



**MONASH** University

Department of Econometrics and Business Statistics

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

**STR: A Seasonal-Trend  
Decomposition Procedure Based  
on Regression**

Alexander Dokumentov and Rob J Hyndman

June 2015

Working Paper 13/15

# STR: A Seasonal-Trend Decomposition Procedure Based on Regression

**Alexander Dokumentov**

Department of Econometrics and Business Statistics,  
Monash University, VIC 3800

Australia.

Email: [alexander.dokumentov@monash.edu](mailto:alexander.dokumentov@monash.edu)

**Rob J Hyndman**

Department of Econometrics and Business Statistics,  
Monash University, VIC 3800

Australia.

Email: [rob.hyndman@monash.edu](mailto:rob.hyndman@monash.edu)

8 June 2015

**JEL classification:** C10,C14,C22

# STR: A Seasonal-Trend Decomposition Procedure Based on Regression

## Abstract

We propose new generic methods for decomposing seasonal data: STR (a Seasonal-Trend decomposition procedure based on Regression) and Robust STR. In some ways, STR is similar to Ridge Regression and Robust STR can be related to LASSO. Our new methods are much more general than any alternative time series decomposition methods. They allow for multiple seasonal and cyclic components, and multiple linear regressors with constant, flexible, seasonal and cyclic influence. Seasonal patterns (for both seasonal components and seasonal regressors) can be fractional and flexible over time; moreover they can be either strictly periodic or have a more complex topology. We also provide confidence intervals for the estimated components, and discuss how STR can be used for forecasting.

**Keywords:** time series decomposition, seasonal data, Tikhonov regularisation, ridge regression, LASSO, STL, TBATS, X-12-ARIMA, BSM

## 1 Introduction

Time series decomposition is an important task in all national statistics agencies, and in many other contexts in which seasonal variation in time series data is observed. It is the basis for all seasonal adjustment procedures, it is widely used in forecasting, and it underpins the analysis of business cycles.

The first attempts to decompose time series into various components can be dated as early as 1884 when Poynting proposed price averaging as a tool to eliminate trend and seasonal fluctuations (Makridakis et al., 1998). Later his approach was extended by Hooker (1901), Spencer (1904) and Anderson and Nochmals (1914). Further research in that direction included Copeland (1915), who was the first to attempt to extract the seasonal component, until Macauley (1930) proposed a method which became “classical” over time. The work of Macauley led to the Census II method, which became widely used after a computer program developed in 1955 significantly simplified the calculations (Shishkin, 1957). The Census II method has continued to evolve, and various techniques and features have been added such as robustness, calendar effects, regressors, ARIMA extensions, and extensive diagnostics. Widely used versions of this approach are X-11 (Shishkin et al., 1967), X-11-ARIMA (Dagum, 1988), X-12-ARIMA (Findley et al., 1998) and X-13ARIMA-SEATS (Findley, 2005). X-13-ARIMA-SEATS includes a version of the TRAMO/SEATS procedure for seasonal adjustment which was developed at the Bank of Spain (see Monsell and Aston (2003)).

A different approach was followed by Cleveland et al. (1990) who developed STL (Seasonal-Trend decomposition using Loess) which has become widely used outside the national statistics agencies, largely because of its availability in R (R Core Team, 2015). This method uses iterative Loess smoothing to obtain an estimate of the trend and then Loess smoothing again to extract a changing additive seasonal component.

Burman (1980) argued that there were too many seasonal adjustment methods and noted that all but one were ad hoc methods. Since that time, several model-based methods for seasonal decomposition have been developed including the TRAMO/SEATS procedure mentioned above, the BATS and TBATS models of De Livera et al. (2011), and various structural time series model approaches (Harvey, 1990; Commandeur et al., 2011). The big advantage of using a model for such methods is that it allows the easy calculation of confidence and prediction intervals, which are not available in many ad hoc methods.

Despite this long history, and the availability of many time series decomposition algorithms and models, there are many time series characteristics that are not addressed in any of these approaches.

The major deficiencies of the main decomposition methods are as follows:

- Inability to provide a meaningful and simple statistical model (for many methods).
- Inability (or difficulty) to calculate confidence intervals (for many methods).
- Inability to take into account regressors (for some methods).
- Inability to take into account fractional seasonality (for most methods).
- Inability to take into account multiple seasonality (for most methods).
- Inability to take into account complex seasonality and regressors which affect data in a seasonal manner (for all methods).

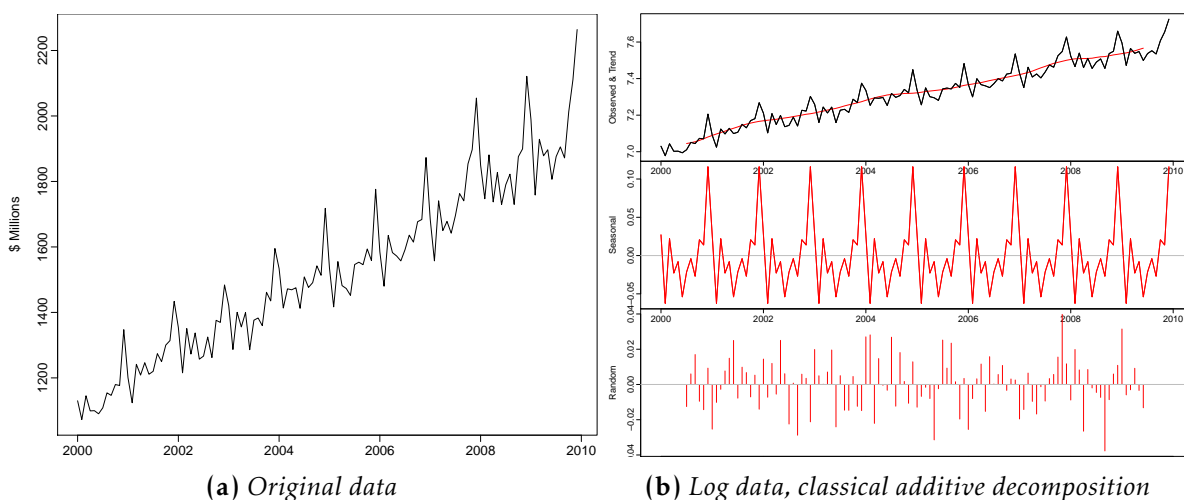
As we can see, currently a variety of methods are available, although few of them have the clarity, simplicity and generality to allow them to handle the many problems which arise with seasonal data decomposition. We aim to fill this gap with our new approach. It is clear, generic, model-based, robust (if required) and is simple — we show that the problem of seasonal decomposition can be re-cast in the framework of ordinary least squares or quantile regression. Moreover our approach allows new features (such as predictors with seasonally varying effects) that have not been developed before. In our opinion, our new STR method is the most generic framework currently available for decomposition of seasonal data.

The structure of this article is as follows: In Section 2 we provide a very simple motivating example which clarifies the main intuitions behind the idea of seasonal decomposition. In Sections 3 and 4 we develop the simplest STR model and show how it can be reduced to ordinary least squares (OLS). In Section 4.2 we show how to efficiently calculate leave-one-out cross validation and propose to use cross validation for estimation of the smoothing parameters. In Section 5 we extend our model to the case when seasonality is considered as a smooth function in two dimensions and defined over a cylinder. In Section 6 we consider cases of multiple seasonality, and we introduce predictors as well as allow them to be “flexible” and “seasonal”. In Section 8 by improving the performance of the method we also solve the problem of fractional seasonality, we introduce the concept of seasonality with complex topology, and we show how to forecast using our model. In Section 9 we introduce RSTR — a robust version of STR. In Section 10 we provide a complex example to highlight the features and capabilities of our approach. Finally in Section 11 we discuss the benefits and disadvantages of this new approach.

We also propose a way to take into account cyclicity in the data and note that it can be handled similarly to seasonality.

## 2 A simple motivating example

Seasonal time series is often assumed to consist of a few components (see for example [Cleveland et al. \(1990\)](#) or [Ladiray and Quenneville \(2001\)](#)). In this initial example, we consider the simplest case when a time series is assumed to consist of three components: seasonal, trend and remainder. The seasonal component is usually assumed to have a repeating pattern which changes very slowly or stays constant over time. The trend component is usually considered to change faster than the seasonal component. The remainder component is the most quickly changing part. The whole time series is calculated as a function of these three parts. Often they are simply added (additive seasonality) or multiplied (multiplicative seasonality) to each other to recover the original data.

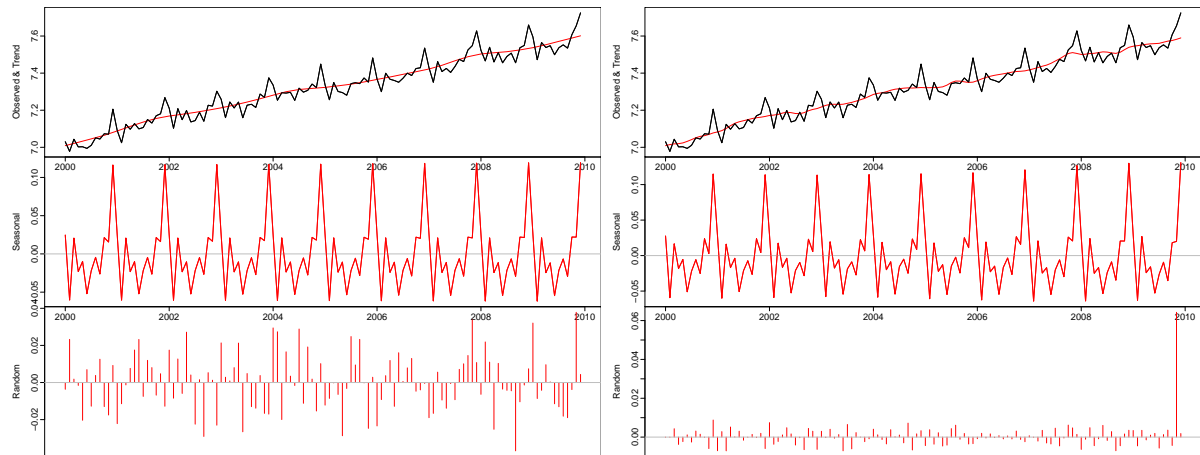


**Figure 1:** Supermarket and grocery stores turnover in New South Wales.

The time series shown in Figure 1a is supermarket and grocery stores turnover in New South Wales from 2000 to 2009 inclusive. It provides a classic example of data with multiplicative seasonality.

There are a few popular methods to deal with such data. One of them is the classical decomposition method for additive seasonality. This method assumes that the seasonal component is additive and is not changing over time. To be able to apply this method to data with multiplicative seasonality, logs are taken as shown in Figure 1b. In this paper, we consider only

additive decomposition, applied to either the original or the logged data. See [McElroy \(2010\)](#) for a discussion of the benefits of a direct implementation of multiplicative decomposition.



**Figure 2:** Supermarket and grocery stores log turnover in New South Wales decomposed with STL and X-13ARIMA-SEATS.

STL and X-13-ARIMA-SEATS are two other very well known methods; an application of each of them to the data in [Figure 1](#) is presented in [Figure 2](#). Both methods are iterative in nature and rather complex in details; see [Cleveland et al. \(1990\)](#) for STL and [Findley et al. \(1998\)](#) for the capabilities and features of X-12-ARIMA, a predecessor of X-13-ARIMA-SEATS.

Many decomposition methods currently available, including STL and X-13-ARIMA-SEATS, lack an underlying stochastic model. As a consequence confidence intervals are not readily available for the decomposed components. The methods which have models underpinning them (for example BSM, BATS, TBATS; see [Harvey \(1990\)](#) and [De Livera et al. \(2011\)](#)) still do not provide such an option, probably due to the complexity of implementing such a procedure.

In contrast, our new proposed method STR provides a statistical model and we can easily calculate confidence intervals for the decomposed components.

### 3 A simple STR model

The first and the simplest STR model describes a time series  $Y_t$  consisting of three parts:

$$Y_t = T_t + S_t + R_t, \quad (1)$$

where  $T_t$  is the trend,  $S_t$  is the additive seasonal component, and  $R_t$  is the “remainder” component. Time  $t \in \{1, \dots, n\}$  and we have  $k$  seasons:  $1, \dots, k$ . Let us also denote function  $\text{sn}(t) \in \{1, \dots, k\}$  which transforms time  $t$  into the corresponding season  $\text{sn}(t)$ .

We assume that the seasonal component  $S_t$  has some stable repeating pattern. At time  $t$  we observe (indirectly, through equation (1)) only one element of the seasonal pattern. It is reasonable to ask what the other components of this pattern are at that very moment  $t$ . In other words we propose to define along with  $S_t$ , other “invisible” components responsible for seasons other than  $\text{sn}(t)$ . In this way, we treat the seasonal pattern as two-dimensional,  $\langle S_{i,t} \rangle_{i=1}^k$ , and we assume that  $S_t = S_{\text{sn}(t),t}$  (here and further  $S$  with one index will represent a vector of seasons from (1) and  $S$  with two indexes will represent a matrix of seasonal shapes). Thus (1) can be rewritten as:

$$Y_t = T_t + S_{\text{sn}(t),t} + R_t \quad (2)$$

where  $S = [S_{s,t}]$  is  $k \times n$  matrix,  $k$  is number of seasons and  $n$  is length of the time series.

This representation allows us to impose simple constraints on the seasonal patterns represented by matrix  $S = [S_{s,t}]$ . The whole model can be described as follows:

- The remainder terms  $R_t$  are i.i.d.  $\mathcal{N}(0, \sigma_R^2)$ ;
- The trend terms are smooth such that  $\Delta^2 T_t = T_{t+1} - 2T_t + T_{t-1}$  are i.i.d.  $\mathcal{N}(0, \sigma_T^2)$ ;
- The seasonal terms must sum to zero so that, for any  $t$ , they have the property  $\sum_s S_{s,t} = 0$ ;
- Each of the seasonal terms are also smoothly changing over time, so that  $\forall t$  vectors  $\langle \Delta_t^2 S_{s,t} \rangle_{s=1}^k = \langle S_{s,t+1} - 2S_{s,t} + S_{s,t-1} \rangle_{s=1}^k$  are i.i.d. by  $t$  and distributed  $\mathcal{N}(0, \sigma_S^2 \Sigma_S)$  inside the vectors, where  $\Sigma_S$  is a  $k \times k$  matrix which can be considered the covariance matrix of  $k$  random variables  $\xi_s = \eta_s - \frac{1}{k} \sum_{r=1}^k \eta_r$  obtained from i.i.d.  $\mathcal{N}(0, 1)$  random variables  $\eta_1, \dots, \eta_k$ ;
- The parameters of the model are given by  $\sigma_R, \sigma_T, \sigma_S, T_0, T_1, \langle S_{s,0} \rangle_{s=1}^k$  and  $\langle S_{s,1} \rangle_{s=1}^k$  (or  $\langle S_{s,n} \rangle_{s=1}^k$ ).

It should be noted that in this, the simplest form of STR, the seasonal component is presented as a two-dimensional array only to ensure that the seasonal components add to zero for any  $t$ . Later we will apply additional restrictions on the seasonal component and the two-dimensional structure will become even more important.



Another way to present the model (2) is to write it in the form of a state-space model in the spirit of [Harvey \(1985\)](#) or [Harvey and Peters \(1990\)](#), or as a multivariate ARIMA model. For example,  $T_t$  can be considered an ARIMA(0,2,0) process and the seasonal component is a multivariate ARIMA(0,2,0) process with contemporaneously correlated errors (but no autocorrelation).

## 4 Estimating an STR model

### 4.1 Maximum likelihood estimation

We define vector  $\ell$  as  $\ell = \langle T_t \rangle_{t=1}^n$  and  $S^-$  as matrix  $S$  without the last row. Since each column of  $S$  must sum to 0, we can write  $S = P_S S^-$  for some matrix  $P_S$ . Let us also define  $s = \text{vec}(S^-)$  to be a vector of length  $n(k-1)$  representing the seasonal components, and  $\Xi = I_{n-2} \otimes \Xi^-$  where  $\Xi^-$  is a  $(k-1) \times (k-1)$  matrix obtained from  $\Sigma_S$  by removing the last row and the last column. There also exists a matrix  $P$  such that  $\text{vec}(S) = Ps$ .

Thus, the minus log likelihood function for this model is given (up to a constant) by

$$-\log(\mathcal{L}) = \left\| \frac{y - Qs - \ell}{\sigma_R} \right\|_{L_2}^2 + \left\| \frac{\Xi^{-\frac{1}{2}} D_s s}{\sigma_S} \right\|_{L_2}^2 + \left\| \frac{D_\ell \ell}{\sigma_L} \right\|_{L_2}^2, \quad (3)$$

where

- $Q$  is a  $n \times n(k-1)$  matrix that computes elements  $\langle S_{\text{sn}(t),t} \rangle_{t=1}^n$  from vector  $s = \text{vec}(S^-)$ ;
- $D_s$  is a  $(n-2)(k-1) \times n(k-1)$  matrix that computes second differences along the time dimension:  $\langle \Delta_t^2 S_{s,t}^- \rangle_{s=1}^k = \langle S_{s,t+1}^- - 2S_{s,t}^- + S_{s,t-1}^- \rangle_{s=1}^{k-1}$  for  $2 \leq t \leq n-1$ ;
- $D_\ell$  is  $(n-2) \times n$  matrix that calculates second differences of  $\ell$ :  $\Delta^2 T_t = T_{t+1} - 2T_{t-1} + T_{t-1}$  for  $2 \leq t \leq n-1$ .

Thus, maximum likelihood estimates are obtained by minimizing

$$\left\| y - Qs - \ell \right\|_{L_2}^2 + \left\| \frac{\sigma_R}{\sigma_S} \Xi^{-\frac{1}{2}} D_s s \right\|_{L_2}^2 + \left\| \frac{\sigma_R}{\sigma_L} D_\ell \ell \right\|_{L_2}^2 \quad (4)$$

over  $s$  and  $\ell$ .

We can also note that (3) corresponds to the minus log likelihood function for the following linear model (here we use an approach similar to that described in [Dokumentov and Hyndman](#)

(2013)):

$$y_{ext} = X\beta + \varepsilon, \quad (5)$$

where  $\beta = [s', \ell']'$  is a vector of unknown coefficients,  $s$  is a vector of seasonal components;  $\ell$  is a vector containing the trend and  $\varepsilon$  is a vector of i.i.d. errors. Observations are defined by matrix

$$X = \begin{bmatrix} Q & I \\ \lambda_s \Xi^{-\frac{1}{2}} D_s & 0 \\ 0 & \lambda_\ell D_\ell \end{bmatrix}$$

for some fixed parameters  $\lambda_s$  and  $\lambda_\ell$ . Predictors are defined by vector  $y_{ext} = [y', 0']'$ , which is vector  $y$  extended with zeros to make it conform to matrix  $X$  defined above. All errors are i.i.d.  $\mathcal{N}(0, \sigma^2)$  for some unknown  $\sigma$ .

Since STR model and the linear model (5) have the same likelihood functions, their maximum likelihood solutions will be identical and are given by

$$\hat{\beta} = (X'X)^{-1} X' y_{ext} = \hat{\beta} = (X'X)^{-1} [Q \ I]' y. \quad (6)$$

Taking into account that the covariance matrix of  $\varepsilon_{ext} = y_{ext} - X\hat{\beta}$  has the following form:

$$\Sigma = \begin{bmatrix} \sigma^2 I_n & 0 \\ 0 & 0 \end{bmatrix},$$

where  $\sigma$  is the standard deviation of the residuals corresponding to the  $[Q \ I]$  part of matrix  $X$ , the covariance matrix of solution  $\hat{\beta}$  can be calculated as

$$\Sigma_{\hat{\beta}} = (X'X)^{-1} X' \Sigma X (X'X)^{-1} = \sigma^2 (X'X)^{-1} [Q \ I]' [Q \ I] (X'X)^{-1}. \quad (7)$$

We estimate  $\sigma$  using cross-validated residuals (see Section 4.2) instead of residuals of the fit. This ensures that  $\sigma$  is not underestimated if the model is over-fitted.

The trend component  $\hat{T}$  and the corresponding confidence intervals can be obtained directly from  $\hat{\beta}$  and  $\Sigma_{\hat{\beta}}$ . In order to obtain the seasonal components  $\hat{S}$ ,  $\hat{\beta}$  needs to be transformed with some matrix  $R$  to recalculate every last seasonal component that was removed earlier. Using matrix  $R$ , a new covariance matrix and confidence intervals for  $\hat{S}$  can also be calculated.

Let us consider a variation of model (5) where we allow the errors to be correlated with covariance matrix  $\Sigma_y$ . Then the covariance matrix of  $\varepsilon_{ext} = y_{ext} - X\hat{\beta}$  has the form

$$\Sigma = \begin{bmatrix} \Sigma_y & 0 \\ 0 & 0 \end{bmatrix}, \quad (8)$$

and if matrix  $\Sigma_y$  is invertible, the solution is

$$\hat{\beta} = (X'WX)^{-1}[Q \ I]'\Sigma_y^{-1}y$$

where

$$W = \begin{bmatrix} \Sigma_y^{-1} & 0 \\ 0 & I \end{bmatrix}.$$

The covariance matrix of the solution will be:

$$\Sigma_{\hat{\beta}} = (X'WX)^{-1}[Q \ I]'\Sigma_y^{-1}[Q \ I](X'WX)^{-1}.$$

## 4.2 Smoothing parameter estimation

The model (5) also requires specification or estimation of the parameters  $\lambda_s$  and  $\lambda_\ell$ . We propose using leave-one-out cross validation to estimate them. Using a model with minimal cross validation will allow choosing the model which ‘‘absorbs’’ information as much as possible and noise (uncorrelated errors) as little as possible.

Since model (5) is a linear model, the leave-one-out cross validation residuals can be calculated using (Seber and Lee, 2003)

$$cv_i = \frac{y_i - \hat{y}_i}{1 - h_{ii}},$$

where  $y_i$  is the  $i$ th element of vector  $y$ ,  $\hat{y}_i$  is the  $i$ th element of vector  $\hat{y} = Hy$  and  $h_{ii}$  is the  $i$ th diagonal element of the hat matrix  $H = X(X'X)^{-1}X'$ . Therefore, we can use the well-known formula for cross validation for linear regression (see for example Ruppert et al. (2003)):

$$SSE(cv) = \sum_{i=1}^n \left( \frac{y_i - \hat{y}_i}{1 - h_{ii}} \right)^2.$$

By minimising  $SSE(cv)$ , STR finds optimal parameters  $\lambda_s$  and  $\lambda_\ell$ . The problem of minimising  $SSE(cv)$  can be complex, for example with many local minima, and we have no method which guarantees finding the global minima. However, in practice rather simple methods often work well. We use R and the Nelder-Mead method as implemented in the `optim()` function from the “stats” package in R (R Core Team, 2015) for such optimisation.

On the other hand, this approach does not work for model (8) when the errors are correlated, or for very big matrices  $X$  due to the difficulties of inverting  $X'X$ . For such scenarios we use K-fold cross validation, as discussed in Section 9.

## 5 STR with two-dimensional restrictions on seasonal changes

Optimization problem (4) can be extended to constrain second discrete derivatives  $\frac{\partial^2}{\partial s \partial t}$  and  $\frac{\partial^2}{\partial s^2}$  of the seasonal component (in our case it is a two-dimensional surface) in addition to  $\frac{\partial^2}{\partial t^2}$ . This will imply a new model, but let us describe the optimisation problem first:

$$(s, \ell) = \arg \min \left[ \|y - Qs - \ell\|_{L_2}^2 + \|\lambda_{tt} D_{tt} s\|_{L_2}^2 + \|\lambda_{st} D_{st} s\|_{L_2}^2 + \|\lambda_{ss} D_{ss} s\|_{L_2}^2 + \|\lambda_\ell D_\ell \ell\|_{L_2}^2 \right], \quad (9)$$

where, similar to (4):

- $y$  is a vector of observations of length  $n$ ;
- $s$  is a vector of two-dimensional seasonal components (an  $n \times (k-1)$  matrix) presented as a vector of length  $n(k-1)$ ;
- $\ell$  is a vector of length  $n$  representing the trend component;
- $D_{tt}$ ,  $D_{st}$  and  $D_{ss}$  are matrices which compute second differences for the two-dimensional seasonal component along the time, time-season and season dimensions, respectively;
- $D_\ell$  is a matrix which calculates second differences for trend  $\ell$ ;
- $\lambda_{tt}$ ,  $\lambda_{st}$ ,  $\lambda_{ss}$  and  $\lambda_\ell$  are parameters to be selected.

It is now evident that the seasonal two-dimensional surface actually has the topology of a tube, where dimension  $t$  spreads infinitely into two directions, but dimension  $s$  is “circular” (season 1 is connected to season 2, season 2 is connected to season 3, ..., season  $(k-1)$  is connected to season  $k$  and season  $k$  is connected back to season 1). It should be noted that matrices  $D_{st}$  and  $D_{ss}$  take the “tube” topology into account to calculate proper “circular” differences

Similar to the problem discussed in Section 4, this optimisation problem corresponds to the following linear regression problem and the model:

$$y_{ext} = X\beta + \varepsilon, \quad (10)$$

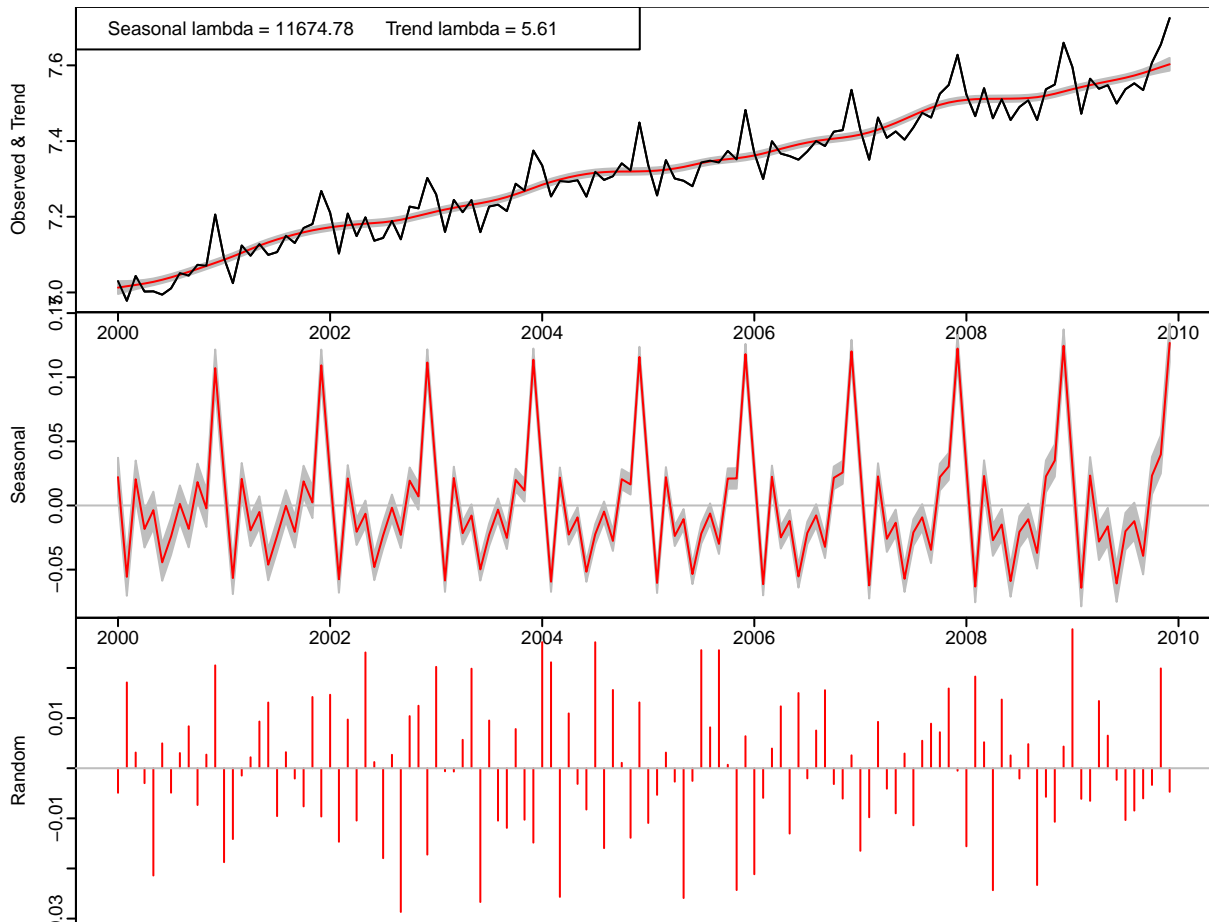
where  $y_{ext} = [y', 0']'$ ,  $\beta = [s', \ell']'$  is a vector of unknown coefficients,  $s$  is a vector of seasonal components,  $\ell$  is a vector containing the trend, and

$$X = \begin{bmatrix} Q & I \\ \lambda_{tt}D_{tt} & 0 \\ \lambda_{st}D_{st} & 0 \\ \lambda_{ss}D_{ss} & 0 \\ 0 & \lambda_{\ell}D_{\ell} \end{bmatrix}.$$

All errors are i.i.d.  $\mathcal{N}(0, \sigma^2)$  for some unknown  $\sigma$ .

Using (6) and (7) as in Section 4, we find the solution of this new problem and the corresponding confidence intervals. Using cross validation (Sections 4.2 and 9) and an optimisation procedure, we can find good values for  $\lambda_{tt}$ ,  $\lambda_{st}$ ,  $\lambda_{ss}$  and  $\lambda_{\ell}$ .

Let us consider an example of the described decomposition. Figure 3 shows the same data as discussed in Section 1, but now decomposed using STR. As we can see, the result of the decomposition is similar but confidence intervals are now provided for the trend and the seasonal component.



**Figure 3:** Supermarket and grocery stores log turnover in New South Wales decomposed with STR. The original data are plotted in black; decomposed trend, seasonal and random components are in red; confidence intervals are in grey.

## 6 STR with multiple seasonality and static, flexible and seasonal predictors

Let us revisit the concept of seasonality. Originally we considered it as a single repeating pattern which affects data in an additive manner. Such a pattern is usually a consequence of some scheduled or naturally repeating activity corresponding to the day of the week, hour of the day, etc. (see for example [Makridakis et al. \(2008\)](#)).

Nevertheless, time series representing real world activity are often affected by *several* schedules representing different seasonal periods. For example, electricity demand typically depends on the time of day, day of week, and day of year ([Hyndman and Fan, 2010](#)). However, in Turkey, electricity demand is also affected by the Islamic Hijri calendar due to various religious activities such as fasting during the month of Ramadan ([De Livera et al., 2011](#)). Another complicating

factor is that the periods of these calendars are fractional. More complex examples can include patterns with periods changing in length with time.

Regressors can also be important factors affecting time series beyond seasonal patterns. For example [Findley and Soukup \(2000\)](#), [Findley et al. \(2009\)](#) and [Hood and Feldpausch \(2011\)](#) consider the effect of various moving holidays on human activities. These effects are usually assumed to be linear and constant over time, although [Bell and Martin \(2004\)](#) considers time-varying coefficients. The magnitude of the impact of such regressors on a time series is important in further analysis and decision making in economics or finance.

In this section we consider time series that are affected by multiple seasonal patterns and linearly by regressors of three types. The first type is constant effect regressors where it is assumed the associated coefficients are constant over time. The second type is regressors with coefficients whose magnitudes change smoothly over time but where the coefficients do not exhibit any seasonal pattern. The third type is regressors with coefficients whose magnitudes change smoothly over time in a seasonal pattern. As far as we know, this last type of time-varying coefficient is new in the decomposition of time series.

We consider a time series  $Y_t$  consisting of the following parts:

$$Y_t = T_t + \sum_{i=1}^r S_{it} + \sum_{i=1}^m P_{it} + \sum_{i=1}^{\tilde{m}} \tilde{P}_{it} + \sum_{i=1}^{\tilde{\tilde{m}}} \tilde{\tilde{P}}_{it} + R_t, \quad (11)$$

where:

- $T_t$  is the smoothly changing trend;
- $S_{it}$  are seasonal components with corresponding seasonal periods  $k_i$ ;
- $P_{it}$  are additive components linearly depending on predictors  $p_i$  with constant coefficients;
- $\tilde{P}_{it}$  are additive components linearly depending on predictors  $\tilde{p}_i$  with time-varying but non-seasonal coefficients;
- $\tilde{\tilde{P}}_{it}$  are additive components linearly depending on predictors  $\tilde{\tilde{p}}_i$  with time-varying coefficients, where the coefficients have seasonal patterns with corresponding seasonal periods  $\tilde{\tilde{k}}_i$ ;
- $R_t$  is the “random” error.

Because this description is rather vague, we define decomposition and the model uniquely through a minimization problem which extends the optimisation problems and the corresponding models of Sections 3 and 5.

In this new optimisation problem defined below, we assume the existence of multiple seasonal periods and dependence on a number of regressors. We also assume that such dependence can change over time for some of the regressors (we call them flexible and seasonal regressors depending on whether the seasonal pattern appears in the changes).

$$\begin{aligned}
 (s, \ell, \beta, \tilde{\beta}, \tilde{\tilde{\beta}}) = \arg \min & \left\{ \left\| y - \sum_{i=1}^r Q_i s_i - \ell - P\beta - \sum_{i=1}^{\tilde{m}} \tilde{P}_i \tilde{\beta}_i - \sum_{i=1}^{\tilde{\tilde{m}}} \tilde{\tilde{P}}_i \tilde{\tilde{\beta}}_i \right\|_{L_2}^2 \right. \\
 & + \sum_{i=1}^r \left( \|\lambda_{tt_i} D_{tt_i} s_i\|_{L_2}^2 + \|\lambda_{st_i} D_{st_i} s_i\|_{L_2}^2 + \|\lambda_{ss_i} D_{ss_i} s_i\|_{L_2}^2 \right) + \|\lambda_{\ell} D_{tt} \ell\|_{L_2}^2 \\
 & \left. + \sum_{i=1}^{\tilde{m}} \|\tilde{\lambda}_i D_{tt} \tilde{\beta}_i\|_{L_2}^2 + \sum_{i=1}^{\tilde{\tilde{m}}} \left( \|\tilde{\tilde{\lambda}}_{tt_i} \tilde{\tilde{D}}_{tt_i} \tilde{\tilde{\beta}}_i\|_{L_2}^2 + \|\tilde{\tilde{\lambda}}_{st_i} \tilde{\tilde{D}}_{st_i} \tilde{\tilde{\beta}}_i\|_{L_2}^2 + \|\tilde{\tilde{\lambda}}_{ss_i} \tilde{\tilde{D}}_{ss_i} \tilde{\tilde{\beta}}_i\|_{L_2}^2 \right) \right\}, \quad (12)
 \end{aligned}$$

where

- $y$  is a vector of observations of length  $n$ ;
- $Q_i$  are  $n \times n(k_i - 1)$  matrices which compute observable seasonal elements from seasonal vectors  $s_i$ , each of which represents the corresponding two-dimensional seasonal component;
- $P$  is an  $n \times m$  matrix of static predictors, where every predictor occupies a single column;
- $\beta$  is an  $m$ -vector of coefficients of the static regressors;
- $\tilde{P}_i = \text{diag}(\tilde{p}_i)$  for  $1 \leq i \leq \tilde{m}$  is the  $i$ th predictor matrix with values of the  $i$ th predictor arranged along the diagonal and all other values equal to zero;
- $\tilde{\beta}_i$  is the  $i$ th vector of changing coefficients for the  $i$ th flexible regressor;
- $\tilde{\tilde{P}}_i = \text{diag}(\tilde{\tilde{p}}_i)$  for  $1 \leq i \leq \tilde{\tilde{m}}$  is the  $i$ th predictor matrix with values of the  $i$ th predictor arranged along the diagonal and all other values equal to zero;
- $\tilde{\tilde{\beta}}_i$  is the  $i$ th vector of changing coefficients for the  $i$ th seasonal regressor;
- $D_{tt}$  is a matrix taking second differences of a vector representing the trend or flexible coefficients;



- $D_{tt_i}$ ,  $D_{st_i}$  and  $D_{ss_i}$  are matrices taking second differences of the  $i$ th seasonal component in the time, time-season and season dimensions;
- $\tilde{D}_{tt_i}$ ,  $\tilde{D}_{st_i}$  and  $\tilde{D}_{ss_i}$  are matrices taking second differences of the  $i$ th seasonal component in time, time-season and season dimensions;
- $\lambda_{tt_i}$ ,  $\lambda_{st_i}$ ,  $\lambda_{ss_i}$ ,  $\lambda_\ell$ ,  $\tilde{\lambda}_i$ ,  $\tilde{\lambda}_{tt_i}$ ,  $\tilde{\lambda}_{st_i}$ ,  $\tilde{\lambda}_{ss_i}$  are the parameters.

We can also note that the optimisation problem can easily be adapted in case some (or many) observations  $y$  are missing (although the model does not allow the predictors to be missing for observed values of  $y$ ). The adaptation involves excluding missing values in vector  $y$  and the corresponding rows in the matrices which are in front of vectors  $s_1, \dots, s_r$ ,  $\ell$ ,  $\tilde{\beta}$ ,  $\tilde{\beta}_1, \dots, \tilde{\beta}_{\tilde{m}}$ ,  $\tilde{\beta}'_1, \dots, \tilde{\beta}'_{\tilde{m}}$  (assuming that in front of  $\ell$  there is an identity matrix  $I_n$ ) in expression (12).

This provides a very natural way to forecast by treating future observations as missing, and then estimating for forecast horizon  $h$ . If the covariance matrix of  $y$  can be estimated, the confidence intervals also can be found the standard for linear regression way.

Although this is a very generic form of STR, in practice it does not need to be that complex. In most cases, the minimization problem contains only a few of the terms in (12).

The optimisation problem (12) corresponds to the following linear model:

$$y_{ext} = X\beta + \varepsilon, \quad (13)$$

where  $y_{ext} = [y', 0']'$ ,

$$\beta = [s'_1, \dots, s'_r, \ell', \beta', \tilde{\beta}'_1, \dots, \tilde{\beta}'_{\tilde{m}}, \tilde{\beta}'_1, \dots, \tilde{\beta}'_{\tilde{m}}]'$$
 (14)

is a vector of unknown coefficients and

$$X = \begin{bmatrix}
 Q_1 & \dots & Q_r & I & R & \tilde{R}_1 & \dots & \tilde{R}_{\tilde{m}} & \tilde{R}_1 & \dots & \tilde{R}_{\tilde{m}} \\
 \lambda_{tt_1} D_{tt_1} & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\
 \lambda_{st_1} D_{st_1} & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\
 \lambda_{ss_1} D_{ss_1} & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\
 0 & \ddots & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\
 0 & \dots & \lambda_{tt_r} D_{tt_r} & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\
 0 & \dots & \lambda_{st_r} D_{st_r} & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\
 0 & \dots & \lambda_{ss_r} D_{ss_r} & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\
 0 & \dots & 0 & \lambda_\ell D_{tt} & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\
 0 & \dots & 0 & 0 & 0 & \tilde{\lambda}_1 D_{tt} & \dots & 0 & 0 & \dots & 0 \\
 0 & \dots & 0 & 0 & 0 & 0 & \ddots & 0 & 0 & \dots & 0 \\
 0 & \dots & 0 & 0 & 0 & 0 & \dots & \tilde{\lambda}_m D_{tt} & 0 & \dots & 0 \\
 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & \tilde{\lambda}_{tt_1} \tilde{D}_{tt_1} & \dots & 0 \\
 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & \tilde{\lambda}_{st_1} \tilde{D}_{st_1} & \dots & 0 \\
 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & \tilde{\lambda}_{ss_1} \tilde{D}_{ss_1} & \dots & 0 \\
 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \ddots & 0 \\
 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & \tilde{\lambda}_{tt_{\tilde{m}}} \tilde{D}_{tt_{\tilde{m}}} \\
 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & \tilde{\lambda}_{st_{\tilde{m}}} \tilde{D}_{st_{\tilde{m}}} \\
 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & \tilde{\lambda}_{ss_{\tilde{m}}} \tilde{D}_{ss_{\tilde{m}}}
 \end{bmatrix} \quad (15)$$

for some fixed parameters

$$\lambda = \left[ \lambda_{tt_1}, \lambda_{st_1}, \lambda_{ss_1}, \dots, \lambda_{tt_r}, \lambda_{st_r}, \lambda_{ss_r}, \lambda_\ell, \tilde{\lambda}_1, \dots, \tilde{\lambda}_{\tilde{m}}, \tilde{\lambda}_{tt_1}, \tilde{\lambda}_{st_1}, \tilde{\lambda}_{ss_1}, \dots, \tilde{\lambda}_{tt_{\tilde{m}}}, \tilde{\lambda}_{st_{\tilde{m}}}, \tilde{\lambda}_{ss_{\tilde{m}}} \right].$$

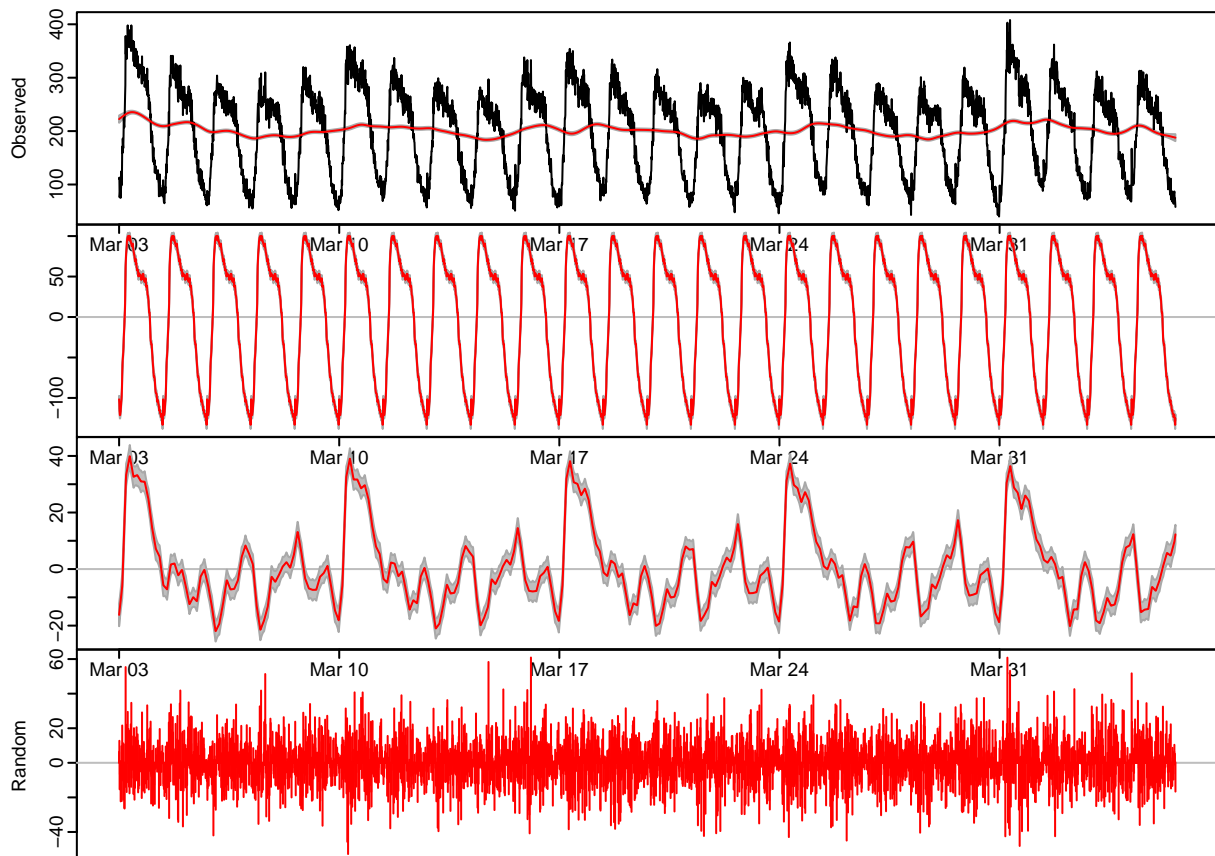
If some values of vector  $\lambda$  are zeros, the corresponding rows of matrix  $X$  can (and should) be removed (as they have no effect and removing them improves computation time). All errors are i.i.d.  $\mathcal{N}(0, \sigma^2)$  for some unknown  $\sigma$ .

We need to note that the combined number of coefficients (the length of  $\beta$  in equation (14)) that we need to estimate is usually much larger than the number of observations (the length of  $y$ ). This does not cause problems of estimation since the coefficients are regularised (restricted), and therefore the estimation is performed against observations presented by  $y_{ext}$  (regression (13)), where  $y_{ext}$  is longer than the number of estimated coefficients.

We can find the solution of this new problem and the corresponding confidence intervals using (6) and (7), similarly to the models discussed in Sections 4 and 6. Using the approach of Section 4.2, we can find good values for  $\lambda$ , although the task becomes more difficult as we need to optimize over a larger parameter space.

An example of decomposition of data with multiple seasonality is presented in Figure 4. The data was originally published in Weinberg et al. (2007). It consists of 10140 observations, although we only use the first 4056 observations (working days from 3 March 2003 to 3 April 2003) to demonstrate our method. The data has two seasonal patterns: the daily pattern has a period of 169 observations and the weekly pattern has a period of  $169 \times 5 = 845$  observations.

CV MSE = 209.7, Lambda predictors = (2.81e-05,0,0) (0,7.79,104) (30.6,262,2.73)



**Figure 4:** The number of calls to a big North American retail bank per 5-minute interval (from 7:00 a.m. to 9:05 p.m., weekdays) decomposed with STR. The original data are plotted in black; decomposed trend, seasonal and random components are in red; confidence intervals are in grey.

Another example, with multiple seasonality and time varying regressors can be found in Section 10.

## 7 Functional coefficient models

Multiple seasonality and multiple regressors (both flexible and seasonal) bring another problem: the length of  $\beta$  (defined in (14)) and the dimensions of  $X$  (defined in (15)) become too large for easy computation. This problem becomes more evident as the model we consider becomes more complex.

To solve this problem, let us consider approximating the varying trend  $\ell$  as a smoothly varying function of time using a linear combination of smooth basis functions. Thus we can write

$$\ell(t) = \sum_{j=1}^q \gamma_j \phi_j(t).$$

Then for some matrix  $\Phi$ , vector  $\ell$  can be presented as  $\ell = \Phi \gamma$ .

Similarly seasonal functional components  $s_i(t)$  can be defined and approximated using a linear combination of smooth basis functions:

$$s_i(t, s) = \sum_{j=1}^{q_i} \gamma_j^{(i)} \phi_j^{(i)}(t, s).$$

Then for some matrices  $\Phi^{(i)}$ , vectors  $s_i$  can be presented as  $s_i = \Phi^{(i)} \gamma^{(i)}$ .

Time varying and seasonally varying components of formula (11) can also be presented through functional components and functional coefficients in a similar way.

The basis functions here could be splines, Fourier terms (in such cases it is reminiscent of [Livera et al., 2010](#)), wavelets etc. In this article we use only piecewise linear regression splines.

Finally, noting that regularisation components can be written as norms of linear transformations of the corresponding gamma coefficients, the minimisation problem (12) can be rewritten in terms of functional coefficients:

$$\begin{aligned} (\gamma^{(\cdot)}, \gamma, \beta, \tilde{\gamma}, \tilde{\gamma}) = \arg \min & \left\{ \left\| y - \sum_{i=1}^r \Phi^{(i)} \gamma^{(i)} - \Phi \gamma - P \beta - \sum_{i=1}^{\tilde{m}} \tilde{\Phi}_i \tilde{\gamma}_i - \sum_{i=1}^{\tilde{m}} \tilde{\tilde{\Phi}}_i \tilde{\tilde{\gamma}}_i \right\|_{L_2}^2 \right. \\ & + \sum_{i=1}^r \left( \|\lambda_{tt_i} \Psi_{tt_i} \gamma_i\|_{L_2}^2 + \|\lambda_{st_i} \Psi_{st_i} \gamma_i\|_{L_2}^2 + \|\lambda_{ss_i} \Psi_{ss_i} \gamma_i\|_{L_2}^2 \right) + \|\lambda \ell \Psi_{tt} \gamma\|_{L_2}^2 \\ & \left. + \sum_{i=1}^{\tilde{m}} \|\tilde{\lambda}_i \Psi_{tt} \tilde{\gamma}_i\|_{L_2}^2 + \sum_{i=1}^{\tilde{m}} \left( \|\tilde{\tilde{\lambda}}_{tt_i} \tilde{\tilde{\Psi}}_{tt_i} \tilde{\tilde{\gamma}}_i\|_{L_2}^2 + \|\tilde{\tilde{\lambda}}_{st_i} \tilde{\tilde{\Psi}}_{st_i} \tilde{\tilde{\gamma}}_i\|_{L_2}^2 + \|\tilde{\tilde{\lambda}}_{ss_i} \tilde{\tilde{\Psi}}_{ss_i} \tilde{\tilde{\gamma}}_i\|_{L_2}^2 \right) \right\}, \quad (16) \end{aligned}$$

where

- the various  $\gamma$  vectors are functional coefficients which are used to represent the corresponding components of decomposition defined by formula (11);
- the various  $\Phi_i$  matrices transform the corresponding functional  $\gamma$  coefficients into corresponding components of representation (11);
- the various  $\Psi$  matrices allow the calculation of second derivatives as linear transformations of the corresponding functional coefficients;
- $P$  and  $\beta$  are defined in Section 6.

This approach leads to a reduction in the number of estimated coefficients to 3–4 times the length of the time series (see note in Section 6), and it dramatically improves the computational performance.

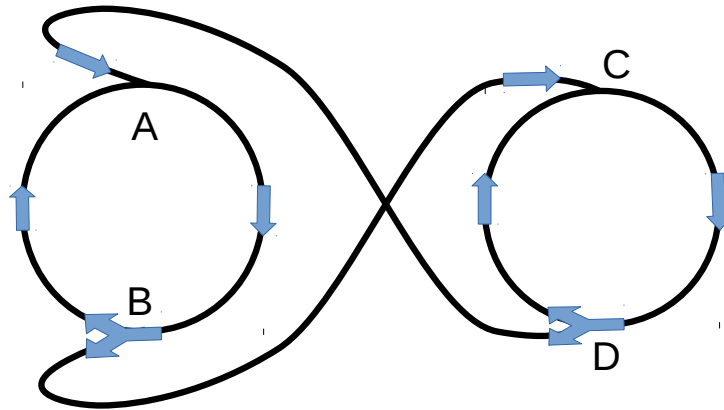
Interestingly, this approach directs us, with no additional effort, to a solution of another problem, namely seasonality with a fractional or varying period. As we can note from (16), seasonality is hidden in the  $\Psi$  matrices, which take second discrete seasonal derivatives of seasonal components. When seasonal components are functions, such matrices can be written for fractional seasonality, since even in this case second derivatives are linear functions of the corresponding functional coefficients.

## 8 Seasonality with complex topology

A seasonal two-dimensional surface can have a topology different from the “tube”. Again, our functional approach leads naturally to modelling seasonality with more complex topology. Let us again clarify with another example. Suppose we are going to model some social behaviour (electricity demand for instance) during working days and holidays, including weekends. The topology modelling the human behaviour is shown in Figure 5.

The left circle represents a working day, the right circle represents a holiday. They are connected by lines representing transition periods. Points A and C represent hour 0 of a day, points B and D represent hour 12. Every day has 24 hours and the Transition periods take 12 hours.

According to the diagram, a working day can follow a working day. Otherwise a working day can flow until hour 12 when transition period type 1 starts (line B – C), which goes for 12 hours and



**Figure 5:** *Topology of complex seasonality modelling human behaviour during working days, holidays and transition periods.*

traverses into a holiday (the right circle). Similarly, a holiday can follow a holiday. Otherwise a holiday flows until hour 12 when transition period type 2 starts (line D – A), which goes for 12 hours and traverses into a working day (the left circle).

Equivalently, the topology shown in Figure 5 can be described as two connected cylinders.

Differencing matrices  $D_{ss}$  and  $D_{st}$  (or  $\Psi_{ss}$  and  $Psi_{st}$ ) are defined in the standard way to restrict derivatives  $\frac{\partial^2}{\partial s^2}$  and  $\frac{\partial^2}{\partial s \partial t}$  for all data points except A, B, C and D. At points A, B, C and D, second derivatives can be restricted in various ways. One of them is to regularise derivatives twice: once for each path in the diagram. Matrices  $D_{ss}$  and  $D_{st}$  (or  $\Psi_{ss}$  and  $Psi_{st}$ ) should be defined accordingly to reflect such an approach.

Finally we should mention that the approaches of this Section and of the previous Section 7 do not change the form of matrix  $X$  (defined by equation (15)), although they change slightly the form of the sub-matrices which comprise  $X$ .

An example of decomposition with complex topology can be found in Section 10.

## 9 Robust STR: the model, solution, confidence and forecasting intervals, optimal smoothing parameters

The Robust STR (referred to as RSTR from hereon) uses a linear model which is identical to model (5) and its variations in the subsequent sections, apart from the error distribution; RSTR assumes a double exponential distribution for the residuals, and trend, seasonal and predictor coefficient changes, instead of the normal distribution assumed in the STR model. The double exponential distribution leads to a different minimization problem. In this case, problem (5) is

translated into the following (problems (10) and (13) are translated similarly):

$$(s, \ell) = \arg \min \left[ \|y - Qs - \ell\|_{L_1} + \left\| \lambda_s \Xi^{-\frac{1}{2}} D_s s \right\|_{L_1} + \left\| \lambda_t D_t \ell \right\|_{L_1} \right]$$

which (similar to the reduction in Section 4) can be written as a quantile regression:

$$\beta = \arg \min \|y_{ext} - X\beta\|_{L_1}, \quad (17)$$

where  $y_{ext}$ ,  $X$  and  $\beta$  are defined as before.

There is no known analytical solution for problem (17). We solve it numerically using quantile regression software (Koenker, 2013). The confidence intervals also cannot be expressed analytically and the following procedure is used to calculate them.

We follow the ideas from Dokumentov and Hyndman (2014) and reuse the following Monte-Carlo style algorithm in order to find  $p$ -confidence intervals for the coefficients  $\beta$  of problem (17).

1. Take  $m$  draws of vectors  $\delta_i$  of length of  $\ell$ , which have elements i.i.d.  $\mathcal{N}(0, \sigma_r^2)$ . We denote a set of  $m$  draws by  $\delta = \bigcup_{i=1}^m \{\delta_i\}$ .
2. Create a set of “distorted” observations  $y_\delta = y + \delta$ , then find a set of solutions  $\beta(y_\delta)$  for them.
3. For every  $1 \leq j \leq \text{length}(\ell)$ , the  $\left(\frac{p}{2}\right)$  and  $\left(1 - \frac{p}{2}\right)$  quantiles of the set  $\beta_j(y_\delta)$  will be the approximate  $p$ -confidence intervals for  $j$ th element of the solution.

It should be ensured that  $m$  is big enough to be able to calculate interval boundaries with the required level of precision.

The optimal smoothing parameters  $\lambda_s$  and  $\lambda_t$  are found using  $m$ -fold cross validation procedure.

For  $m$ -fold cross validation with gaps  $g \in \mathbb{N}$ , we split the data set into  $m$  subsets such that the observation at time  $1 \leq t$  belongs subset  $i$  ( $0 \leq i < m$ ) if and only if

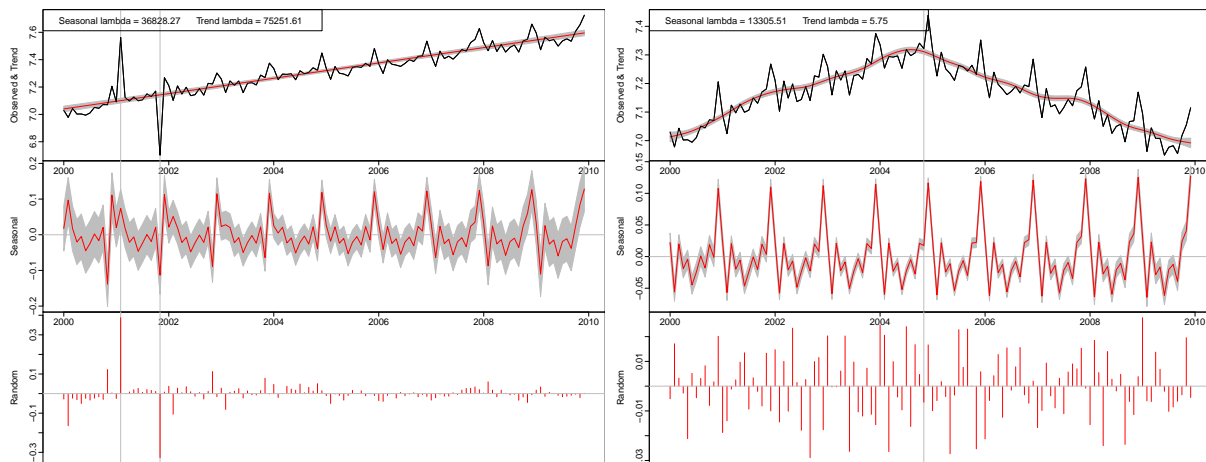
$$(t - 1) \bmod (mg) \in [ig, \dots, (i + 1)g - 1].$$

With  $g = 1$  this rule gives reasonable sparsity of the subsets and we speculate that the result of such  $m$ -fold cross validation will not differ much from the result of pure leave-one-out cross validation.

Although, to exclude situations when trend flexibility is exaggerated because of high correlation of nearest observation in the data and because only single observation are missed when  $g = 1$ ,  $g$  can be set comparable to number of observations in seasonal patterns. We experimented with  $g$  between 1 and 169 in different scenarios.

For optimisation, we use R and the method “optim” (Nelder-Mead) in the “stats” R package (R Core Team, 2015).

To show advantages of Robust STR we provide some examples which compare decomposition using STR and Robust STR on some artificial data. Let us create two datasets by spoiling data of Supermarket and grocery stores with outliers and abruptly changing the trend downwards. We will then decompose both new datasets with STR. The new data sets and the results of decomposition are shown in Figure 6.

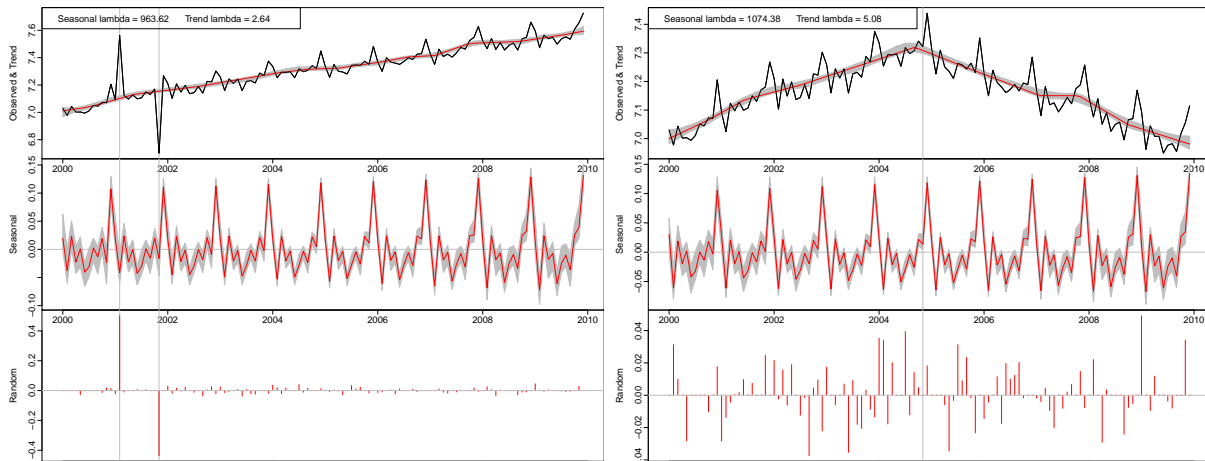


**Figure 6:** Supermarket and grocery stores log turnover in New South Wales distorted by outliers and trend change then decomposed with STR. The original data are plotted in black; the decomposed trend, seasonal and random components are in red; confidence intervals are in grey.

As we can see, outliers dramatically changed the results of the decomposition and the confidence intervals. Very high values of  $\lambda_s$  and  $\lambda_t$  show that STR degraded to a purely linear trend model with seasonal components also estimated linearly. For the second data set, the change in the trend was smoothed and other components were relatively unchanged by STR.

To make STR work better in the presence of outliers we used Robust STR. The results of decomposition using Robust STR for both distorted data sets are shown in Figure 7.

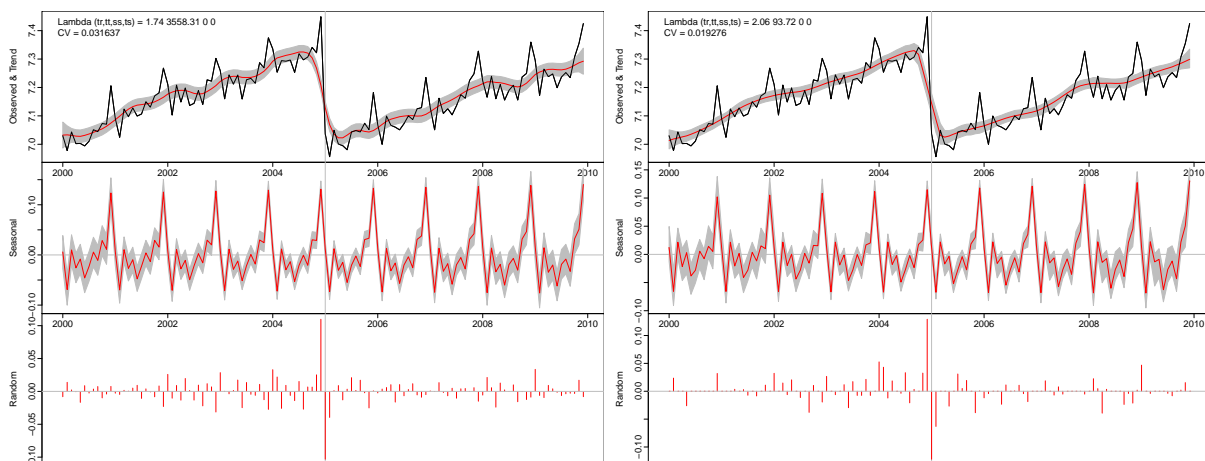




**Figure 7:** Supermarket and grocery stores log turnover in New South Wales distorted by outliers and trend change then decomposed with Robust STR. The original data is plotted in black; decomposed trend, seasonal and random components are in red; confidence intervals are in grey.

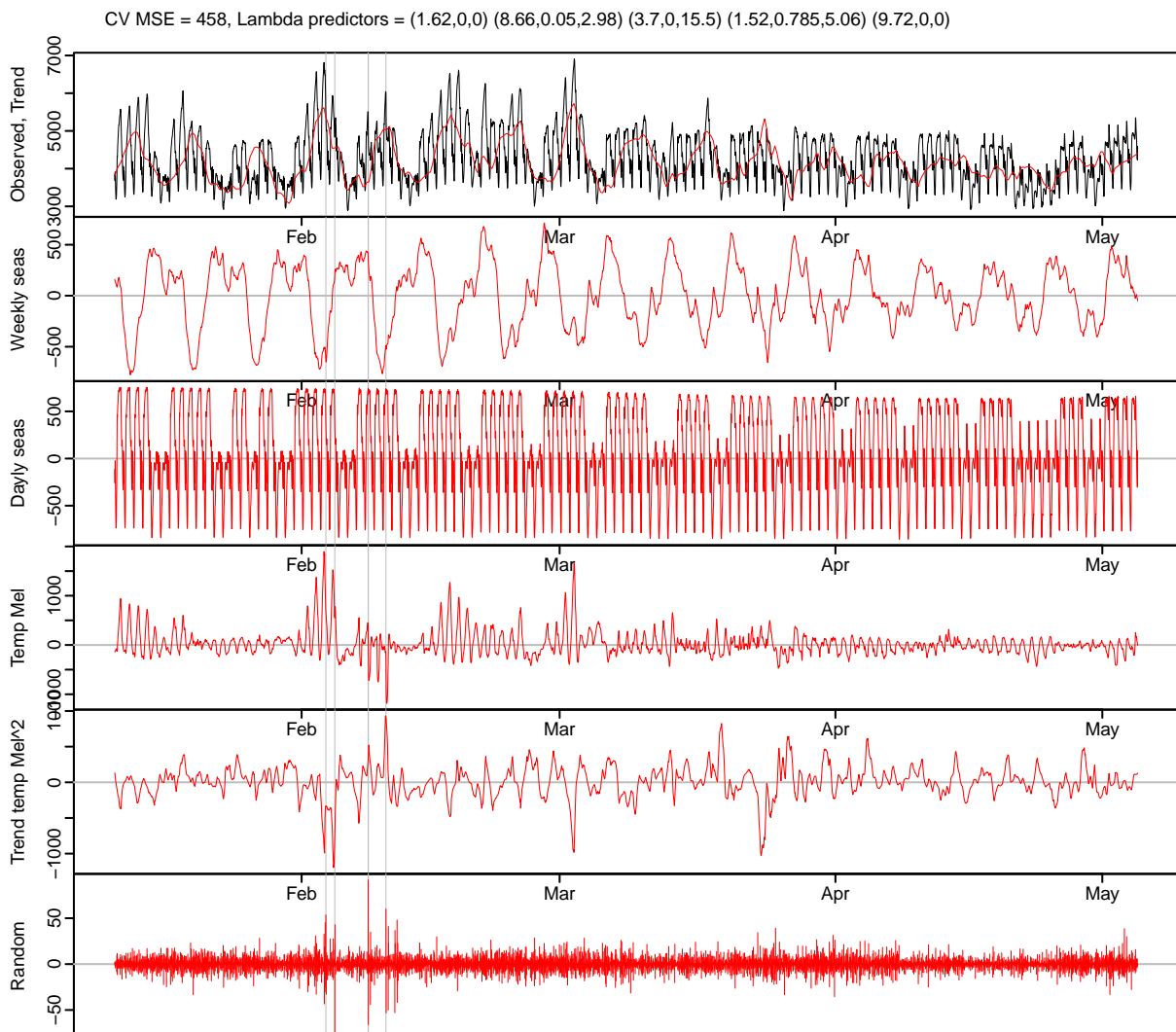
As we can see Robust STR works extremely well with outliers. It also finds the break of the trend quite well, although the confidence intervals are slightly wider than for STR. We speculate that this is due to the different distribution of errors assumed by the method.

Let us check performance of Robust STR on some other data. Let us change the data of Supermarket and grocery stores with a quick level change. Such level changes are called shifts. Figure 8 demonstrates that Robust STR deals rather well with abrupt shifts in the trend (See Section 11 for further discussion).



**Figure 8:** Supermarket and grocery stores log turnover in New South Wales distorted by an abrupt shift in the trend decomposed with STR and Robust STR. The original data is plotted in black; decomposed trend, seasonal and random components are in red; confidence intervals are in grey.

## 10 Example: electricity consumption with temperature predictors



**Figure 9:** Peak electricity demand in Victoria decomposed with STR. The original data is plotted in black; the decomposed trend is plotted in red over the data in the first panel; the subsequent windows contain the weekly seasonal pattern, daily seasonal pattern with complex topology, the effect of temperature in Melbourne which has a daily seasonal pattern with complex topology, the effect of squared temperatures in Melbourne which is allowed to vary over time, and the residuals. Five grey vertical lines (the third and fourth lines are very close) represent the residuals with the largest absolute values.

In this section we present one rather complicated example of time series decomposition. The data is electricity consumption in Victoria during 115 days starting on 10 January 2000. The data set comprises maximum electricity demand in Victoria during 30 minute periods (48 observations per day). In addition, for every 30 minute period, we have the concurrent value of the air temperature at the Melbourne City weather station. We use these temperatures and their squared values as predictors.

In this example we assume that the data set has two seasonal patterns. The first pattern is a weak seasonal pattern representing the specific demand features attributable to a particular day of the week. The second pattern is a daily seasonal pattern with topology of Figure 5. Such a topology is required to be able to distinguish working days and holidays/weekends and to be able to make transitions between them. The pattern reflects the tendency to have more electricity demand during standard working hours and lower demand at nights. It also reflects the tendency to have different demand patterns during working days and holidays/weekends.

Figure 9 shows the graph of the time series decomposed with STR. The  $\lambda$  coefficients were chosen semi-automatically (the starting point for the minimization procedure was chosen according to the previous experiments involving minimization of the same problem with fewer predictors). Five-fold cross validation for the optimal smoothing parameters yielded  $RMSE = 21.4$ .

Two seasonal patterns and two regressors are used for the decomposition. Therefore the data is represented as the sum of six components: trend, weekly seasonality, daily seasonality with a complex topology, temperature with a daily seasonality having a complex topology, squared temperature which is time-varying but non-seasonal, and the remainder.

The shape of all components is difficult to interpret since they affect each other in a rather complex manner (for example the trend and the seasonal components also play the role of an intercept for two predictors). Although difficult to interpret, the decomposition is still interesting from at least two points of view.

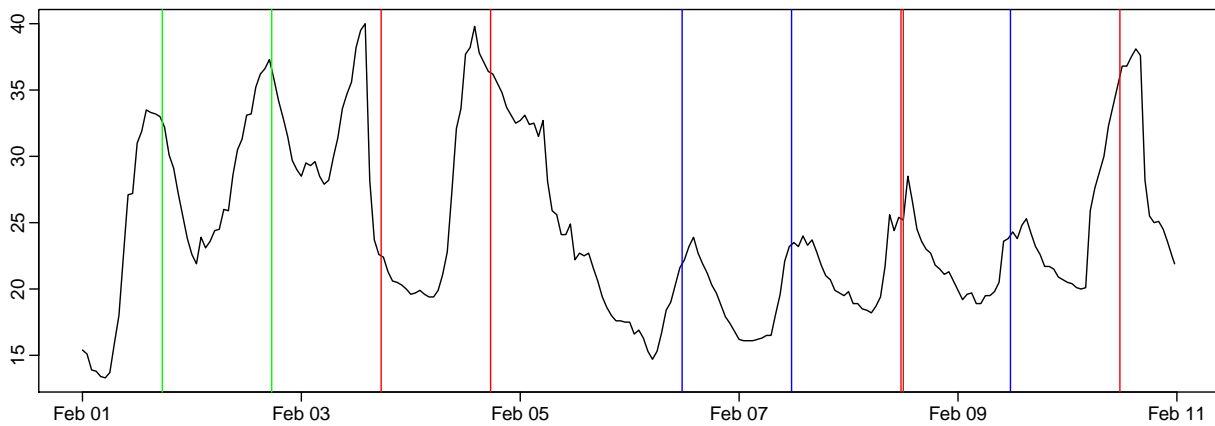
The first is the point of view of a forecaster, who might not be very interested in interpreting the internal structure, but will use the decomposition “as is” for prediction.

The second is the point of view of a researcher, who is interested in discovering events affecting (or correlated with) electricity consumption. In this case, investigation of the residuals of the decomposition can provide some light. For example, the five biggest residuals in absolute value (see Table 1 and the grey vertical lines in Figure 9) can be investigated.

Date	Period	Time period	Residual
3 February 2000, Thursday	36	17:30 – 18:00	53.6
4 February 2000, Friday	36	17:30 – 18:00	-73.9
8 February 2000, Tuesday	24	11:30 – 12:00	91.6
8 February 2000, Tuesday	25	12:00 – 12:30	-66.0
10 February 2000, Thursday	24	11:30 – 12:00	59.9

**Table 1:** Five biggest residuals in absolute value after a STR decomposition.

Melbourne is famous for its unstable weather. At least three of these five outliers can be explained by unusual weather during those days.



**Figure 10:** *The temperature in Melbourne starting from 1 February 2000 and during the following ten days. Four red lines and one brown line mark the times of the outliers. The green lines mark 5:30 pm (corresponding to the first two outliers) and the three blue lines mark 11:30 am.*

A positive outlier occurred at 5:30 pm on Thursday 3 February 2000, probably because it was end of one of the hottest days in Melbourne ( $40^{\circ}\text{C}$ ) followed by a sudden drop in temperature (see the first red line on Figure 10; the two previous green lines mark the same time on the previous two days). We can speculate that although the temperature dropped at 5:30 pm, the buildings stayed hot because they had heated up during the previous three very hot days and two nights. Therefore the electricity consumption was higher than expected by the model.

A negative outlier at 5:30 pm on Friday 4 February 2000 happened near the end of a very hot day in Melbourne ( $39.8^{\circ}\text{C}$ ) which was then followed by cooler weather (see the second red line on Figure 10). We can speculate that although the temperature did not drop (which made the model predict that electricity consumption will be high) the buildings did not require much cooling since the previous night was rather cold and probably also many people went out of building since it was a Friday night.

A positive outlier at 11:30 am on Thursday 10 February 2000 probably happened because of an early and extreme temperature rise in the beginning of the day. It might have led to a “shock” effect which was followed by a greater use of air conditioners and higher electricity consumption than anticipated by the model.

The outliers at 11:30 am and 12:00 pm (the first one which is positive, immediately followed by a negative one) are more difficult to explain by extreme temperature changes alone, although some sudden temperature jumps are observed on that day (8 February 2000). We could speculate

that some other weather or economic events happened that day, although currently we do not have a satisfactory explanation of these two outliers.

## 11 Concluding remarks and discussion

In this article we introduced a new approach for seasonal-trend decomposition and provided a few examples, including rather complex, for decomposition using this approach. We showed that the new method allows for multiple and complex seasonality, provides confidence intervals, finds smoothing parameters, and allows regressors to be taken into account with coefficients that are possibly time-varying and seasonal. We have also proposed a robust version of the method.

The main disadvantage of our proposed method is its slow speed in cases when many seasonal components or seasonal predictors are used. Although, as was mentioned in Section 6, this problem can be mostly overcome with the assumption that the seasonal components and coefficients for flexible predictors do not change quickly, so that rather sparse knots can be used to achieve good performance without compromising the quality of decomposition.

In Section 9 we demonstrated that the RSTR variation of our approach deals rather well with outliers and shifts in the trend. In spite of the presence of a few very big residuals and some sudden changes in trend direction, the overall separation of the data into trend, seasonal and random components remained good. Here we would like to note that such behaviour is attributable to some good properties of the  $L_1$  norm (and the double exponential distribution associated with it).

In Section 6 we showed that it is relatively easy to take smooth seasonality into account using terms that involve second differences. For example, in (9), we use the second term  $\|\lambda_{tt}D_{tt}s\|_{L_2}^2$  to calculate and restrict second derivatives along the time dimension. Interestingly, cyclicity (i.e., smooth aperiodic fluctuations) can also be taken into account using a similar technique. In this case, and using (9), the term responsible for cyclicity will be  $\|\lambda_{tt}^{(cyc)}(\alpha^2D_{tt} + I)\ell_{cyc}\|_{L_2}^2$ , where  $D_{tt}$  is the matrix that takes second differences along the time dimension using nearest observations and  $\alpha$  is the coefficient proportional to an average cycle length. We are going to investigate this idea further in future research.

De Livera et al. (2011) show that multiple seasonality can be identified in various ways. In this work we do not consider the identifiability problem, assuming that any seasonal representation which minimises cross validated error will suit.

The  $L_1$  and  $L_2$  norms used in minimization problems can be mixed according to different assumptions on the distributions of the residuals of the model, and assumptions about trend or seasonal component changes. In such cases, the minimization problem can be reduced to the LASSO minimization problem.

A mixture of norms can be useful, for example, in cases when trend and seasonal patterns are smooth and also the noise component has outliers. It can also be useful when the trend changes abruptly, but the seasonal components are smooth, and the noise is distributed normally or, at least, has no outliers. In all such cases the mixture of norms can lead to better performance of the model.

Finally we need to note that our new approach allows us to deal with shifts (rapid changes in the level of the data) separately, treating them as another component along with trend, seasonal, random and cyclic parts. The main difference between, for example, the trend component and this new shift component is how they are regularised in the corresponding minimization problems: while the second differences are regularised for the trend component, for the shift component the first differences are regularised instead (in both cases using  $L_1$  norm). We plan to investigate this approach further in our subsequent research.

## References

- Anderson, V. O. and Nochmals, U. (1914). The elimination of spurious correlation due to position in time or space. *Biometrika*, pages 269–279.
- Bell, W. R. and Martin, D. E. (2004). Modeling time-varying trading-day effects in monthly time series. In *Proceedings of the Joint Statistical Meetings*, pages 8–12.
- Burman, J. P. (1980). Seasonal adjustment by signal extraction. *Journal of the Royal Statistical Society. Series A (General)*, pages 321–337.
- Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I. (1990). Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3–73.
- Commandeur, J. J. F., Koopman, S. J., and Ooms, M. (2011). Statistical software for state space methods. *Journal of Statistical Software*, 41(1):1–18. <http://www.jstatsoft.org/v41/i01>.
- Copeland, M. T. (1915). Statistical indices of business conditions. *The Quarterly Journal of Economics*, pages 522–562.
- Dagum, E. B. (1988). *The X11ARIMA/88 Seasonal Adjustment Method: Foundations and User's Manual*. Statistics Canada, Time Series Research and Analysis Division.
- De Livera, A. M., Hyndman, R. J., and Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106(496):1513–1527.
- Dokumentov, A. and Hyndman, R. J. (2013). Two-dimensional smoothing of mortality rates. Technical report, Monash University. <http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2013/wp26-13.pdf>.
- Dokumentov, A. and Hyndman, R. J. (2014). Low-dimensional decomposition, smoothing and forecasting of sparse functional data. Technical report, Monash University. <http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2014/wp16-14.pdf>.
- Findley, D. and Soukup, R. (2000). Modeling and model selection for moving holidays. *Proceedings of the Business and Economic Statistics Section of the American Statistical Association, Alexandria: American Statistical Association Alexandria*, pages 102–107.
- Findley, D. F. (2005). Some recent developments and directions in seasonal adjustment. *Journal of Official Statistics*, 21(2):343.

- Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C., and Chen, B.-C. (1998). New capabilities and methods of the X-12-ARIMA seasonal-adjustment program. *Journal of Business & Economic Statistics*, 16(2):127–152.
- Findley, D. F., Monsell, B. C., and Hou, C.-T. (2009). Stock series holiday regressors generated by flow series holiday regressors. *Statistics*, (04).
- Harvey, A. and Peters, S. (1990). Estimation procedures for structural time series models. *Journal of Forecasting*, 9(2):89–108.
- Harvey, A. C. (1985). Trends and cycles in macroeconomic time series. *Journal of Business & Economic Statistics*, 3(3):216–227.
- Harvey, A. C. (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge university press.
- Hood, C. C. H. and Feldpausch, R. (2011). Experiences with user-defined regressors for seasonal adjustment. Technical report, Catherine Hood Consulting. [http://www.catherinehood.net/papers/chood\\_asa2011.pdf](http://www.catherinehood.net/papers/chood_asa2011.pdf).
- Hooker, R. H. (1901). The suspension of the berlin produce exchange and its effect upon corn prices. *Journal of the Royal Statistical Society*, 64(4):574–613.
- Hyndman, R. J. and Fan, S. (2010). Density forecasting for long-term peak electricity demand. *Power Systems, IEEE Transactions on*, 25(2):1142–1153.
- Koenker, R. (2013). *quantreg: Quantile Regression*. <http://cran.r-project.org/package=quantreg>.
- Ladiray, D. and Quenneville, B. (2001). Seasonal adjustment with the X-11 method. *Lecture notes in statistics*, 158:220.
- Livera, A. M. D., Hyndman, R. J., and Snyder, R. D. (2010). Forecasting time series with complex seasonal patterns using exponential smoothing. (October).
- Macauley, F. R. (1930). The smoothing of time series. *National Bureau of Economic Research*, pages 121–136.
- Makridakis, S., Wheelwright, S. C., and Hyndman, R. J. (2008). *Forecasting methods and applications*. John Wiley & Sons.



- Makridakis, S. G., Wheelwright, S. C., and Hyndman, R. J. (1998). *Forecasting: methods and applications*. John Wiley & Sons, New York, 3rd edition. <http://robhyndman.com/forecasting/>.
- McElroy, T. S. (2010). A nonlinear algorithm for seasonal adjustment in multiplicative component decompositions. *Studies in Nonlinear Dynamics & Econometrics*, 14(4).
- Monsell, B. C. and Aston, J. A. (2003). Toward x-13? Technical report, U. S. Census Bureau. <http://www.census.gov/ts/papers/jsm2003bcm.pdf>.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge University Press, New York.
- Seber, G. A. F. and Lee, A. J. (2003). *Linear regression analysis*. John Wiley & Sons, second edition.
- Shishkin, J., Young, A. H., and C, M. J. (1967). The x-11 variant of the census ii method seasonal adjustment program. *Bureau of the Census, Technical Paper*, (15).
- Shiskin, J. (1957). Electronic computers and business indicators. In *Electronic Computers and Business Indicators*. National Bureau of Economic Research.
- Spencer, J. (1904). On the graduation of the rates of sickness and mortality presented by the experience of the Manchester Unity of Oddfellows during the period 1893-97. *Journal of the Institute of Actuaries*, 38(4):334–343.
- Weinberg, J., Brown, L. D., and Stroud, J. R. (2007). Bayesian forecasting of an inhomogeneous poisson process with applications to call center data. *Journal of the American Statistical Association*, 102(480):1185–1198.