LONG SEASONAL PERIODS MODELING

JIŘÍ PROCHÁZKA, SAMUEL FLIMMEL, MILAN JANTOŠ, MILAN BAŠTA
University of Economics, Prague, Faculty of Informatics and Statistics,
Department of Statistics and Probability,
W. Churchill Sq. 4, Prague, Czech Republic
e-mail: xproj16@vse.cz, flis00@vse.cz, janm11@vse.cz, milan.basta@vse.cz

Abstract

In recent years there has been a growing need in practice to model and forecast time series with more complex seasonal patterns, such as seasonality with long periods, multiple seasonality. Unfortunately, most of the methods for seasonality modeling are focused on shorter seasonal periods. The goal of this paper is to present and illustrate different ways of modeling long seasonal patterns. Special focus is given on models employing a basis expansion such as the Fourier basis. An illustration is provided making use of an electricity demand dataset. The real potential of the presented methods lies in their applications to data associated with performance marketing, supply management and related areas.

Key words: *long seasonal period, multiple seasonality, exponential smoothing, Fourier basis, BATS, TBATS.*

1. Introduction

Lots of companies are coming into contact with high volume of data that are recorded with high frequency. For efficient use of such datasets, there is a need to develop models, which can be estimated in a fraction of a second and which can also provide accurate predictions. This data usage can be used to gain a comparative advantage over competitors. For example, in performance marketing this can be used for efficient campaign planning or for planning of supply management of high speed goods etc.

The above mentioned time series often contain simple or multiple seasonal patterns with different periods. Plenty of methods have been suggested in the literature for the purpose of seasonality modeling. However, most of them are designed for short seasonal periods and need not be appropriate for modeling seasonality with long periods. The goal of this paper is to present and illustrate different ways of modeling long seasonal patterns with a special focus given on models based on basis expansion such as the Fourier basis. An illustration is provided employing an electricity demand dataset.

The paper is organized as follows. Section 2 introduces methods used for simple seasonality modeling, i.e. for modeling seasonal components with only one seasonal period. The section is divided into two parts. In the first part standard approaches to simple seasonality modeling are introduced. In the second part, an alternative approach based on basis expansion methods is outlined. Section 3 discusses multiple seasonality modeling, i.e. modeling seasonal components with more than one seasonal period. Section 4 is focused on models with complex seasonal patterns, namely the BATS and the TBATS models. Benefits of different approaches are discussed in Section 5, where the approaches are used to model the half-hourly electricity demand time series. The results are summarized in Section 6.

2. Simple Seasonality Modeling

A time series is often assumed to contain several components: the trend component, the seasonal component and the error component. If the seasonal component contains only one seasonal pattern (the period being denoted as L), we speak about simple seasonality. Simple seasonality is usually present in monthly ($L = 12$) or quarterly ($L = 4$) time series. Should the seasonal component contain multiple seasonal patterns with different periods (L_1, L_2, …), multiple seasonality is present in the time series. Multiple seasonality can often be found in time series observed, for example, on a daily basis over several years where the seasonal component may consist of two seasonal patterns with periods $L_1 = 7$ and $L_2 = 365$.

If the period of a seasonal pattern (either in the case of simple or multiple seasonality) is rather long – say, (much) greater than 12 – we speak about long seasonal periods.

2.1. Standard Methods for Simple Seasonality Modeling

Common methods used for seasonality modeling are linear regression models with seasonal dummy variables, SARIMA models and models based on exponential smoothing. All of these models were primarily constructed for modeling simple seasonality with rather short periods L.

Probably the simplest model is the one which assumes that a seasonal time series $\{X_t: t = 1, \ldots, N\}$ of length N can be written as

$$X_t = B_t + S_t + E_t, \qquad t = 1,\ldots, N, \tag{1}$$

where $\{B_t: t = 1, \ldots, N\}$, $\{S_t: t = 1, \ldots, N\}$ and $\{E_t: t = 1, \ldots, N\}$ stand for a non-seasonal deterministic component (such as a deterministic trend), a seasonal deterministic component and for a stationary ARMA process. Subsequently, in the context of Equation (1), the seasonal component $\{S_t\}$ can be modeled employing dummy variables. Specifically,

$$S_t = \sum_{k=1}^{L-1} \beta_k D_{k,t}, \qquad t = 1,\ldots, N, \tag{2}$$

where β_k (for $k = 1, \ldots, L-1$) is a parameter and $\{D_{k,t}: t = 1, \ldots, N\}$ (for $k = 1, \ldots, L-1$) is dummy-variable time series which is equal to one for those times t which correspond to the kth season in the seasonal pattern, otherwise it is equal to zero. An obvious problem with this seasonality model is the fact that it is necessary to estimate $L-1$ parameters, which can be a computational problem in the case of very long seasonal periods (such as those with $L = 365$ etc.). Moreover, if the length of the time series N is not much larger than L, the estimators of the β_k parameters (for $k = 1, \ldots, L-1$) will be highly variable, the estimated seasonal pattern will be "non-smooth", and the predictions from the model inaccurate.

Another possible modeling approach is based on SARIMA models, which are a generalization of ARIMA models that allows for handling seasonality. Specifically, a SARIMA$(p,d,q)(P,D,Q)_L$ model denotes a SARIMA model with non-seasonal orders (autoregressive order, order of integration and moving-average order) given as p, d and q and seasonal orders given as P, D and Q, assuming seasonality with period L (Hyndman and Athanasopoulos, 2013). For example, a SARIMA$(1,0,0)(1,0,0)_{365}$ model can be written as

$$(1-\phi \mathrm{B})(1-\Phi \mathrm{B}^{365}) X_t = e_t, \tag{3}$$

where B denotes the backshift operator (for more details about backshift notation see e.g. Hyndman and Athanasopoulos 2013), ϕ and Φ are parameters and $\{e_t\}$ is a white noise process. A big disadvantage of the SARIMA approach is that one is losing L observations if $D = 1$, i.e. if seasonal differencing is applied. Another problem is that SARIMA models cannot guarantee the smoothness of the estimated seasonal pattern.

Another approach to simple seasonality modeling is to employ models based on exponential smoothing. The most frequently used model is presumably Holt-Winters exponential smoothing which is capable of handling both additive and multiplicative seasonality. Since the time of Winter's modification of the original Holt's model, many other modifications have been further suggested which include the possibility to combine additive and multiplicative components, add a dumping parameter for the trend etc. Probably the biggest improvement of Holt-Winters exponential smoothing is the innovative state space approach (Hyndman, 2008) which allows for an easier estimation of model parameters as well as of the initial values, and is capable of capturing the temporal evolution of the parameters. Generally speaking the Holt-Winters approach decomposes $\{X_t\}$ into a level $\{l_t\}$, trend $\{b_t\}$, seasonal component $\{s_t\}$, and an error $\{\varepsilon_t\}$. In the case of an additive model, Holt-Winters model based on the innovation state space approach without the dumping parameter is given as

$$\begin{aligned} X_t &= l_{t-1} + b_{t-1} + s_{t-L} + \varepsilon_t , \\ l_t &= l_{t-1} + b_{t-1} + \alpha\varepsilon_t , \\ b_t &= b_{t-1} + \beta\varepsilon_t , \\ s_t &= s_{t-L} + \gamma_\omega \varepsilon_t , \end{aligned} \tag{4}$$

where $\{\varepsilon_t\}$ is a sequence of normally independently distributed random variables with zero mean and constant variance (i.e. Gaussian white noise) and α, β and γ_ω are smoothing parameters for the level, trend and seasonal component (Hyndman and Athanasopoulos, 2013). Besides the smoothing parameters, it is necessary to estimate L initial values corresponding to the initial state of the seasonal cycle. This can be computationally demanding. Moreover, the predicted seasonal pattern need not be smooth.

2.2. Fourier, Spline and Wavelet Basis Functions

Let us again assume the model of Equation (1). Another commonly used approach to representing the seasonal component of the equation is the decomposition of the component into Fourier basis functions (sines and cosines). Specifically, for an odd value of L, the seasonal component $\{S_t\}$ can be expressed as

$$S_t = \sum_{K=1}^{\lfloor L/2 \rfloor} \alpha_K \sin\left(2\pi \frac{K}{L} t\right) + \sum_{K=1}^{\lfloor L/2 \rfloor} \beta_K \cos\left(2\pi \frac{K}{L} t\right), \quad t = 1, \dots, N. \tag{5a}$$

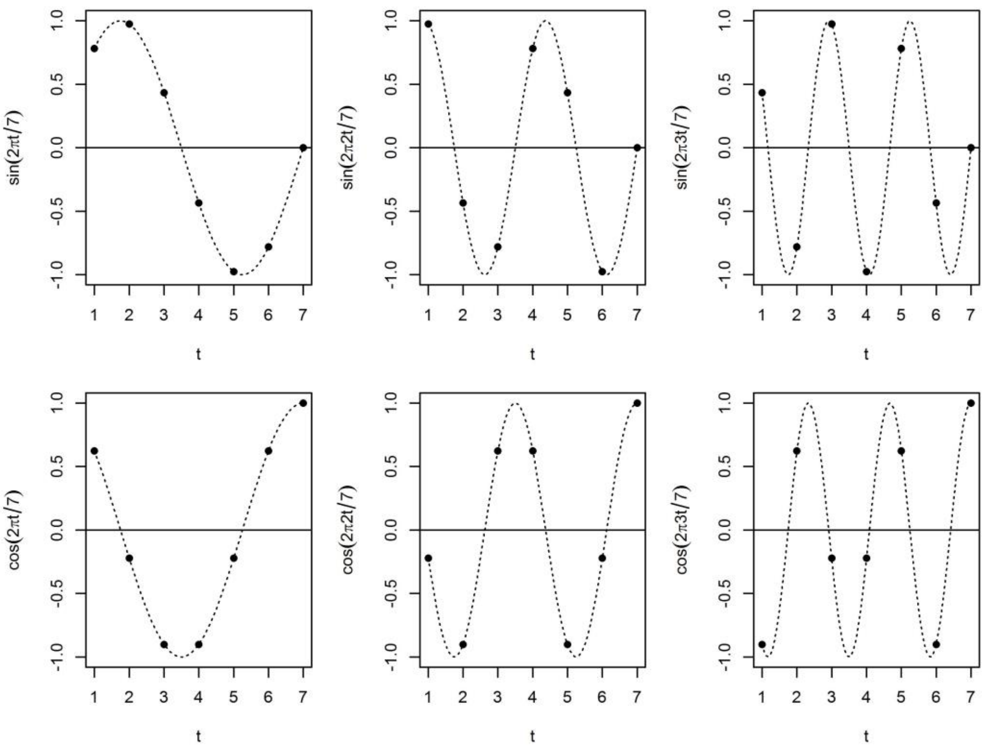
For an even value of L, we can write

$$S_t = \sum_{K=1}^{L/2-1} \alpha_K \sin\left(2\pi \frac{K}{L} t\right) + \sum_{K=1}^{L/2} \beta_K \cos\left(2\pi \frac{K}{L} t\right), \quad t = 1, \dots, N. \tag{5b}$$

α_Ks and β_Ks in both the equations are parameters. The basis functions which are associated with low values of K capture low-frequency features of the seasonal component, whereas those basis functions associated with high values of K capture high-frequency features of the seasonal component (see also Figure 1). Note that Equations (5a) and (5b) do not provide the usual decomposition (into sines and cosines) of the time series $\{X_t: t = 1, \ldots, N\}$ from which the periodogram could be obtained. Rather, the equations provide the representation of a periodic seasonal pattern with the period of length L. Consequently, the frequencies of sines and cosines are of the form K/L in Equations (5a) and (5b) rather than K/N as would the case in the usual decomposition of $\{X_t: t = 1, \ldots, N\}$.

The representation of $\{S_t\}$ given in Equations (5a) and (5b) allows for the representation of a periodic seasonal component of any shape and is practically equivalent to the representation given in Equation (2). However, in real-life analysis, we often do not have to use all the basis functions given in Equations (5a) and (5b). A good approximation of the seasonal component can be achieved if we use only the sines and cosines with the lowest frequencies (i.e. those corresponding to $K = 1, 2, \ldots$) since many real-life seasonal components do not contain major high-frequency features and are of a "smooth" shape. The reduction of the number of sines and cosines (by assuming only those with the lowest frequencies) also leads to the reduction of the number of parameters (α_Ks and β_Ks) which have to be estimated. This can be beneficial for forecasting purposes due to the bias-variance trade-off issue.

Figure 1: The six basis functions of Equation (5a) for $L = 7$



Note: The first row corresponds to sines, the second row to cosines. The columns correspond to $K = 1$, 2 and 3. Only the first 7 values are plotted for each function since the functions are periodic with the period equal to 7.

Source: the authors.

One of the disadvantages of the Fourier basis is the fact that the basis functions are not localized in time. Consequently, the Fourier basis does not allow for the representation of local characteristics present in the seasonal pattern. For this purpose, basis functions based on regression splines or wavelets (see e.g. Gencay et al., 2001) could potentially be used instead. Alternatively, penalized splines (P-splines) can be employed for seasonality modeling (Lee and Durbán, 2012).

3. Multiple Seasonality Modeling

Only models for simple seasonality have been discussed so far. However, many economic time series contain multiple seasonality. In order to model multiple seasonality, the previously mentioned models for simple seasonality can be slightly modified. Specifically, let us assume a time series where the seasonal component consists of M seasonal patterns with periods $L_1, L_2, \ldots, L_M$.

The models of simple seasonality based on dummy variables and the SARIMA approach (see Section 2.1) can be straightforwardly extended to handle multiple seasonality. In the former case, there will be M sets of dummy variables, each set being associated with one of the M seasonal patterns. In the latter case, there will be M seasonal parts in the model, each part corresponding to one of the M seasonal patterns.

The model of Equation (4) can be generalized to handle multiple seasonality as follows

$$\begin{aligned} X_t &= l_{t-1} + b_{t-1} + \sum_{i=1}^{M} s_{t-L_i}^{(i)} + \varepsilon_t, \\ l_t &= l_{t-1} + b_{t-1} + \alpha\varepsilon_t, \\ b_t &= b_{t-1} + \beta\varepsilon_t, \\ s_t^{(1)} &= s_{t-L_1}^{(1)} + \gamma_{L_1}\varepsilon_t, \\ &\vdots \\ s_t^{(M)} &= s_{t-L_M}^{(M)} + \gamma_{L_M}\varepsilon_t. \end{aligned} \tag{6}$$

The number of initial values that have to be estimated can be a computational problem for the model of Equation (6). For example, if there are two seasonal patterns (i.e. $M = 2$) with periods L_1 and L_2, we must estimate $L_1 + L_2$ initial values. If L_1 is divisible by L_2, the number of initial values to be estimated can be slightly reduced (e.g. Gould, 2008).

Another approach to handle the issue with a large number of parameters is to assume the model of Equation (1) and represent the seasonal component of this model in terms of Fourier basis functions analogously to Section 2.2., i.e.

$$S_t = \sum_{i=1}^{M}\left[\sum_{K=1}^{h_{i1}} \alpha_{iK} \sin\left(2\pi \frac{K}{L_i} t\right) + \sum_{K=1}^{h_{i2}} \beta_{iK} \cos\left(2\pi \frac{K}{L_i} t\right)\right], \tag{7}$$

where $h_{i1} = h_{i2} = \text{floor}(L_i/2)$ if L_i is odd[1], $h_{i1} = L_i/2 - 1$ if L_i is even, and $h_{i2} = L_i/2$ if L_i is even. Analogously to the case of simple seasonality, we do not have to use all the frequencies for each component and it may be sufficient to use only the lowest ones. Again, such an approach can be beneficial in terms of the bias-variance trade-off issue and the forecast accuracy.

4. BATS and TBATS

The BATS and TBATS models are extensions of the approaches mentioned above. When employing these models, the observed time series is transformed using the Box-Cox transformation at first. The motivation behind this transformation is to handle nonlinearity which can be present in the original time series.

[1] The function floor(x) returns the greatest integer that is less than or equal to x.

4.1 BATS Model

The acronym BATS stands for Box-Cox transformation, ARMA errors, and Trend and Seasonality. The original time series is transformed using the Box-Cox transformation. The transformed time series is denoted as $X_t^{(\omega)}$, where ω is a parameter of the transformation which has to be estimated. Subsequently, the BATS model employs a slight modification of the Holt-Winters method. Specifically, following de Livera et al. (2012) the BATS model can be written as

$$
\begin{aligned}
X_t^{(\omega)} &= \begin{cases} \dfrac{X_t^{\omega}-1}{\omega}, & \omega \neq 0, \\ \log X_t, & \omega = 0, \end{cases} \\
X_t^{(\omega)} &= l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{M} s_{t-L_i}^{(i)} + d_t, \\
l_t &= l_{t-1} + \phi b_{t-1} + \alpha d_t, \\
b_t &= (1-\phi) b + \phi b_{t-1} + \beta d_t, \\
s_t^{(i)} &= s_{t-L_i}^{(i)} + \gamma_i d_t, \\
d_t &= \sum_{i=1}^{p} \varphi_i d_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t,
\end{aligned}
\tag{8}
$$

where ω is the parameter of the Box-Cox transformation, l_t is the local level of the time series at time t, b is the long run trend, b_t is the short run trend at time t, $s_t^{(i)}$ is the ith seasonal pattern at time t, d_t denotes an ARMA(p, q) process of orders p and q at time t, and ε_t is a Gaussian white noise at time t. α, β and γ_i (for $i = 1, 2, \ldots, M$) are smoothing parameters. ϕ is the so-called dumping parameter and allows for the flattening of the trend in the long run. To sum up, the BATS model has the following parameters $\{\omega, \phi, p, q, L_1, L_2, ..., L_M\}$. A special case of the BATS model is the BATS(1, 1, 0, 0, L_1) model which corresponds to the classical Holt-Winters simple exponential smoothing. The BATS model inherits a lot of bad properties of the classical Holt-Winters method including the problem with long seasonal periods.

4.2 TBATS Model

The TBATS model is useful when a time series with a complicated seasonal pattern is encountered. The first letter T in TBATS stands for the trigonometric representation. Specifically, the problem with the overparametrization of the BATS model is resolved by decomposing the seasonal pattern into trigonometric functions. Unlike in Equation (8), the TBATS model brings further modifications such as the use of the innovative state space approach etc. For a more detailed description of the TBATS model see de Livera et al. (2012). The TBATS model has the following parameters $\{\omega, \phi, p, q, \{L_1;K_1\}, \{L_2;K_2\}, ..., \{L_M;K_M\}\}$. Parameters ω, ϕ, p, q and $L_1, \ldots, L_M$ are consistent with the BATS model, and parameters $K_1, \ldots, K_M$ denote the number of frequencies used within each seasonal component. Each seasonal pattern is represented by a specific decomposition which means that one can use different number of frequencies ($K_1, \ldots, K_M$) for each seasonality component.

To sum up:

1) The TBATS model reduces the parameter space without a negative impact on the predictive accuracy.
2) The TBATS model can deal with nested and non-nested multiple seasonal components (the term non-nested seasonal components means that the length of the longer seasonal component is not divisible by the length of the shorter one.).
3) The TBATS model handles nonlinear features in time series.
4) The TBATS model can take any autocorrelation in residuals into account.
5) The TBATS model provides a much simpler estimation procedure.

As was already mentioned, the TBATS model is very similar to the BATS model. The main difference between the two models is that in the TBATS model each seasonal pattern is represented by a specific decomposition using the Fourier basis. This decomposition is also extended by the state space approach which means that the parameters of every single decomposition are perceived as time series. Following de Livera et al. (2012) the seasonal component in the TBATS model can be represented as

$$\begin{aligned} s_t^{(i)} &= \sum_{j=1}^{K_i} s_{j,t}^{(i)}, \\ s_{j,t}^{(i)} &= s_{j,t-1}^{(i)} \cos\left(\frac{2\pi j}{L_i}\right) + s_{j,t-1}^{*(i)} \sin\left(\frac{2\pi j}{L_i}\right) + \gamma_1^{(i)} d_t, \\ s_{j,t-1}^{*(i)} &= -\, s_{j,t-1} \sin\left(\frac{2\pi j}{L_i}\right) + s_{j,t-1}^{*(i)} \cos\left(\frac{2\pi j}{L_i}\right) + \gamma_2^{(i)} d_t, \end{aligned} \tag{9}$$

where $\gamma_1^{(i)}$ and $\gamma_2^{(i)}$ (for $i = 1, 2, \ldots, M$) are smoothing parameters (as in Equation (8)), d_t again denotes an ARMA(p, q) process and $K_i = L_i/2$ when L_i is even and $K_i = (L_i - 1)/2$ when L_i is odd.

5. Electricity Demand Modeling

Short term electricity demand modeling is crucial in terms of production planning. It is necessary for the safety and reliability of the electric system. Making use of the Taylor time series (e.g. Taylor, 2003) from the R forecast package (e.g. Hyndman, 2008), we will illustrate the implementation of different methods for seasonality modeling mentioned in the previous sections. The time series captures half-hourly electricity demand in England and Wales from Monday 5 June 2000 to Sunday 27 August 2000 (see Figure 2).

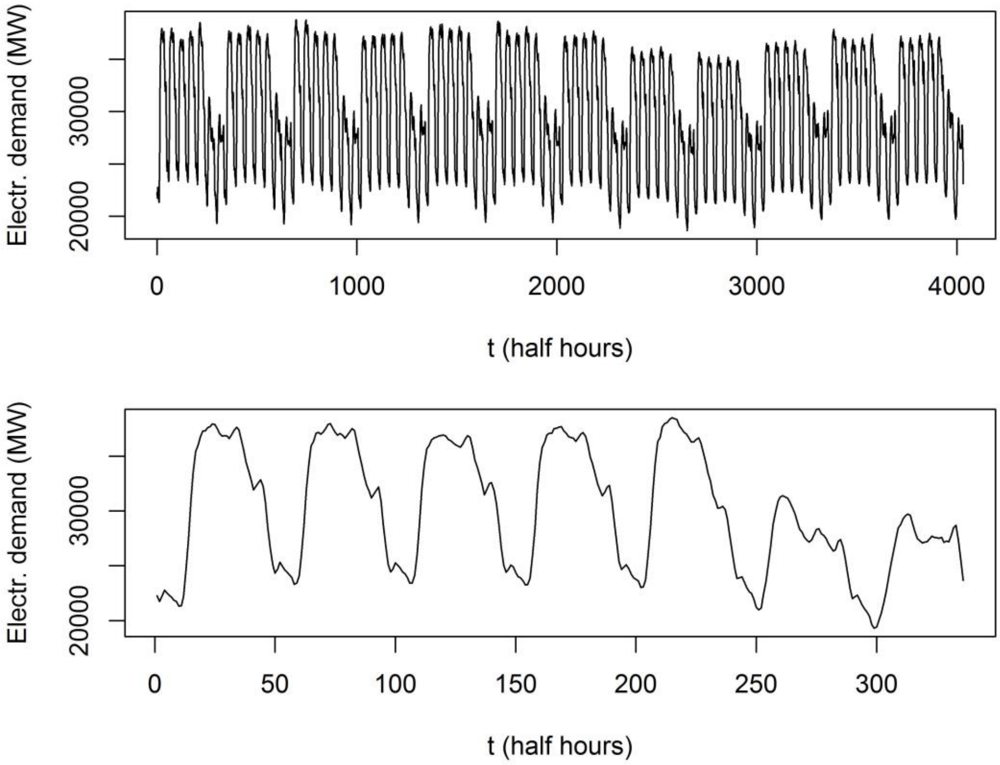
The time series consists of 4032 observations covering 84 days (12 weeks). In the upper part of Figure 2 the seasonal pattern with L = 336 (one week) is visible. From the bottom part of Figure 2 the seasonal pattern with L = 48 (one day) is visible. In order to compare the different approaches to seasonality modeling, the time series will be split into a training part represented by the first ten weeks and an evaluation part represented by the last two weeks. Five approaches to modeling seasonality will be employed. Namely,

1) The model of Equation (1) combined with Equation (2) with L = 48. The B_t component in Equation (1) was set to zero and E_t is assumed to be an ARMA(3,2) process. The order of the ARMA process was chosen based on the “auto arima” procedure (see e.g. Hyndman and Khandakar, 2009).
2) The model of Equation (1) combined with Equation (5b) with L = 336 where only the lowest frequencies of sines and cosines are employed (see the text below Equation (5b)).

Specifically, based on the Akaike information criterion the frequencies corresponding to the first 40 lowest values of K are assumed.

3) Double-seasonal Holt-Winters model with L_1 = 336 and L_2 = 48 (by the double-seasonal Holt-Winters model we mean the Holt-Winters model with two seasonal components).
4) BATS model (the parameter of the Box-Cox transformation, and the smoothing and dumping parameters are estimated using maximum likelihood, other parameters being estimated using Kalman filter. The order of the ARMA process is determined using automatic selection based on the Akaike information criterion).
5) TBATS model. The estimation procedure is very similar to the BATS model. Moreover, in the TBATS model the number of frequencies for each seasonal component is determined using automatic selection based on Akaike information criterion.

Figure 2: Half-hourly electricity demand in England and Wales.



Note: Top figure shows the entire time series. Bottom figure shows only the first 336 observations (7 days).

Source: the authors.

Following the estimation procedure, predictions two weeks ahead will be compared with the evaluation part, root mean squared error (RMSE) being used as an evaluation criterion.

Regarding computational complexity, the BATS model (Model 4) is unsuitable for this type of time series. Specifically, the estimation of the BATS model has not converged.

Table 1: Values of RMSE for the assumed models.

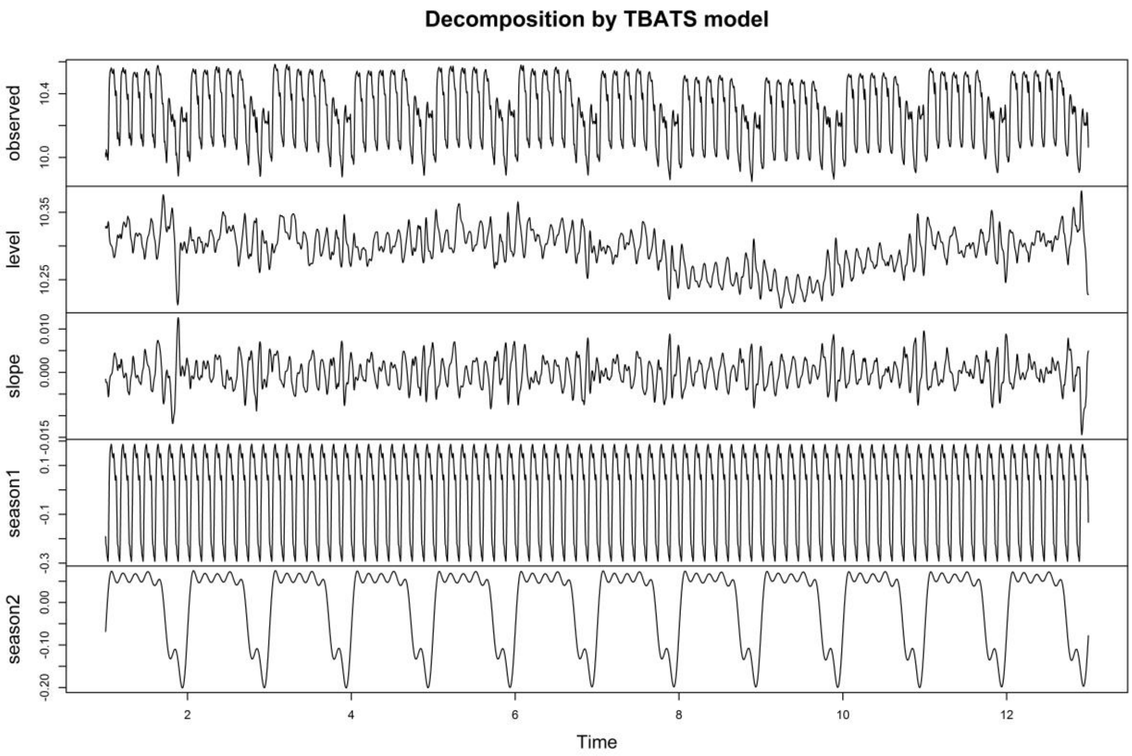
Characteristic	Model 1	Model 2	Model 3	Model 4	Model 5
RMSE	3260.30	672.95	1097.15	-	1382.88

Source: the authors.

The highest RMSE was achieved for Model 1. This is not surprising since Model 1 is the simplest model capturing only the short seasonal pattern and employing many parameters at the same time. On the other hand, Model 2 has achieved the lowest RMSE. Very good results have also been obtained for Model 3. Model 5 performs similarly to Model 3.

Model 5 can also be used to provide the decomposition of the original time series, see Figure 3. The figure consists of five subplots. The subplots describe the temporal dynamics of (starting from the second subplot) the level term, the trend term, the seasonal pattern with $L_1 = 48$, and the seasonal pattern with $L_2 = 336$.

Figure 3: Decomposition of electricity demand time series using the TBATS model.



Source: the authors.

6. Conclusion

This paper summarized the most common methods used for modeling seasonality with long seasonal periods. In Section 5 we demonstrated that even a simple model, i.e. Model 2, can give better predictions than more complex models (BATS, TBATS etc.). On the other hand, the complexity of the BATS and TBATS models guarantees suitability for a wider range of time series. To make a more reliable evaluation of the forecast accuracy of different models, it would be necessary to compare the models on a larger set of time series.

As in Section 5, the electricity demand time series was used only for the purpose of illustration. Further research will be focused on more practical applications in the area of supply management and performance marketing. The models described above could be used for modeling demand in any area where the construction of hourly prediction is relevant. The reason for constructing such predictions could be the efficient supply planning or efficient planning of marketing campaigns.

Potential improvements of the approaches employing Fourier basis functions can also be obtained by using different basis functions such as splines, wavelets etc. Those improvements along with the currently used models could help develop systems that can help companies achieve comparative advantage over competitors.

Acknowledgements

This paper is supported by the grant F4/67/2016 (Modelování sezonních časových řad s velkou délkou sezónnosti) which has been provided by the Interní grantová agentura Vysoké školy ekonomické v Praze (Internal Grant Agency of the University of Economics in Prague).

References

[1] DE LIVERA, A. M., HYNDMAN, R. J., SNYDER, D. R. 2012. Forecasting time series with complex seasonal patterns using exponential smoothing. In Journal of the American Statistical Association, 2012, vol. 106, pp. 1513-1527.

[2] EVERETTE S., GARDNER, J. 1985. Exponential smoothing : The state of the art. In Journal of Forecasting, 1985, vol. 4, iss. 1, pp. 1-28.

[3] GENCAY, R., SELCUK, F., WHITCHER, B. 2001. Differentiating intraday seasonalities through wavelet multi-scaling. In Physica A, 2001, vol. 289, iss. 3-4 , pp. 543-556.

[4] GOULD, P. G. et al. 2008. Forecasting time series with multiple seasonal patterns. In The European Journal of Operational Research, 2008, vol. 191, pp. 207-222.

[5] HYNDMAN, R. J., ATHANASOPOULOS, G. 2013. Forecasting: principles and practice. Melbourne : OTexts, 2013. http://otexts.org/fpp/2/3.

[6] HYNDMAN, R. J., KHANDAKAR, Y. 2008. Automatic time series forecasting : the forecast package for R. In Journal of Statistical Software, 2008, pp. 1-22.

[7] HYNDMAN, R. J. 2008. Forecasting with exponential smoothing : The state space approach. New York : Springer, 2008

[8] LEE, D.-J., DURBÁN, M. 2012. Seasonal modulation mixed models for time series forecasting. In DES - Working Papers. Statistics and Econometrics, 2008, pp. 12-25.

[9] TAYLOR, J. 2003. Short-term electricity demand forecasting using double seasonal exponential smoothing. In Journal of the Operational Reseach Society, 2003, vol. 54, iss. 8, pp. 799-805.