series can "borrow strength" from the group.

The rules select DGSI for six groups and WGSI for one group. We checked whether DGSI was the best for those six groups and whether WGSI was the best for the other group. Table 9 presents the relative MSE based on ISI (DGSI/ISI and WGSI/ISI), with the best result for each forecasting horizon in bold font.

For the six groups for which DGSI was recommended by the rule, DGSI was better than ISI and WGSI for all forecasting horizons. For the one group where WGSI was recommended, WGSI was generally better than ISI and DGSI. Although the rules were derived from a stationary mean, stationary seasonality model with fixed origins, they are robust when applied with rolling origins, as validated by results under the MSE and symmetric MAPE.

6. Comparing the rules

Miller and Williams (2003) suggested guidelines, based on their simulation experiments, for choosing the best method between ISI, the James-Stein shrinkage method, and the Lemon-Krutchkov method. In this section, we compare the rules we developed by Chen and Boylan (2007) with the Miller and Williams guidelines. We calculated the relative MSE of Miller and Williams' guidelines to our rules (see Table 10).

Overall, our rules are better than the Miller and Williams guidelines for all forecasting horizons, as all of the values are greater than one. The results are consistent over all three lengths of data history. The results are also consistent with expectations, because our rules are based on MSE.

In Table 11, we present average symmetric MAPEs to provide another indication of forecasting accuracy.

Our rules do not consistently outperform Miller and Williams' guidelines under sMAPE. However, our rules are better for the cumulative forecasts and when there are four years of data.

7. Conclusions

This paper provides further empirical evidence on the performance of the ISI, DGSI, and WGSI methods with the largest empirical dataset used thus far. Measured by the MSE and sMAPE, both GSI methods outperformed

Table 11
Average sMAPE results

	Methods	Point forecasts				Cumulative forecasts		
		1	3	6	9	3	6	9
2 years	R	46.03	46.23	48.35	48.43	**31.77**	**26.85**	**24.79**
	RMW	**45.26**	**45.98**	**47.84**	**46.47**	31.92	27.56	25.19
3 years	R	45.33	46.02	48.82	47.83	**32.86**	**29.05**	**26.50**
	RMW	**45.00**	**46.00**	**48.79**	**47.82**	33.81	29.71	26.96
4 years	R	**45.67**	**46.04**	**49.05**	**48.01**	**34.23**	**30.54**	**27.90**
	RMW	45.96	46.59	49.25	48.69	34.87	31.02	28.45

ISI, supporting Bunn and Vassilopoulos' (1993) findings. We found that DGSI was the best universal application, disagreeing with Bunn and Vassilopoulos (1993), who found that WGSI was the best. One of the reasons for this may relate to the definition of seasonal homogeneity, although further research is needed to clarify this.

Miller and Williams (2003) proposed two shrinkage methods: James-Stein and Lemon-Krutchkoff. We applied these two methods to the same empirical dataset. We found that both shrinkage methods performed better than ISI. Overall, 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, the two approaches need not be perceived as strict competitors (see Bunn & Vassilopoulos, 1999). Linking the two approaches offers the potential for additional gains in forecast accuracy. This is the subject of current research by the authors.

Empirical results on the theoretical rules developed by the authors for choosing the best method between the ISI and GSI methods are also included in this paper. We compared our rules to the Miller and Williams (2003) guidelines. Overall, our rules are consistently better than those guidelines when the forecast errors are measured by the MSE. When sMAPEs are used, the results are mixed. Our rules are better for cumulative forecasts and when there is a longer data history. Overall, our rules are competitive with Miller and Williams' (2003) guidelines.

Acknowledgements

We are indebted to an associate editor and two anonymous referees for their detailed comments and helpful suggestions.

Appendix A. Theoretical rules for choosing between the ISI and GSI methods

The rules are based on the following model:

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

where i is a suffix representing the SKU or the location; a suffix t represents the year, and $t=1,2,\ldots r$; a suffix h represents the seasonal period, and $h=1,2,\ldots q$;

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 is the same for all SKUs or locations under consideration; and

ε_{ith} 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 ε_{ith} and ε_{jth} at the same time period. Autocorrelations and cross correlations at different time periods are assumed to be zero.

The rules can be summarised as:

$$\text{if } \min\left(\frac{\sigma_i^2}{\mu_i^2}, \frac{1}{m^2}\left(\frac{\sigma_1^2}{\mu_1^2}+\frac{\sigma_2^2}{\mu_2^2}+\ldots+\frac{\sigma_m^2}{\mu_m^2}+2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\frac{1}{\mu_j}\frac{1}{\mu_l}\rho_{jl}\sigma_j\sigma_l\right), \frac{\sigma_A^2}{\mu_A^2}\right) = \frac{\sigma_i^2}{\mu_i^2},$$

then choose the ISI method.

$$\text{if } \min\left(\frac{\sigma_i^2}{\mu_i^2}, \frac{1}{m^2}\left(\frac{\sigma_1^2}{\mu_1^2}+\frac{\sigma_2^2}{\mu_2^2}+\ldots+\frac{\sigma_m^2}{\mu_m^2}+2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\frac{1}{\mu_j}\frac{1}{\mu_l}\rho_{jl}\sigma_j\sigma_l\right), \frac{\sigma_A^2}{\mu_A^2}\right) = \frac{1}{m^2}\left(\frac{\sigma_1^2}{\mu_1^2}+\frac{\sigma_2^2}{\mu_2^2}+\ldots+\frac{\sigma_m^2}{\mu_m^2}+2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\frac{1}{\mu_j}\frac{1}{\mu_l}\rho_{jl}\sigma_j\sigma_l\right),$$

then choose the DGSI method.

$$\text{if } \min\left(\frac{\sigma_i^2}{\mu_i^2}, \frac{1}{m^2}\left(\frac{\sigma_1^2}{\mu_1^2}+\frac{\sigma_2^2}{\mu_2^2}+\ldots+\frac{\sigma_m^2}{\mu_m^2}+2\sum_{j=1}^{m-1}\sum_{l=j+1}^{m}\frac{1}{\mu_j}\frac{1}{\mu_l}\rho_{jl}\sigma_j\sigma_l\right), \frac{\sigma_A^2}{\mu_A^2}\right) = \frac{\sigma_A^2}{\mu_A^2},$$

then choose the WGSI method.

References

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

Bunn, D. W., & Vassilopoulos, A. I. (1999). Comparison of Seasonal Estimation Methods in Multi-item Short-term Forecasting. *International Journal of Forecasting*, *15*, 431–443.

Chen, H. (2005). *Comparing Individual with Group Seasonal Indices to Forecast Subaggregate Demand*. PhD thesis, Buckinghamshire Chilterns University College, Brunel University, UK.

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.

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

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

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

Gorr, W., Olligschlaeger, A., & Thompson, Y. (2003). Short-term Forecasting of Crime. *International Journal of Forecasting*, *19*, 579−594.

Makridakis, S., & Hibon, M. (1995). *Evaluating Accuracy (or error) Measures*. Working paper 95/18/TM, INSEAD, France.

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*, 529−549.

Ouwehand, P., van Donselaar, K. H., & de Kok, A. G. (2005). *The Impact of the Forecasting Horizon when Forecasting with Group Seasonal Indices*. Working paper 162, Eindhoven University of Technology, the Netherlands.

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

Huijing Chen is Lecturer in Management Science at the University of Salford. She completed an MSc. in Management Science and Operational Research at the University of Warwick in 2001 and received her Ph.D. from Brunel University in 2005. Her current research interests are in short-term seasonal demand forecasting.

John Boylan is Professor of Management Science at Buckinghamshire New University. He holds degrees from Oxford and Warwick Universities and has published papers on short-term forecasting in a variety of academic and practitioner-oriented journals. His current research interests relate to demand forecasting in the supply chain.