ISSN 1440-771X

MONASH University

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

A state space model for exponential smoothing with group seasonality

Pim Ouwehand, Rob J Hyndman, Ton G. de Kok and Karel H. van Donselaar

June 2007

Working Paper 07/07

A state space model for exponential smoothing with group seasonality

Pim Ouwehand, Department of Technology Management,
Eindhoven University of Technology,
P.O. Box 513,
5600 MB Eindhoven,
The Netherlands
Email: p.ouwehand@tm.tue.nl

Rob J Hyndman
Department of Econometrics and Business Statistics,
Monash University, VIC 3800
Australia.
Email: Rob.Hyndman@buseco.monash.edu.au

Ton G. de Kok, Department of Technology Management,
Eindhoven University of Technology,
P.O. Box 513,
5600 MB Eindhoven,
The Netherlands
Email: a.g.d.kok@tm.tue.nl

Karel H. van Donselaar, Department of Technology Management,
Eindhoven University of Technology,
P.O. Box 513,
5600 MB Eindhoven,
The Netherlands
Email: k.h.v.donselaar@tm.tue.nl

11 June 2007

JEL classification: C53,C22,C52

A state space model for exponential smoothing with group seasonality

Abstract: We present an approach to improve forecast accuracy by simultaneously forecasting a group of products that exhibit similar seasonal demand patterns. Better seasonality estimates can be made by using information on all products in a group, and using these improved estimates when forecasting at the individual product level. This approach is called the group seasonal indices (GSI) approach, and is a generalization of the classical Holt-Winters procedure. This article describes an underlying state space model for this method and presents simulation results that show when it yields more accurate forecasts than Holt-Winters.

Keywords: common seasonality, demand forecasting, exponential smoothing, Holt-Winters, state space model.

1 Introduction

In business, often demand forecasts for many hundreds or thousands of items are required, and thus must be made in an automatic fashion. For this reason, simple extrapolative methods are widely used in practice. The standard methodology is to forecast each product's demand separately. However, data at this level is usually subject to a relatively large amount of noise. More accurate forecasts can be made by considering groups of products that have similar demand patterns.

In this article, our particular interest is in groups of products with similar seasonal patterns for which forecasts are made by exponential smoothing methods. Since exponential smoothing characterizes seasonality by a set of seasonal indices, we refer to this as the *group seasonal indices* (GSI) approach. In this approach, seasonal indices are estimated using information on all products in a group and used when forecasting at the individual product level. This improves the quality of the seasonality estimates, thereby resulting in more accurate forecasts. Because there is only one observation for each seasonal index per complete cycle (e.g., a year), there is opportunity for improving seasonality estimates. In this article, we present a statistical framework for this GSI approach, which is a generalization of the classical Holt-Winters procedure (Holt, 1957; Winters, 1960).

Publications on group seasonality approaches can be traced back to Dalhart (1974), who first proposed to estimate seasonality from aggregate data. Later, this was extended by Withycombe (1989) and Bunn and Vassilopoulos (1993, 1999). All studies report the improvement potential of their methods over standard methods. However, the experiments are only on a small scale and only focus on short term forecasts. Besides, seasonal indices are assumed to be fixed through time. Once they are estimated, they are not updated in subsequent periods. In the current article, an approach is studied that is based on smoothing of the seasonal estimates. The approach updates level and trend components at the item level, while the seasonal component is updated using a pooled seasonality estimate. Empirical results (Dekker et al., 2004; Ouwehand et al., 2004) have shown significant improvement potential over the Holt-Winters method.

All earlier publications present only empirical and simulation experiments to establish the potential improvement of a GSI approach. In Chen (2005), a first theoretical comparison is given of the methods of Dalhart (1974) and Withycombe (1989). These methods are compared with a traditional individual seasonal indices method and conditions are derived under which one method is preferred to the other. However, data processes are considered for which the methods studied are not necessarily the most appropriate choices.

In the next section, we develop a statistical framework that specifies the data processes for which our GSI method is the optimal forecasting approach. It provides a statistical basis for the GSI approach by describing an underlying state space model for the GSI method. For data following this model, GSI generates forecasts with minimal forecast error variances. The model is an extension of the model underlying the Holt-Winters method, described in Ord et al. (1997) and Koehler et al. (2001).

In Section 3, we present a simulation study that determines in which situations GSI yields better forecasts than Holt-Winters, and that gives an indication of how much the accuracy can be improved under various parameter settings and types of demand patterns. The main results are that GSI performs better than Holt-Winters if there is more similarity in a group's seasonal patterns, under larger amounts of noise, for larger and more homogeneous groups, for longer forecast horizons, and when less historical data is available.

2 Group seasonal indices

2.1 Model

In this section, we present a theoretical framework for the GSI approach. By identifying an underlying model, we determine for which data processes GSI is the optimal method. For data following the underlying model, the method generates forecasts with a minimal Mean Squared Error (MSE). The model is a generalization of the models underlying the Holt-Winters method, as described in Ord et al. (1997) and Koehler et al. (2001). The model underlies the GSI method, which pools the time series to improve forecasts, and which is a generalization of the Holt-Winters method.

Our focus is on multiplicative seasonality. Although seasonality could include the additive case, it is less likely that in practice a group of time series can be found with the same additive seasonality. Since the GSI method is thus nonlinear, there will not be an ARIMA model underlying this method. However, the class of state space models has provided a way to underpin this method. Since the resulting models are also nonlinear, the standard Kalman filter does not apply. Nonlinear state space models with multiple sources of error are usually estimated using quasi-maximum likelihood methods in conjunction with an extended Kalman filter.

In Ord et al. (1997) a class of nonlinear state space models is introduced that have only one disturbance term. These simpler models can be estimated by a conditional maximum likelihood procedure based on exponential smoothing instead of an extended Kalman filter. One of the models in this class is the model underlying the multiplicative Holt-Winters procedure. The minimum mean squared error updating equations and forecast function for this model correspond to those of the HW method. The model is based on a single source of error (Snyder, 1985), and a multiplicative error term (Ord and Koehler, 1990). The model has a single noise process describing the development of all time series components, and is specified by

$$y_t = (\ell_{t-1} + b_{t-1})s_{t-m} + (\ell_{t-1} + b_{t-1})s_{t-m}\varepsilon_t \tag{1a}$$

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha_1(\ell_{t-1} + b_{t-1})\varepsilon_t \tag{1b}$$

$$b_t = b_{t-1} + \alpha_2(\ell_{t-1} + b_{t-1})\varepsilon_t \tag{1c}$$

$$s_t = s_{t-m} + \alpha_3 s_{t-m}\varepsilon_t \tag{1d}$$

where y_t denotes the times series, ℓ_t is the underlying level, b_t the growth rate, and s_t the seasonal factor. The number of seasons per year is equal to m. Furthermore, ε_t are serially uncorrelated disturbances with mean zero and variance σ_ε^2. Solving the measurement equation for ε_t and substituting for ε_t in the transition equations gives the error-correction form for these models:

$$\ell_t = \ell_{t-1} + b_{t-1} + \alpha_1 e_t/s_{t-m} \tag{2a}$$

$$b_t = b_{t-1} + \alpha_2 e_t/s_{t-m} \tag{2b}$$

$$s_t = s_{t-m} + \alpha_3 e_t/(\ell_{t-1} + b_{t-1}) \tag{2c}$$

with $e_t = y_t - (\ell_{t-1} + b_{t-1})s_{t-m}$. The initial states ℓ_0, b_0 and $s_{-m+1}, \dots, s_0$ and the parameters have to be estimated, after which consecutive estimates of ℓ_t, b_t and s_t can be calculated from these formulae. After estimation, the transition equations show great similarity with the error-correction form of classical multiplicative Holt-Winters:

$$\begin{aligned}
\hat{y}_t(h) &= (\hat{\ell}_t + h\hat{b}_t)\hat{s}_{t+h-m} && \text{(3a)}\\
\hat{\ell}_t &= \hat{\ell}_{t-1} + \hat{b}_{t-1} + \hat{\alpha}\hat{e}_t/\hat{s}_{t-m} && \text{(3b)}\\
\hat{b}_t &= \hat{b}_{t-1} + \hat{\alpha}\hat{\beta}\hat{e}_t/\hat{s}_{t-m} && \text{(3c)}\\
\hat{s}_t &= \hat{s}_{t-m} + \hat{\gamma}(1-\hat{\alpha})\hat{e}_t/\hat{\ell}_t && \text{(3d)}
\end{aligned}$$

The only difference is the denominator on the right-hand side of the updating equation for the seasonal indices, but since $\ell_t \approx \ell_{t-1} + b_{t-1}$, the difference is only minor.

Extending the above to the multivariate case, we identify the following model underlying the GSI approach:

$$\begin{aligned}
y_{i,t} &= (\ell_{i,t-1} + b_{i,t-1})s_{t-m} + (\ell_{i,t-1} + b_{i,t-1})s_{t-m}\varepsilon_{i,t} && \text{(4a)}\\
\ell_{i,t} &= \ell_{i,t-1} + b_{i,t-1} + \alpha_i(\ell_{i,t-1} + b_{i,t-1})\varepsilon_{i,t} && \text{(4b)}\\
b_{i,t} &= b_{i,t-1} + \alpha_i\beta_i(\ell_{i,t-1} + b_{i,t-1})\varepsilon_{i,t} && \text{(4c)}\\
s_t &= s_{t-m} + \gamma s_{t-m}\sum_{i=1}^{N} w_i\varepsilon_{i,t} && \text{(4d)}
\end{aligned}$$

where $i = 1, \ldots, N$ denotes the items in the product group and $\ell_{i,t}$ and $b_{i,t}$ denotes their level and trend components. All items share a common seasonality, denoted by s_t. All items have normally distributed disturbances $\varepsilon_{i,t}$. The w_i are weights that sum to 1. This model can be rewritten in error-correction form as

$$\begin{aligned}
\ell_{i,t} &= \ell_{i,t-1} + b_{i,t-1} + \alpha_i e_{i,t}/s_{t-m} && \text{(5a)}\\
b_{i,t} &= b_{i,t-1} + \alpha_i\beta_i e_{i,t}/s_{t-m} && \text{(5b)}\\
s_t &= s_{t-m} + \gamma\sum_{i=1}^{N}\frac{w_i e_{i,t}}{\ell_{i,t-1} + b_{i,t-1}} && \text{(5c)}
\end{aligned}$$

with $e_{i,t} = y_{i,t} - (\ell_{i,t-1} + b_{i,t-1})s_{t-m}$. The forecasting method resulting from this model has the following updating equations:

$$\begin{aligned}
\hat{\ell}_{i,t} &= \alpha_i\frac{y_{i,t}}{\hat{s}_{t-m}} + (1-\alpha_i)(\hat{\ell}_{i,t-1} + \hat{b}_{i,t-1}) && \text{(6a)}\\
\hat{b}_{i,t} &= \beta_i(\hat{\ell}_{i,t} - \hat{\ell}_{i,t-1}) + (1-\beta_i)\hat{b}_{i,t-1} && \text{(6b)}
\end{aligned}$$

$$\hat{s}_t = \gamma \sum_{i=1}^{N} \frac{w_i y_{i,t}}{\hat{\ell}_{i,t-1} + \hat{b}_{i,t-1}} + (1-\gamma)\hat{s}_{t-m} \tag{6c}$$

A h-step ahead forecast for item i, made at time t, is given by $\hat{y}_{i,t}(h) = (\hat{\ell}_{i,t} + h\hat{b}_{i,t})\hat{s}_{t+h-m}$. To see why this model indeed yields these equations for forecasts and updates of state variables, first consider the single source of error version of the random walk plus noise model:

$$y_t = \ell_{t-1} + \varepsilon_t \tag{7a}$$

$$\ell_t = \ell_{t-1} + \alpha\varepsilon_t \tag{7b}$$

Solving the first equation for ε_t and substituting in the second one gives

$$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1} \tag{8}$$

which can be further worked out to $\ell_t = (1-\alpha)^t \ell_0 + \alpha \sum_{j=0}^{t-1}(1-\alpha)^j y_{t-j}$. The consecutive values of the ℓ_t are thus determined by a starting value ℓ_0 and the observations $y_1, y_2, \ldots$. This means that once we have an estimate for ℓ_0, namely $\hat{\ell}_0$, subsequent estimates $\hat{\ell}_1, \hat{\ell}_2, \ldots$ can easily be computed every period as new observations become available. In other words, an estimate for ℓ_t can be computed by taking its conditional expectation, given the choice of the starting value and parameters, and the observations:

$$\hat{\ell}_t = E(\ell_t | \hat{\ell}_0, \alpha, y_1, \ldots, y_t) \tag{9}$$

Since all necessary quantities are known, no actual expectation has to be taken and $\hat{\ell}_t$ can simply be computed using recursion (8). In this way, we get minimum mean square error estimates of the ℓ_t's, conditional on $\hat{\ell}_0$. A minimum mean square error forecast is then obtained by taking the conditional expectation of (7a), $E(y_{t+h}|\hat{\ell}_0, \alpha, y_1, \ldots, y_t) = \hat{\ell}_t$.

This idea extends to the model for GSI. The error-correction form in (5a)-(5c) is the equivalent of that for the random walk plus noise model in (8). Again, once starting values $\hat{\ell}_{i,0}$, $\hat{b}_{i,0}$ and $\hat{s}_{-m+1}, \ldots, \hat{s}_0$ have been provided, subsequent estimates of $\ell_{i,t}$, $b_{i,t}$ and s_t can be calculated as observations become available. The forecast function is equal to $(\hat{\ell}_{i,t} + h\hat{b}_{i,t})\hat{s}_{t+h-m}$. The initial states and the smoothing parameters can be estimated by for example using a maximum likelihood approach, or by minimizing some other criterion such as MSE.

The method (6a)-(6c) resulting from model (4a)-(4d) is a generalization of the Holt-Winters procedure (3a)-(3d). The updating equations for level and trend are the same as those for HW. The updating equation for the seasonal component, however, makes use of all time series $i = 1, \ldots, N$. It updates the previous estimate $\hat{s}_{t-m}$ by weighting it with a new estimate $\sum_{i=1}^{N} \frac{w_i y_{i,t}}{\hat{\ell}_{i,t-1}+\hat{b}_{i,t-1}}$. This new estimate is a weighted average of N estimates obtained from all time series independently. By pooling these estimates using weights w_i, the forecast accuracy can be improved.

The weights w_i can be chosen to minimize forecast errors, measured by for example MSE, or can be specified in a variety of other ways. For example, taking $w_i = \frac{1}{N}$ gives equal weight to all error terms and thus all time series. Taking a simple average means all time series are considered to have the same amount of noise and thus get the same smoothing parameter $\gamma \frac{1}{N}$. If this is not the case, highly variable series may corrupt the estimated seasonal component. In general, noisier series should thus get lower weights. The lower the relative noise $\varepsilon_{i,t}$, the higher the weight w_i should be, and thus also the higher the smoothing parameter γw_i should be.

The variability of a time series can be measured by the variance of relative noise, equal to $Var(\frac{e_{i,t}}{(\ell_{i,t-1}+b_{i,t-1})s_{t-m}}) = Var(\varepsilon_{i,t}) = \sigma_i^2$. Weights could thus be taken to be $w_i = \frac{\sigma_i^{-2}}{\sum_{j=1}^{N} \sigma_j^{-2}}$, giving higher weights when there is less noise. We will use these latter weights in our simulations below.

A special case of the model arises when we make the weights time-dependent and take them equal to the proportion of a time series in the aggregate times series: $w_{i,t} = \frac{\ell_{i,t-1}+b_{i,t-1}}{\sum_{j=1}^{N}(\ell_{j,t-1}+b_{j,t-1})}$. This results in the updating equations for level and trend equal to (6a)-(6b), but seasonal equation (6c) replaced by

$$\hat{s}_t \quad = \quad \gamma \frac{\sum_{i=1}^{N} y_{i,t}}{\sum_{j=1}^{N} \hat{\ell}_{j,t-1} + \hat{b}_{j,t-1}} + (1-\gamma)\hat{s}_{t-m} \tag{10}$$

This choice of weights gives equal weight to all time series in a simple summation. In other words, seasonality is now estimated from aggregate data, while level and trend are estimated from disaggregate data.

Since model (4a)-(4d) allows the weights to be chosen in a variety of ways, it generalizes earlier GSI approaches. Dalhart (1974) computed seasonal indices separately for all N time series and then averaged them to get a composite estimate. This approach corresponds to setting $w_i = \frac{1}{N}$. Withycombe (1989) computed the seasonal indices from aggregate data, where demand was weighted by selling price per item p_i. The rationale for this is that we are more concerned with forecasts errors for higher valued items. This approach corresponds to setting

$$w_{i,t} = \frac{p_i(\ell_{i,t-1} + b_{i,t-1})}{\sum_{j=1}^{N} p_j(\ell_{j,t-1} + b_{j,t-1})} \tag{11}$$

giving seasonal updating equation

$$\hat{s}_t = \gamma \frac{\sum_{i=1}^{N} p_i y_{i,t}}{\sum_{j=1}^{N} p_j(\hat{\ell}_{j,t-1} + \hat{b}_{j,t-1})} + (1-\gamma)\hat{s}_{t-m} \tag{12}$$

Problems can arise if the time series that are grouped are not expressed in the same units of measurement. In practice, the unit of measurement in which a series is recorded is often arbitrary.For example, demand can be expressed in terms of single items, turnover, or packing sizes. If some time series are expressed in different units, different weights w_i result and this can determine the quality of the forecasts. One way of avoiding this problem is to express all time series in the same units. However, this is not always easy to do.

Another option is to ensure the model and method are unit-free and do not have this problem. This means that the unit of measurement in which the time series is expressed does not influence the outcome of the model and method, in particular the seasonal equations of both the model (4d) and the method (6c). These two equations contain the relative errors ($\varepsilon_{i,t}$) or $y_{i,t}/(\ell_{i,t-1} + b_{i,t-1})$ for all series, which are independent of the unit of measurement. The weights w_i, however, are not necessarily unit-free, causing the equations to be dependent on the unit of measurement. For example, $w_{i,t} = \frac{\ell_{i,t-1}+b_{i,t-1}}{\sum_{j=1}^{N}(\ell_{j,t-1}+b_{j,t-1})}$ or $w_i = \frac{p_i}{\sum_{j=1}^{N} p_j}$ are dependent on the unit of measurement of each of the series. On the other hand, $w_{i,t} = \frac{p_i(\ell_{i,t-1}+b_{i,t-1})}{\sum_{j=1}^{N} p_j(\ell_{j,t-1}+b_{j,t-1})}$, $w_i = \frac{1}{N}$ or $w_i = \frac{\sigma_i^{-2}}{\sum_{j=1}^{N} \sigma_j^{-2}}$ are unit-free weights. The latter option is unit-free since it depends on the relative errors, and is therefore used in the simulations below.

2.2 Estimation

Before we can make forecasts, we need estimates for initial values $\ell_{i,0}$, $b_{i,0}$ and $s_{-m+1}, \ldots, s_0$, and for parameters α_i, β_i, γ and w_i. These can be obtained by maximizing the conditional likelihood function or by minimizing some criterion that measures forecast errors such as the MSE. In Hyndman et al. (2002) several estimation methods were compared for fitting exponential smoothing state space models on data from the M3-competition, and minimizing the MSE was found on average to result in slightly more accurate forecasts than maximizing the likelihood. Although these ways of obtaining estimates may be feasible approaches for simpler models like the model underlying HW, the GSI model contains many parameters, and thus finding optimal values in such a high dimensional ($m + 1 + 5N$ dimensions) parameter space may be very time-consuming.

Instead of finding optimal estimates for all parameters and initial states by a nonlinear optimization algorithm, we can use a two-stage procedure and first obtain initial estimates of the states and then optimize the smoothing parameters. This is a common procedure for exponential smoothing methods and means that the smoothing parameters are optimized conditional on the values of the initial states.

A two-step heuristic solution divides the historical data (T periods) into two sections: an initialization period (I) and an optimization period (O), with $T = I + O$. Heuristic estimates for initial states are obtained using sample I, and a nonlinear optimization algorithm is used to find optimal parameter settings over sample O. There are several ways in which this can be done. Below, we describe the procedure that is used in the simulation experiments in the next section.

Classical decomposition by ratio-to-moving-averages (RTMA) is used to obtain estimates for ℓ_0, b_0 and $s_{-m+1}, \ldots, s_0$. For HW, we apply this procedure to each time series separately. For the GSI method, we also need estimates of the w_i. In the forecasting method, we use weights equal to $\hat{w}_i = \frac{\hat{\sigma}_i^{-2}}{\sum_{j=1}^{N} \hat{\sigma}_j^{-2}}$. To estimate the $\hat{\sigma}_i^2$, we fit a single source of error HW model to each time series separately, by applying the corresponding HW method. Although this method is not optimal if we assume the data follows the model underlying GSI, it still gives reasonably good estimates of $\varepsilon_{i,t}$. These fitted errors can be calculated using the smoothed state estimates and the observation equation: $\hat{\varepsilon}_{i,t} = \frac{y_{i,t} - (\hat{\ell}_{i,t} - \hat{b}_{i,t})\hat{s}_{t-m}}{(\hat{\ell}_{i,t} - \hat{b}_{i,t})\hat{s}_{t-m}}$.

This gives the following estimate for σ_i^2:

$$\hat{\sigma}_i^2 = \frac{\sum_{t=1}^{I} \hat{\varepsilon}_{i,t}^2}{I-3} = \frac{\sum_{t=1}^{I} \left(\frac{y_{i,t} - (\hat{\ell}_{i,t-1} + \hat{b}_{i,t-1})\hat{s}_{t-m}}{(\hat{\ell}_{i,t-1} + \hat{b}_{i,t-1})\hat{s}_{t-m}} \right)^2}{I-3} \tag{13}$$

We use [I-number of smoothing parameters] in the divisor, as suggested by Bowerman et al. (2005). Next, since the GSI model assumes the seasonal component is common to all time series, we use the estimates of w_i to find initial estimates of the common seasonal pattern by computing $\hat{s}_k = \sum_{i=1}^{N} \hat{w}_i \hat{s}_{i,k}$ for $k = -m+1, \dots, 0$.

After obtaining estimates for the initial states, each forecasting algorithm is run over I and O, so that the impact of poor initial estimates is offset during period I. The MSE is then calculated and minimized over period O, by applying a nonlinear optimization algorithm. The same smoothing parameters are used throughout both I and O, but only measured over O. The starting values of the smoothing parameters are all taken to be 0.5. Furthermore, they are constrained to $0 \leq \alpha_i, \beta_i, \gamma \leq 1$. Since for GSI all time series are interrelated via the common seasonality equation, the sum of MSE's is minimized, while for HW the MSE is optimized for each time series separately.

We normalize the seasonal indices after each update of the estimates, as argued in Archibald and Koehler (2003). Although the model does not make such an assumption on the seasonal indices, their interpretation would be lost if they do no longer sum up to m.

3 Simulation study

We are interested in determining in which situations the GSI method is more accurate than HW and how the forecast accuracy of both methods depends on parameter settings and time series characteristics. In this section, we describe a simulation study that investigates the properties of the GSI method. The reason for carrying out a simulation study is that obtaining analytical expressions for forecast accuracy and prediction intervals has proven to be difficult. Derivation of exact expressions like those that exist for

Holt-Winters (Hyndman et al., 2005) becomes mathematically intractable. Approximations like those in Koehler et al. (2001) assume that the seasonal component is unchanging in the future. Under this assumption, the GSI model yields the same expressions as for HW.

3.1 Parameter settings and data simulation

The model developed in the previous section allows random simulation of data for which the GSI method is optimal. However, here we generate data from a slightly modified model:

$$y_{i,t} = (\ell_{i,t-1} + b_{i,t-1}) s_{i,t-m} v_{i,t} \tag{14a}$$

$$\ell_{i,t} = (\ell_{i,t-1} + b_{i,t-1})(1 + \alpha_i (v_{i,t} - 1)) \tag{14b}$$

$$b_{i,t} = b_{i,t-1} + (\ell_{i,t-1} + b_{i,t-1}) \alpha_i \beta_i (v_{i,t} - 1) \tag{14c}$$

$$s_t = s_{t-m} + \gamma s_{t-m} \sum_{i=1}^{N} w_i (v_{i,t} - 1) \tag{14d}$$

$$s_{i,t} = s_t + d_{i,t} \tag{14e}$$

Firstly, this model replaces $1+\varepsilon_{i,t}$ by $v_{i,t}$. In model (4a)-(4d), the disturbances are assumed to be normally distributed. Since errors are multiplicative, this could result in negative time series values . Therefore, we use $v_{i,t} \sim \Gamma(a,b)$ with $a = 1/b$ and $b = \sigma_i^2$, so that $v_{i,t}$ has mean $ab = 1$ and variance $ab^2 = \sigma_i^2$. Although this reduces the probability of $y_{i,t}$ becoming negative or problems due to division by zero, these can still occur if $\ell_{i,t} + b_{i,t} \leq 0$. If this happens, $\ell_{i,t} + b_{i,t}$ is truncated and set equal to a small number.

Secondly, this model includes possible dissimilarity in seasonal patterns by letting the seasonal patterns of each of the time series $i = 1, \ldots, N$ have a deviation $d_{i,t}$ from the common pattern. If the seasonal patterns are not identical, GSI will no longer be optimal, and HW may give more accurate forecasts. For $d_{i,t} = 0$, the equations of this model, when put in error-correction form, show equivalence with the GSI method (and with HW for $N = 1$). The deviations $d_{i,t}$ are assumed to be deterministic and periodic, i.e. $d_{i,t+m} = d_{i,t}$. In this way, the extent of dissimilarity remains the same over time and is not affected by the noise processes. In this simulation study we determine how the magnitude of $d_{i,t}$ affects the performance of GSI relative to that of HW. Although the $d_{i,t}$

are deterministic, we assume they are initially drawn from a normal distribution. If we draw the $d_{i,t}$ from a normal distribution with mean zero and standard deviation σ_d, then about 95% of seasonal indices $s_{i,t}$ should be in the interval $(s_t - 2\sigma_d, s_t + 2\sigma_d)$. The value of σ_d thus determines a bandwidth within which the seasonal indices of all series in the group lie.

In the simulations, data processes and types of forecasts are varied. Below, we discuss some of the corresponding model and forecast parameter settings. All parameter settings are summarized in Tables 1 and 2. Many parameters are only scale parameters and thus their actual settings are not important. Their values relative to that of others are relevant. More precisely, the results are determined by the amount of noise relative to the underlying pattern. This determines the quality of the estimates and thus of the forecasts. For example, the value of the level $\ell_{i,t}$ of a time series is not relevant on its own, but the ratio between $\varepsilon_{i,t}$ and $\ell_{i,t}$ is. The same applies for the settings of other parameters and will be discussed below.

This allows for a reduction in the number of parameters to be considered, resulting in 1600 remaining combinations of parameter settings that are examined. Some parameters are varied systematically on a grid of values. Other parameters are varied randomly on a continuous interval (parameters are randomly drawn from this interval). For some parameters that are varied randomly, the interval from which they are drawn is varied systematically.

Table 1: *Parameters that remain fixed*

Parameter	Description	Level
m	Number of seasons	12
$\ell_{1,0}$	Initial level of first item	100
$c_{\max}$	Maximum initial trend	0.008
$s_{\{-m+1,\dots,0\}}$	Initial seasonal pattern	Sinusoid
A	Amplitude of seasonal pattern	0.2

Table 2: *Levels at which parameters are varied*

Parameter	Description	Level
Time series		
T	Length of historical data	$4m$, $6m$
$r_{\max}$		$1, 4$
r_i	Ratio between $\ell_{i,0}$ and $\ell_{1,0}$	$\in [1, r_{\max}]$
c_i	Ratio between $b_{i,0}$ and $\ell_{i,0}$	$\in [-c_{\max}, c_{\max}]$
α_i		$\in [0, 1]$
β_i		$\in [0, 1]$
γ		$\in [0, 1]$
w_i	Weights	$\in [0, 1]$
σ_i^2	Variance of noise	$\in [0, \sigma_{\max}^2]$
$\sigma_{\max}$		$0.01, 0.03, 0.05, 0.07$
Product groups		
N	Number of items in group	$2, 4, 8, 16, 32$
σ_d	Dissimilarity in seasonal patterns	$0, 0.01, 0.03, 0.05, 0.07$
Forecasting		
h	Forecast horizon	$1, 4, 8, 12$

Level and trend

In order to generate simulated data, seed values $\ell_{i,0}$, $b_{i,0}$ and $s_0, \dots, s_{-m+1}$ are specified, after which data from the model is generated for $t = 1, \dots, T$. The starting values of the time series are normalized by setting the level of the first time series equal to $\ell_{i=1,t=0} = 100$. The levels of the other series are set by specifying the ratio between the initial level $\ell_{i,0}$ of item i and that of the first item: $r_i = \frac{\ell_{i,0}}{\ell_{1,0}}$. The r_i are randomly drawn from $[1, r_{\max}]$ and used to set $\ell_{i,0} = r_i \ell_{1,0}$. The ratios are used for initial values when generating data. However, the levels evolve according to random walk processes, as well as due to trends. Especially under large amounts of noise or if trends move in opposite directions, the actual ratios may thus deviate from the r_i. For the regression analyses later on, we will therefore use an average ratio: $r_i' = (r_i + \frac{\ell_{i,T}}{\ell_{1,T}})/2$. Besides, to avoid using all the r_i's in a regression, we will summarize the group by regressing on μ_r and σ_r^2, which characterize the mean of r_i' and its variation among the group of items.

The initial growth rate $b_{i,0}$ is set by randomly drawing $c_i \in [-c_{\max}, c_{\max}]$ and setting $b_{i,0} = c_i \ell_{i,0}$, so that we end up with an initial trend between $-c_{\max} \cdot m \cdot 100$ and $c_{\max} \cdot m \cdot 100$

annually. Taking $c_{\max} = 0.008$, $m = 12$ and $\ell_{i,0} = 100$, this gives an initial trend between approximately -10 and $+10$. Only one value of $c_{\max}$ is considered since the performance of GSI does not depend on the trend, but on the amount of noise to which the trend is subject. Since the trend development is incorporated in the level, the trends are not used in the regression. The parameters α_i and β_i are drawn randomly from $(0, 1)$. Once they are drawn, they are summarized for regression purposes by computing the proportion of each parameters in the intervals $(0, 0.33)$, $(0.33, 0.67)$ or $(0.67, 1)$, denoted by $\alpha_0^{0.33}$, $\alpha_{0.33}^{0.67}$, $\alpha_{0.67}^{1.00}$, $\beta_0^{0.33}$, $\beta_{0.33}^{0.67}$ and $\beta_{0.67}^{1.00}$. Both the former three and the latter three thus sum to 1.

Seasonal patterns

We consider a single and fixed initial seasonal pattern $\{\hat{s}_{-m+1}, \dots, \hat{s}_0\}$, with $s_{j-m} = 1 + A \cdot \sin(2\pi j/m)$ for $j = 1, \dots, m$ and with $A = 0.2$. The smoothing parameter γ is drawn randomly from $(0, 1)$. The dissimilarity in seasonal patterns is modeled via parameters $d_{i,t}$, which are drawn from $N(0, \sigma_d^2)$. Thus, σ_d^2 determines the bandwidth within which all seasonal patterns of the group lie and therefore the extent of dissimilarity. If $\sigma_d^2 = 0$ the seasonal patterns are identical.

Weights

For data that are simulated, w_i' are drawn randomly from $[0, 1]$ and then normalized by taking $w_i = \frac{w_i'}{\sum_{i=1}^N w_i'}$, so that $\sum_{i=1}^N w_i = 1$. In the GSI method, weights w_i are chosen to be fixed and equal to $\frac{\sigma_i^{-2}}{\sum_{j=1}^N \sigma_j^{-2}}$.

Noise

We choose noise variances (σ_i^2) on the interval $[0, \sigma^2_{\text{max}}]$. The noisiness of series is determined by the noise process in combination with parameters α, β and γ. These determine the signal-to-noise ratios for the level, trend and seasonals. The value of σ_i above gives information on the volatility of the time series under perfect information about the trend-seasonal cycle, but σ_i in combination with parameters α, β and γ determines the total variation in the series.

3.2 Accuracy measurement

The two-step heuristic solution from the previous section now divides the historical data (T periods) into three sections instead of two: an initialization period (I), an optimization period (O), and a hold-out sample (H), with $T = I + O + H$. Heuristic estimates for initial states are obtained using sample I, and a nonlinear optimization algorithm is used to find optimal parameter settings over sample O. We use $2m$ periods for obtaining initial estimates, and $1m$ and $3m$ (for $T = 4m$ and $T = 6m$ respectively) periods for estimating the smoothing parameters. After optimization, the algorithm is run over the last m periods and accuracy measures are calculated.

For each combination of parameter settings, multiple replications are carried out. We continue making replications until we have found an approximate 95% confidence interval for our accuracy measure. Then, our estimate of the accuracy measure has a relative error of at most 0.1 with a probability of approximately 0.95. In 99% of the cases, less than 200 replications were needed.

From a computational perspective, we want a measure that is insensitive to outliers, negative or (close to) zero values. For these reasons, many measures of accuracy for univariate forecasts are inadequate, since they may be infinite, skewed or undefined, and can produce misleading results (Makridakis and Hibon, 1995; Hyndman and Koehler, 2006). Even the most common measures have drawbacks.

Based on these considerations, we opt for the use of the following accuracy measure. Assume that historical data is available for periods $t = 1, \ldots, T$, and that forecasts are made for $t = T + j$, $j = 1, \ldots, H$, with H the length of the hold-out sample. Based on an evaluation of all common types of accuracy measures, Hyndman and Koehler (2006) advocates the use of scaled errors, where the error is scaled based on the in-sample MAD from a benchmark method. If we take the Naïve method (see e.g., Makridakis et al., 1998) as a benchmark, the scaled error is defined as

$$q_{i,t} = \frac{y_{i,t} - \hat{y}_{i,t}}{\frac{1}{T-1}\sum_{t=2}^{T} |y_{i,t} - y_{i,t-1}|} \tag{15}$$

with $y_{i,t}$ the time series value and $\hat{y}_{i,t}$ its forecast. The Mean Absolute Scaled Error is then simply the mean of $|q_{i,t}|$. In our case we take the mean over the hold-out sample and over all time series:

$$\text{MASE} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{H}\sum_{t=T+1}^{T+H}|q_{i,t}| = \frac{1}{N}\sum_{i=1}^{N}\frac{\text{MAD}_i}{\text{MAD}_{b,i}} \tag{16}$$

with MAD_i the out-of-sample MAD of time series i and $\text{MAD}_{b,i}$ the in-sample MAD of the benchmark method for time series i. Hyndman and Koehler (2006) recommends using the MASE for comparing accuracy across multiple time series, since it resolves all arithmetic issues mentioned above and is thus suitable for all situations.

4 Results

We compare results on the relative accuracy of GSI compared to HW:

$$\text{MASE}^{\text{GSI}}_{\text{HW}} = \frac{\text{MASE}_{\text{GSI}}}{\text{MASE}_{\text{HW}}} \tag{17}$$

We first determine which of the parameters have the strongest impact on the accuracy improvement. We do this by using the analysis of variance procedure (ANOVA). This analysis tells us if distinct levels of each of the parameters result in significantly different values of the accuracy improvement. For example, if at different values of N the values of $\text{MASE}^{\text{GSI}}_{\text{HW}}$ are significantly different, the accuracy improvement depends on the values of N significantly.

The analysis of variance procedure allows us to see which parameters explain the variation in results most in a clear way, and get a good estimate of interaction effects. An interaction effect is the extra effect due to combining explanatory variables that cannot be predicted by knowing the the effects of the variables separately. Significant interaction effects indicate that a combination of explanatory variables is particularly effective. Based on the results in the previous section, there seem to be several interaction effects. For our simulation this means that simply optimizing each parameter separately does not necessarily lead to the combination of parameter settings with the lowest value of $\text{MASE}^{\text{GSI}}_{\text{HW}}$.

Table 3 presents the results for an ANOVA of the simulation results. It includes the input parameters for the simulation (N, T, $r_{\max}$, h, σ_d and $\sigma_{\max}$), as well as the parameters that describe the group of time series in more detail (μ_r, σ_r^2, $\alpha_0^{0.33}$, $\alpha_{0.33}^{0.67}$, $\alpha_{0.67}^{1.00}$, $\beta_0^{0.33}$, $\beta_{0.33}^{0.67}$, $\beta_{0.67}^{1.00}$

Table 3: *ANOVA for simulation results*

Source of variation	*SS*[1]	*df*[2]	*MSS*[3]	F	P
N	64819	1	64819	170260.0	< 0.0001
T	51803	1	51803	136070.0	< 0.0001
$r_{\max}$	583	1	583	1530.4	< 0.0001
σ_d	5167	1	5167	13572.0	< 0.0001
$\sigma_{\max}$	1161	1	1161	3049.6	< 0.0001
h	94	1	94	247.0	< 0.0001
μ_r	0.21	1	0.21	0.5	0.46
σ_r^2	0.12	1	0.12	0.3	0.58
$\alpha_0^{0.33}$	519	1	519	1364.0	< 0.0001
$\alpha_{0.33}^{0.67}$	635	1	635	1668.8	< 0.0001
$\alpha_{0.67}^{1.00}$	1703	1	1703	4474.2	< 0.0001
$\beta_0^{0.33}$	51	1	51	134.5	< 0.0001
$\beta_{0.33}^{0.67}$	34	1	34	88.5	< 0.0001
γ	0.13	1	0.13	0.3	0.56
$I(N,T)^*$	62	1	62	164.0	< 0.0001
$I(N,\sigma_d)$	204	1	204	536.7	< 0.0001
$I(T,\sigma_d)$	139	1	139	364.5	< 0.0001
$I(h,\sigma_d)$	209	1	209	549.4	< 0.0001
$I(\sigma_d,\sigma_{\max})$	61	1	61	161.2	< 0.0001
$I(N,\sigma_{\max})$	362	1	362	951.3	< 0.0001
$I(T,\sigma_{\max})$	213	1	213	560.1	< 0.0001
Residuals	29838	78379	0.38		
Total	157658	78400			

[1] *SS* : sum of squares
[2] *df* : degrees of freedom
[3] *MSS* : mean sum of squares
* $I(a,b)$: interaction effect between a and b

and γ). The latter were calculated from the set of time series that were simulated using the input parameters. In addition, we included the most important interaction effects, i.e. those interaction effects that were both significant and were larger than 50.

Firstly, it appears that N and T show large values of their respective sum of squares (SS). This is not surprising, since they take on larger values than the other parameters. If we take into account the relative magnitudes of the parameters, in particular the SS values of σ_d and $\sigma_{\max}$ are large. This corresponds to the findings in the previous section, where these parameters were found to be important determinants of accuracy improvement. The SS value of h on the other hand is relatively small, indicating that benefits of GSI do not apply to only a particular forecast horizon.

In Table 3, the parameter $\beta_{0.67}^{1.00}$ is omitted. The reason is that at least one of the variables

$\alpha_0^{0.33}$, $\alpha_{0.33}^{0.67}$, $\alpha_{0.67}^{1.00}$, $\beta_0^{0.33}$, $\beta_{0.33}^{0.67}$ or $\beta_{0.67}^{1.00}$ has to be removed from the analysis to avoid singularities (since both $\alpha_0^{0.33} + \alpha_{0.33}^{0.67} + \alpha_{0.67}^{1.00} = 1$ and $\beta_0^{0.33} + \beta_{0.33}^{0.67} + \beta_{0.67}^{1.00} = 1$ and this ANOVA has no intercept). The values of $\beta_0^{0.33}$, $\beta_{0.33}^{0.67}$ or $\beta_{0.67}^{1.00}$ are much lower than those of $\alpha_0^{0.33}$, $\alpha_{0.33}^{0.67}$ and $\alpha_{0.67}^{1.00}$, but would be comparable if for example $\alpha_{0.67}^{1.00}$ was omitted instead of $\beta_{0.67}^{1.00}$. In either case, it appears that if a group has a larger portion of time series with high α_i's or β_i's, the accuracy improvement show more variability. Since also $r_{\max}$ has a high SS value, this shows that groups of time series with more heterogeneous demand patterns show more variation in accuracy improvement. This is not surprising, since more in those cases, the time series are more variable, and thus more difficult to forecast.

For most of the parameters the effects on the results are clearly significant, with very large F-values. However, three parameters do not have a significant effect on the accuracy improvement: μ_r, σ_r^2 and γ. In contrast with the values of the α_i's or β_i's, the value of γ appears to have no clear influence on the accuracy improvement.

Next, we consider the interaction effects. Especially the values of σ_d and $\sigma_{\max}$ in combination with several other parameters can yield an additional accuracy improvement. Especially combination including the input parameters N, T and h appear to explain a significant part of the variation in $\text{MASE}_{\text{HW}}^{\text{GSI}}$.

Now we know which parameters explain most of the variance in the accuracy improvement, let us be more precise about the impact of certain parameters. To investigate the relation between accuracy and parameter settings, we fit a multiple linear regression model by least squares. After examining several regression models, and restricting ourselves to linear models, the following model without intercept appears to give a good fit:

$$\begin{aligned}\text{MASE}_{\text{HW}}^{\text{GSI}} &= \theta_1 N + \theta_2 T + \theta_3 r_{\max} + \theta_4 h + \theta_5 \sigma_d + \theta_6 \sigma_{\max} \\ &\quad + \theta_7 \alpha_0^{0.33} + \theta_8 \alpha_{0.33}^{0.67} + \theta_9 \alpha_{0.67}^{1.00} \\ &\quad + \theta_{10} \beta_0^{0.33} + \theta_{11} \beta_{0.33}^{0.67} + \xi_i \qquad (18)\end{aligned}$$

Table 4 contains the results for this regression. In this regression, we have left out the parameters found insignificant in the analysis of variance, as well as the interaction effects. All parameters have very large t-values and are clearly significant, in particular σ_d and $\sigma_{\max}$. The estimates of the regression coefficients are in line with the results in the previous section and indicate that larger accuracy improvements of GSI over HW can be achieved in case of:

- more similarity in seasonal patterns (smaller values of σ_d)
- larger amounts of noise (larger $\sigma_{\max}$)
- larger groups (larger N)
- longer forecast horizons (larger h)
- less historical data (smaller T)
- more homogeneous groups (smaller $r_{\max}$)

As noted above, one parameter is omitted from the regression. If we leave out $\alpha_{0.67}^{1.00}$, the parameter estimates of $\beta_0^{0.33}$, $\beta_{0.33}^{0.67}$ and $\beta_{0.67}^{1.00}$ are 1.10, 1.03 and 0.92, respectively. If we leave out $\beta_{0.67}^{1.00}$, the estimates of $\alpha_0^{0.33}$, $\alpha_{0.33}^{0.67}$ and $\alpha_{0.67}^{1.00}$ are 1.15, 1.05 and 0.92, respectively (see Table 4). The values of $\beta_0^{0.33}$, $\beta_{0.33}^{0.67}$ in the regression are thus relative to the value of $\beta_{0.67}^{1.00}$. In both cases this shows that GSI especially benefits from a group of time series where a large portion of the series have low values for α_i and β_i, i.e. slowly evolving series. In combination with the result that a low value of $r_{\max}$ results in a lower $\text{MASE}_{\text{HW}}^{\text{GSI}}$, this supports literature that hierarchical approaches are more suitable for homogeneous groups.

Table 4: *Results for regression of $MASE_{HW}^{GSI}$ on parameters (eq. 18)*

Variable	*Coefficient*	*t*	*P*
N	-0.0024	-11.6	< 0.0001
T	0.0037	19.9	< 0.0001
$r_{\max}$	0.0183	12.2	< 0.0001
σ_d	8.6863	98.9	< 0.0001
$\sigma_{\max}$	-7.5374	-75.0	< 0.0001
h	-0.0192	-35.4	< 0.0001
$\alpha_0^{0.33}$	1.1515	67.7	< 0.0001
$\alpha_{0.33}^{0.67}$	1.0467	61.0	< 0.0001
$\alpha_{0.67}^{1.00}$	0.9175	54.0	< 0.0001
$\beta_0^{0.33}$	0.1802	14.5	< 0.0001
$\beta_{0.33}^{0.67}$	0.1141	9.2	< 0.0001
$R^2 = 0.81$			
F-statistic = 29010, P-value = < 0.0001			

We now look at some results in more detail. Table 5 presents results for the difference in the seasonal patterns and the amount of noise present in the group in more detail. The results are broken down with respect to the number of series in a group and averaged

Table 5: *Relative accuracy of GSI compared to HW at various levels of N, σ_d and σ_{max}*

		σ_{max}			
N	σ_d	0.01	0.03	0.05	0.07
2	0	**0.99**	**0.99**	**0.91**	**0.99**
2	0.01	1.33	1.18	1.10	1.04
2	0.03	1.46	1.31	1.26	1.13
2	0.05	1.47	1.35	1.28	1.18
2	0.07	1.44	1.35	1.33	1.33
4	0	**0.94**	**0.94**	**0.89**	**0.88**
4	0.01	1.43	1.21	1.10	1.00
4	0.03	1.61	1.38	1.29	1.22
4	0.05	1.65	1.41	1.38	1.27
4	0.07	1.62	1.52	1.38	1.36
8	0	**0.89**	**0.83**	**0.84**	**0.78**
8	0.01	1.48	1.11	**0.99**	**0.89**
8	0.03	1.69	1.42	1.29	1.13
8	0.05	1.79	1.51	1.31	1.29
8	0.07	1.78	1.61	1.45	1.32
16	0	**0.86**	**0.74**	**0.67**	**0.68**
16	0.01	1.47	1.06	**0.86**	**0.77**
16	0.03	1.78	1.48	1.20	1.02
16	0.05	1.83	1.50	1.31	1.18
16	0.07	1.88	1.59	1.40	1.16
32	0	**0.82**	**0.64**	**0.61**	**0.52**
32	0.01	1.54	1.02	**0.82**	**0.69**
32	0.03	1.82	1.32	1.06	**0.91**
32	0.05	1.90	1.56	1.24	1.01
32	0.07	1.93	1.60	1.35	1.19

over all other parameters. The cases where $\text{MASE}_{\text{HW}}^{\text{GSI}} < 1$ are indicated by bold-faced numbers. The results show that as the amount of noise increases, the accuracy of GSI relative to that of HW increases. Furthermore, the seasonal patterns need not be identical for GSI to improve on HW. If there is enough noise they can be nonidentical to some extent. However, as the dissimilarity increases, the improvement diminishes. If the patterns become too dissimilar, HW performs better in all cases.

Besides this, we see that the average improvement increases as the product groups get larger. If the number of time series increases, the same amount of noise and dissimilarity gives larger improvements. The results for small groups are less pronounced since the results show more variability. With small groups there is a larger chance of substantial differences in accuracy of both methods. For larger groups, the accuracy is averaged over more time series and thus more stable. From Table 5, it appears that in case of small product groups and little noise, it is more likely that HW applied to each series separately results in better forecasts, while for large product groups and substantial amounts of noise, GSI is more likely to perform better.

The level of σ_d (the difference in the seasonal patterns) relative to that of $\sigma_{\max}$ (the maximum amount of noise in the group) has a significant impact on the results. In Figure 1 the results for $\log(\sigma_d/\sigma_{\max})$ are plotted, which shows a clear relationship between $\text{MASE}_{\text{HW}}^{\text{GSI}}$ and the value of $\sigma_d/\sigma_{\max}$. The lower the value of $\sigma_d/\sigma_{\max}$, the larger the accuracy improvement, in particular for larger values of N.

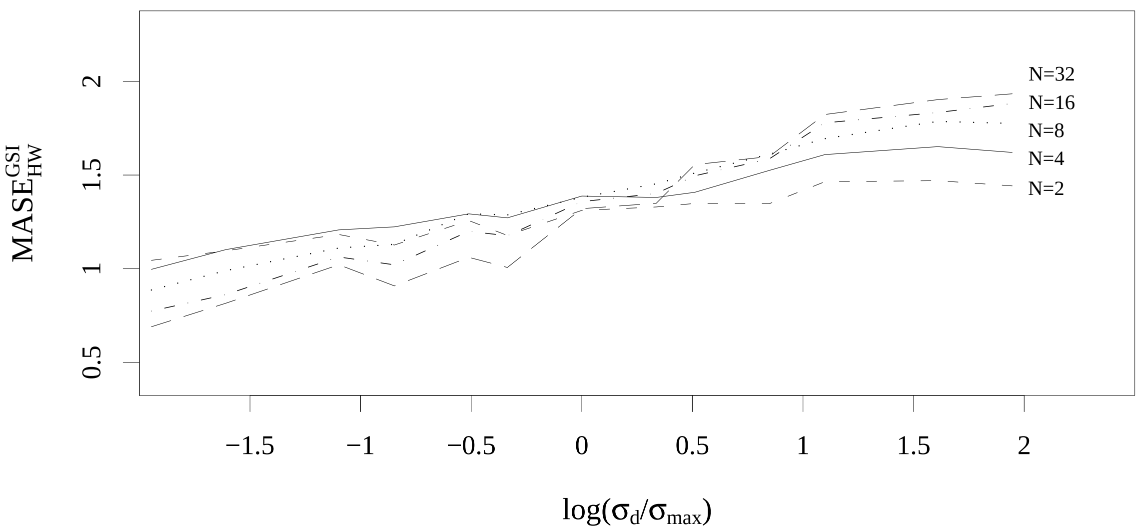

Figure 1: *$MASE_{HW}^{GSI}$ as a function of $log(\sigma_d/\sigma_{max})$*

Since GSI can improve on HW even if the seasonal patterns are not identical, the method is robust to deviations from one of its major assumptions. GSI can also be considered to be more robust than HW if it is less sensitive to outliers. In Table 6, the mean and standard deviation of MASE for both GSI and HW are shown, calculated over all parameter values. As the amount of noise increases, the chance of outliers occurring also increases. Especially for HW, for larger amounts of noise, the mean and standard deviation go up significantly. GSI is thus more robust than HW since the accuracy measures are lower and show less variation.

Table 6: *Robustness results for GSI and HW*

		MASE	
Method	$\sigma_{\max}$	**Mean**	**St.dev.**
GSI	0.01	1.00	0.62
	0.03	1.35	0.90
	0.05	1.60	1.12
	0.07	1.80	1.29
HW	0.01	2.49	55.14
	0.03	16.11	154.89
	0.05	25.91	192.83
	0.07	37.43	226.54

5 Conclusions

This paper presents an approach to making more accurate demand forecasts by using data from related items. Forecast accuracy at the item level can be improved by simultaneously forecasting a group of items sharing a common seasonal pattern. In practice, we would need to find such a group of products, but here we can make use of the hierarchical nature of data within companies. The GSI method and model are generalizations of the Holt-Winters method and its underlying model. An advantage over Holt-Winters is that less historical data is needed to obtain good estimates. Instead of using e.g., a ratio-to-moving-average procedure on several years of data, we now only need one year, taken across several series, to obtain estimates.

A contribution of this article is that it has not only generalized the Holt-Winters method, but also earlier group seasonality methods. The methods in Dalhart (1974) and Withycombe (1989) used aggregation to improve forecast accuracy. We have generalized this to a system with weights, where the former methods are now special cases. Furthermore, instead of empirical comparisons, we have presented a theoretical framework for group seasonality approaches.

In the simulation study, we determined the forecast accuracy of each method for a large number of parameter settings. The main results are that GSI performs better than HW if there is more similarity in seasonal patterns, for larger amounts of noise, for larger and more homogeneous groups, for longer forecast horizons, and with less historical data. There is a clear relationship between the similarity in seasonal patterns and the random variation. Furthermore, it appears that seasonal patterns need not be identical for GSI to give more accurate forecasts than Holt-Winters. If there is substantial random variation compared to the variation among the patterns, Holt-Winters performs poorly, and it becomes beneficial to apply GSI, even if seasonal patterns are somewhat different. This also improves the practical applicability, since in practice it is not very likely that a group of products exhibits exactly the same seasonal behavior, or at least it will be impossible to get an accurate estimate of this.

With these simulation results, we can check, for a given group of items, whether we can expect to obtain a more accurate forecast by GSI than by HW, and indicate how large the forecast error of both methods will be. Based on this, items can be grouped in order to generate more accurate forecasts by the GSI method.

References

Archibald, B.C. and Koehler, A.B. (2003). Normalization of seasonal factors in Winters' methods. *International Journal of Forecasting*, 19(1), 143–148.

Bowerman, B.L., O'Connell, R.T., and Koehler, A.B. (2005). *Forecasting, Time Series, and Regresssion: An Applied Approach*, 4th edition. Thomson Brooks/Cole, Belmont, CA.

Bunn, D.W. and Vassilopoulos, A.I. (1993). Using group seasonal indices in multi-item short-term forecasting. *International Journal of Forecasting*, 9(4), 517–526.

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

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

Dalhart, G. (1974). Class seasonality - a new approach. American Production and Inventory Control Society 1974 Conference Proceedings, reprinted in *Forecasting*, 2nd ed. (APICS, Washington, DC), 11-16.

Dekker, M., van Donselaar, K.H., and Ouwehand, P. (2004). How to use aggregation and combined forecasting to improve seasonal demand forecasts. *International Journal of Production Economics*, 90(2), 151–167.

Holt, C.C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages. Office of Naval Research Memorandum No. 52, Carnegie Institute of Technology, Pittsburgh, Pennsylvania. Published in *International Journal of Forecasting* (2004), 20(1), 5-10.

Hyndman, R.J. and Koehler, A.B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688.

Hyndman, R.J., Koehler, A.B., Ord, J.K., and Snyder, R.D. (2005). Prediction intervals for exponential smoothing using two new classes of state space models. *Journal of Forecasting*, 24(1), 17–37.

Hyndman, R.J., Koehler, A.B., Snyder, R.D., and Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3), 439–454.

Koehler, A.B., Snyder, R.D., and Ord, J.K. (2001). Forecasting models and prediction intervals for the multiplicative Holt-Winters method. *International Journal of Forecasting*, 17(2), 269–286.

Makridakis, S. and Hibon, M. (1995). Evaluating accuracy (or error) measures. *Working paper 95/18*. INSEAD, Fontainebleau, France.

Makridakis, S., Wheelwright, S.C., and Hyndman, R.J. (1998). *Forecasting: methods and applications*, 3rd edition. John Wiley, New York.

Ord, J.K. and Koehler, A.B. (1990). A structural model for the multiplicative Holt-Winters method. *Proceedings of the Decision Sciences Institute*, 555-557.

Ord, J.K., Koehler, A.B., and Snyder, R.D. (1997). Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association*, 92(440), 1621–1629.

Ouwehand, P., van Donselaar, K.H., and de Kok, A.G. (2004). The impact of the forecasting horizon when forecasting with group seasonal indices. *Working paper 162*. Beta research school, Technische Universiteit Eindhoven, the Netherlands.

Snyder, R.D. (1985). Recursive estimation of dynamic linear models. *Journal of the Royal Statistical Society, series B*, 47(2), 272–276.

Winters, P.R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3), 324–342.

Withycombe, R. (1989). Forecasting with combined seasonal indices. *International Journal of Forecasting*, 5(4), 547–552.