Simulation Approach to Improve Performance of the Seasonal Time series Decomposition Method

Shamsuddin Ahmed

Bang College of Business
Kazakhstan Institute of Management Economics and Strategic Research, Almaty,
and
Edith Cowan University, Perth, Australia.
Email: Dr_Shamsuddin_Ahmed@Yahoo.Com

Abstract

A simulation method to model seasonal time series is developed. Using the ratio of seasonal data to nonlinear periodic trend, simulation parameters of the seasonal indices are identified. A decomposed seasonal time series is re-seasonalized to identify the combined random as well as cyclical indices in the seasonal model. These indices are modeled as a multi channel queue system. The seasonal time series is constructed as a multiplicative decomposition simulation model. To improve performance of the time series, the simulation model identifies the best seasonal indices and combined cyclical and random indices. These indices are updated whenever the simulation search improves squared error criteria of the fitted model. The simulated optimized seasonal, random and cyclical indices are used finally to re-seasonalize the data for forecast evaluation. The software is developed to test the performance of the approach using airline travel data available in Box, Jenkins and Reinsel (1994). The proposed simulation method achieves 38.89% improvement in squared error measure against the X-11 decomposition method.

Key words: simulation, seasonal, time series, indices, multiplicative, decomposition

Notations and definitions

n = Total number of data in time series
t = Number of seasons
$i = 1,2,\ldots,n$, index associated with seasonal data
$j = 1,2,\ldots,t$, index associated with season number
$\frac{n}{t}$ = Number of periods in time series data
$k = 1,2,\ldots,\frac{n}{t}$, index associated with periods in data
a_l = coefficient in equations representing trend in data
$l = 1,2,\ldots,m$, index associated with trend equation coefficients
m = total number of coefficients in trend equation
y_i = time series data at time i
$\bar{y}_i$ = estimate of time series data at time i
Ψ_k = trend component in cycle k
$\overline{\Psi}_k$ = estimate of trend component in cycle k
CR_{jk} = combined cyclical and random component in season j and cycle k
Φ_{jk} = seasonal component in season j and cycle k
Φ^*_{jk} = best seasonal component in season j and cycle k
CR^*_{jk} = best combined cyclical and random component in season j and cycle k
λ = arrival rate of entity in queue model
μ = service rate of server in system

1. Introduction

Seasonality in time series is an approximate cyclical pattern with a time period and varying in amplitude. The exogenous factors that influence seasonality can be attributed to weather variations, new technological developments, major calendar events, such as festive seasons and other regular social events. Economic growth or recession in economy contributes to longer cycle cyclical pattern that is different from seasonal pattern. In the multiplicative decomposition method a time series is decomposed into several components [4]. Each component is treated separately and recomposes the time series as the product of these identified components.

Seasonal indices, Φ_{jk}, and combined cyclical and random indices, CR_{jk}, are represented as multi channel queuing system. They are modeled as entity and arrive into the simulation system in bulk. The entities form a queue that has infinite queue capacity. The entities are assigned FIFO processing rule. The arrivals of customer in each channel follow a distribution that is less than the service rate hence, forming a stable queue system. Consequently, the arrival rate, λ, and service rate, μ, are such that the inequality $\lambda \leq \mu$ exists.

Depending on the number of seasons, t, the server calls t number of entities from queue containing Φ_{jk} entities. Similar numbers of entities from CR_{jk} queue are also called for service. They are placed in buffer queues. The server provides them service sequentially as shown in figure 1. The service time of these entities are defined by a uniform distribution. The parameters of the distribution are the value of the indices. The service time distribution is determined from the value of Φ_{jk} and CR_{jk} from equation (4) and (5) as shown in section 3. The service time is taken as uniform distribution with upper and lower limits as the maximum and minimum value found in Φ_{jk} and CR_{jk}.

Each customer in the queue waits for service to be provided by a single server in the system. When the server is free a customer is served and it leaves the system. The entities in buffer are served by FIFO queue processing rule in the order of the seasons. This rule ensures that the indices in a particular time period are processed one after another. For example, in a particular period the first quarter entities are processed first, then the second quarter entities and so on. After getting required services they leave the system. The processing value of the entities provided by the server is equivalent to the reconstruction of time series at that instant. The server calls the next batch of entities only when all the entities in buffer have finished their services. The fresh set of entities is drawn into buffer queue for service to be provided by the server.

Figure 3 displays the fitted model and clearly indicate a very close fit to the original data set. The simulated optimized Φ^*_{jk} are identified for example as 0.8542234, 1.096305 and 0.9305162 corresponding to $t = 1, 6 \text{ and } 12$. In case of CR^*_{jk} the related values for $t = 1, 6 \text{ and } 12$ are 1.000445, 1.014137 and 0.9974069 respectively.

2. Decomposition Method

One way to deal with seasonal data is to apply the X-11 filter [3], developed by Shiskin [7] while working with US census bureau. He attributes the method to papers by Persons [5,6]. The X-11 decomposition method deseasonalizes data using seasonal index obtained from a ratio of data to a moving average or a centred moving average. Using trend component and seasonal index, the data is reseasonalised and forecast is evaluated by extrapolation. In X-11 filter, data should be large to achieve a reasonable level of accuracy. Itigg [2] presents a multiplicative method that determines seasonal indices using decision variables. The principle upon which the method is derived suffers from difficulties when trend in data does not follow compounded growth.

Stationary time series are used in control theory for predicting process using autoregressive (AR), moving average (MA) or integrated autoregressive moving average (ARMA) and mixed integrated autoregressive moving average (ARIMA) models [1]. The process to be predicted is assumed stationary Gaussian. Todorescu [8] showed that a stationary time series could be decomposed into several components by means of an optimized parameter such that the first component is relatively smooth process having a large autocorrelation in comparison with the original time series. The successive components thus derived contained magnitudes of autocorrelations in descending order. This study showed only a few components is needed to approximate the time series. No assumption is made concerning the distribution of the process. The forecast procedure is applicable for computing predictors under relaxed conditions as compared to the currently used adaptive or minimum variance techniques.

3. Seasonal Indices Calculation

Equation 1 is the recomposition of seasonal data using three components $\overline{\Psi}_{k}$, Φ_{jk} and CR_{jk} as $\bar{y}_i$. The quantity Ψ_k in Equation 2 is the accumulated data for all seasons $j = 1,2,.....,t$, in period k. The estimate of Ψ_k is determined in $\overline{\Psi}_{k}$. Equation 3 models Ψ_k and accounts for nonlinear trend in data. Seasonal indices Φ_{jk} in Equation 4 are computed as the ratio of seasonal data y_i to the estimated trend $\overline{\Psi}_{k}$ at period k. The long term cyclical components and random components in seasonal data are determined concurrently in Equation 5 as the ratio of accumulated periodic seasonal data y_i at time $i = i+1$, $\boldsymbol{i}+2,\ldots,\boldsymbol{i}+\boldsymbol{t}$, to the reseasonalized data $(\Phi_{jk}\overline{\Psi}_k)$.

$$\bar{y}_i = \overline{\Psi}_k \Phi_{jk} CR_{jk} \qquad (1)$$

$$\Psi_k = \sum_{i=t(k-1)+1}^{tk} y_i \qquad (2)$$

$$\overline{\Psi}_k = \sum_{l=0}^{m} \frac{\alpha_l (k)^l}{l!}, k = 1,2,\ldots, \tfrac{n}{t} \qquad (3)$$

$$\Phi_{jk} = \frac{y_{t(k-1)+j}}{\overline{\Psi}_k} \qquad (4)$$

$$CR_{jk} = \frac{\sum_{i=t(k-1)+1}^{tk} y_i}{\Phi_{jk}\overline{\Psi}_k}, \; j = 1,2,\ldots,t \qquad (5)$$

4. Modeling Seasonal Indices as Queuing Systems

The notation $\{\Phi_{jk}\}$ and $\{CR_{jk}\}$ represents sequence of entities arriving into the queue systems from two queue Q1 and Q2 as shown in Figure 1. Server s in the system calls t number of entities waiting in queue Q1 and place them in buffer queue B1. The service follows FIFO rule. The t group of entity is equivalent to one time period data at k consisting of t number of seasonal data indexed $i = t(k-1)+j$, $j = 1,2,\ldots,t$, $k = 1,2,\ldots,\frac{n}{t}$. The service time that is required by these $\{\Phi_{jk}\}$ entities is determined from Equation 4. In quarterly time series the number of these entities $\{\Phi_{jk}\} \equiv (\Phi_{11}, \Phi_{21}, \Phi_{31}, \Phi_{41})$ for $k = 1$ and $t = 4$ indices they are associated with (y_1, y_2, y_3, y_4) data. Similarly $\{\Phi_{jk}\} \equiv (\Phi_{12}, \Phi_{22}, \Phi_{32}, \Phi_{42})$ are corresponding to (y_5, y_6, y_7, y_8) data. The system contains total n number of entities consisting of $\{\Phi_{jk}\}$. The order in which the entities arrive determines their corresponding position in season. For example Φ_{31} represents the entity at quarter 3 in time period 1. Similarly Φ_{12} represents the entity at quarter 1 in time period 2. There are total n numbers of $\{\Phi_{jk}\}$ entities in queue Q1. The service times of these entities $\{\Phi_{jk}\}$ are assigned at the time they join the queue based on parameters determined in Equation 4. The $\{CR_{jk}\}$ entities are also generated into the system similarly as in the case of $\{\Phi_{jk}\}$. Also Q2 contains n number of $\{CR_{jk}\}$ entities.

The server s calls t number of entities from Q1 containing $\{\Phi_{jk}\}$ and calls t number of entities from Q2 containing $\{CR_{jk}\}$ on FIFO basis. They await service to be provided by server s in buffer queue B1 and B2. The server process two entities from the buffer queue at a time, and the processing value on these entities is determined in Equation 1.

The entities leave the system after finishing their service. All the entities are processed from buffer queue until the buffer is exhausted. When services to the entities have been completed the next set of t entities from Q1 and t entities from Q2 are drawn into buffer queue B1 and B2. The processing continues until all the n entities from Q1 and all the similar number of n entities from Q2 has been served.

Service times for $\{\Phi_{jk}\}$ entities are assumed to follow uniform distribution. The parameters of the distribution are determined from Equation 4 on the interval defined as $\max\{\Phi_{jk}\}, \forall j,k$ and $\min\{\Phi_{jk}\}, \forall j,k$. The probability distribution function (pdf) for $x_{\Phi_{jk}}$, the service time, is given in Equation 6:

$$f(x_{\Phi_{jk}}) = \begin{cases} \dfrac{1}{\max\limits_{\forall j,k}\{\Phi_{jk}\} - \min\limits_{\forall j,k}\{\Phi_{jk}\}} & \min\limits_{\forall j,k}\{\Phi_{jk}\} \le x_{\Phi_{jk}} \le \max\limits_{\forall j,k}\{\Phi_{jk}\}. \\ 0 & otherwise \end{cases} \qquad (6)$$

The entities are introduced into the system is order. The first quarter indices enter first followed by second quarter, then third quarter and finally fourth quarter in case of quarterly data. They are not allowed to change queue position in the system.

The service time for $\{CR_{jk}\}$ entities is generated from Equation 5. They follow uniform distribution in the interval defined as $\max\{CR_{jk}\}, \forall j,k$ and

$\min\{CR_{jk}\}, \forall j, k$. The pdf for $x_{CR_{jk}}$, the service time, is defined in equation (7).

Figure 1. Modelling Seasonal Indices, Random & Cyclical Components as Queue in Quarterly Seasonal Data

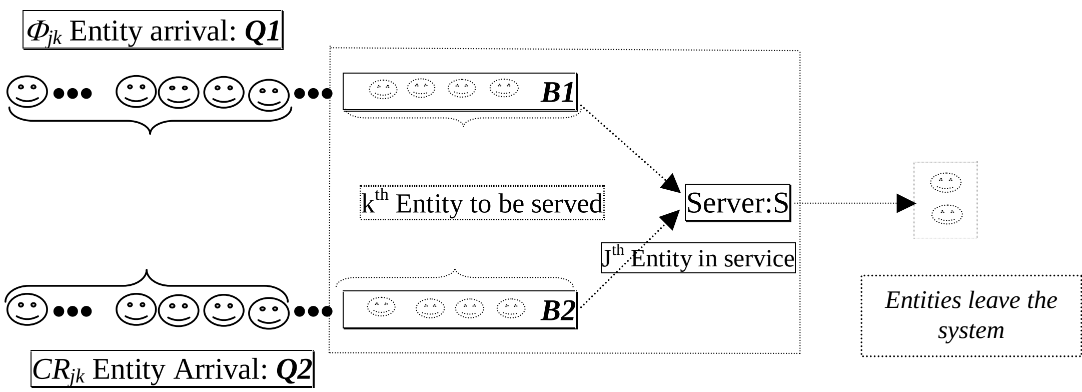
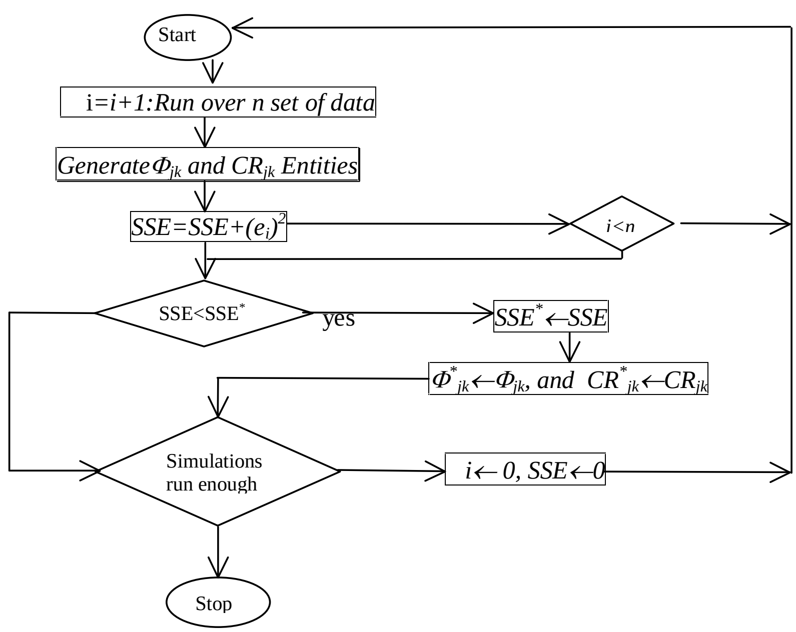


Figure 2. Flow chart for Modelling Seasonal Indices,Random & Cyclical Components as Queue in Quarterly Data



$$f(x_{CR_{jk}}) = \begin{cases} \frac{1}{\max_{\forall j,k}\{CR_{jk}\} - \min_{\forall j,k}\{CR_{jk}\}} & \min_{\forall j,k}\{CR_{jk}\} \le x_{CR_{jk}} \le \max_{\forall j,k}\{CR_{jk}\}. \\ 0 & otherwise \end{cases} \quad (7)$$

At the end of service to all the *n* entities from both the queue the following quantities are determined:

Sum of Square Error (SSE) $=\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \bar{y}_i)^2$; (8)

Mean Square Error (MSE) $= \frac{1}{n}\sum_{i=1}^{n} (y_i - \bar{y}_i)^2$ (9)

Mean Absolute Error (MAE) $= \frac{1}{n}\sum_{i=1}^{n} |y_i - \bar{y}_i|$; (10)

Relative Mean Absolute Percentage Error (RMAPE) $= \frac{1}{n}\sum_{i=1}^{n} \frac{|y_i - \bar{y}_i|}{y_i}\%$. (11)

Figure 2 displays the simulation iteration. The next simulation run begins with a new set of $\{\Phi_{jk}\}$ and $\{CR_{jk}\}$ entities. Equation 1 is evaluated in order. If the current SSE is less than the previous SSE value, the $\{\Phi^*_{jk}\}$ and $\{CR^*_{jk}\}$ values are updated corresponding to the simulation run where minimum SSE value have occurred. The process is repeated sufficiently large number of times to obtain lowest minimum value of SSE and its corresponding $\{\Phi^*_{jk}\}$ and $\{CR^*_{jk}\}$ values. Finally, the estimate of y_i as $\bar{y}_i$ is evaluated with the improved set of indices.

5. Parameter Estimation of the Simulation Model

Parameters that govern the behavior of the queuing model are estimated using Equation 4 and 5. The parameters to be determined are arrival rate, service rate, queue capacity; limits on arrival, initial condition on queue and queue list processing rule.

With infinite capacity, unlimited numbers of entities are brought into the system. Inter arrival time of entity is set to zero. Server processes two entities, $\{\Phi_{jk}\}$ and $\{CR_{jk}\}$, at a time drawing from buffer queues B1 and B2 such that the two entities belong to the same season in a period and are processed sequentially.

The parameters of the simulation model are identified for the Box et al.[1] data set on airline travel. The distribution for $\{\Phi_{jk}\}$ is uniform in the range 0.8671883 ± 1.765 at $t=1$, 1.10204 ± 1.456 at $t=6$ and 0.9416979 ± 1.855 for $t=12$. The corresponding uniform distribution parameters for $\{CR_{jk}\}$ are 1.000483 ± 0.0000575 at $t=1$, 1.014156 ± 0.0000455 at $t=6$ and $0.9974457 \pm .0000855$ at $t=12$. The term $\overline{\Psi}_k$ is modeled with $m=3$, resulting in $\alpha_0 = 103.331700$, $\alpha_1 = 18.922530$ and $\alpha_2 = 0.9976193$. Two entities with similar seasonal attribute values from the buffer queue $B1$ and $B2$ are processed on FIFO basis. These two entities leave the system once their services are complete. Next set of entities is processed until all the t entities are exhausted from the buffer queue. The next set of *t* entities are taken into the buffer queue $B1$ and $B2$ from queue $Q1$ and $Q2$. The parameter for the data set under consideration contains $n=144$, $t=12$ and $k=12$ receptively. The simulation follows the scheme given in Figure 2. the entities $\{\Phi_{jk}\}$ and $\{CR_{jk}\}$ are generated and Equation 1 is evaluated repeatedly n times. Once the index i has reached the value *n*, the current SSE value is compared with previous SSE value. Updated value of $\{\Phi_{jk}\}$ and $\{CR_{jk}\}$ is maintained throughout the simulation whenever the SSE is improved. The simulation is continued sufficiently large number of times to update the values of $\{\Phi_{jk}\}$ and $\{CR_{jk}\}$. At the end of simulation improved value of $\{\Phi_{jk}\}$ and $\{CR_{jk}\}$ are used to resonalize the data and finally the forecast is evaluated.

6. Results

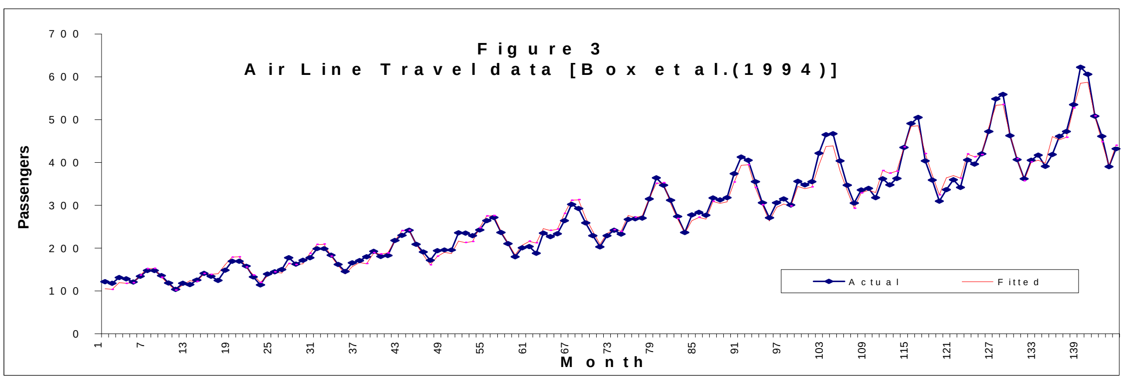
Figure 3 shows the fitted model using simulation approach, which clearly fits well the data. Figure 4 is the error pattern and many disturbances are occurring after 112^{th} month. These disturbances are due to change of trend. Table 1 displays the statistical results of the fitted model against the X-11 decomposition method. The SSE value is reduced 38.89% as compared to X-11 method and demonstrates significant improvement using the suggested simulation approach. The RMAPE and MAE measures improve by 15.51%, and 18.48% respectively. Since no assumptions on error distribution holds the ME cannot be interpreted against X-11 method. The simulation is continued for total of 20,000 runs to improve the SSE criteria.

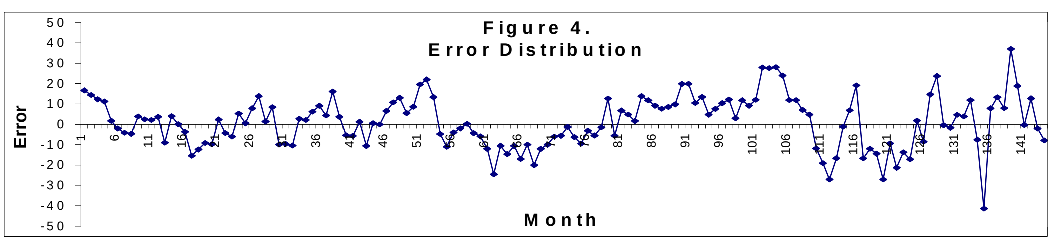
7. Conclusions

Seasonal indices of the time series data set are modeled as multi channel queuing discipline-using FIFO processing rule. Customer in multi channel queue is processed using sequential processing rule one customer at a time from each channel. The parameters of the queue in model are identified using computed parameters in the decomposition time series. The seasonal factors are updated using squared error criteria. It is found that the performance of the new method out performs X-11 decomposition method on Box et al. [1] data set on airline travel. The proposed method achieves 38.89% improvements on SSE criteria.

Table 1. Simulation results

Measures	X-11 filter(Multiplicative Decomposition method)	Simulation approach	% improvement $\frac{Value_{x-11}-Value_{Simulation}}{Value_{x-11}}\%$
Y average	280.368	208.368	-
Number of data used	144	144	-
SSE	36505.33	22306.88	38.89
MAE	12.21	9.95	18.48
MAPE	4.481	3.786	15.51
MSE	253.51	154.91	-





References

1. G.E.P. Box, G.M. Jenkins and G.C. Reinsel, Time Series Analysis Forecasting and Control, 3rd Ed., Prentice Hall, Eaglewood, NY, USA, (1994).

2. P.T. Ittig, A Seasonal Index for Business, Decision Sciences, 28, (2), 335-355, (1997).

3. T. Jaditz, Seasonality: Economic Data and Model Estimation, Monthly Labor Review, (December), 17-22, (1994).

4. Makridakis, S., and Wheelwright, S., Forecasting Methods for Management, 5th Ed. New York, John Wiley, USA, (1989).

5. W. Persons, Indices of Business Conditions, Review of Economics and Statistics, 1, (Jan), 5-107, (1919a).

6. W. Persons, An Index of General Business Conditions, Review of Economics and Statistics, 1, (April), 111-212, (1919b).

7. J. Shiskin, Decomposition of Economic time Series, Science, 128, (3338), 1539-1546, (1958).

8. D. Teodorescu, Time Series Decomposition and Forecasting, Int. J. Control, 50,(5), 1577-1585, (1989).