



INFORMS Journal on Data Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

STR: Seasonal-Trend Decomposition Using Regression

Alexander Dokumentov, Rob J. Hyndman

To cite this article:

Alexander Dokumentov, Rob J. Hyndman (2022) STR: Seasonal-Trend Decomposition Using Regression. INFORMS Journal on Data Science 1(1):50-62. <https://doi.org/10.1287/ijds.2021.0004>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2021, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

STR: Seasonal-Trend Decomposition Using Regression

Alexander Dokumentov,^a Rob J. Hyndman^b
^aLet's Forecast, Parkdale, Victoria 3195, Australia; ^bDepartment of Econometrics & Business Statistics, Monash University, Clayton, Victoria 3800, Australia

Contact: alexander.dokumentov@letsforecast.com (AD); rob.hyndman@monash.edu,  <https://orcid.org/0000-0002-2140-5352> (RJH)

Received: November 17, 2020

Revised: April 4, 2021

Accepted: June 2, 2021

Published Online in Articles in Advance:
December 23, 2021

<https://doi.org/10.1287/ijds.2021.0004>

Copyright: © 2021 INFORMS

Abstract. We propose a new method for decomposing seasonal data: a seasonal-trend decomposition using regression (STR). Unlike other decomposition methods, STR allows for multiple seasonal and cyclic components, covariates, seasonal patterns that may have non-integer periods, and seasonality with complex topology. It can be used for time series with any regular time index, including hourly, daily, weekly, monthly, or quarterly data. It is competitive with existing methods when they exist and tackles many more decomposition problems than other methods allow. STR is based on a regularized optimization and so is somewhat related to ridge regression. Because it is based on a statistical model, we can easily compute confidence intervals for components, something that is not possible with most existing decomposition methods (such as seasonal-trend decomposition using Loess, X-12-ARIMA, SEATS-TRAMO, etc.). Our model is implemented in the R package *str*, so it can be applied by anyone to their own data.

History: Nick Street served as the senior editor for this article.

Funding: This work was supported by the Australian Research Council [Grant CE140100049].

Data Ethics & Reproducibility Note: All data included are publicly available with no data ethics issues. The code capsule is available on Code Ocean at <https://doi.org/10.24433/CO.2481805.v1>, at https://github.com/robjhyndman/STR_paper, and in the e-Companion to this article (available at <https://doi.org/10.1287/ijds.2021.0004>)

Keywords: time series decomposition • seasonal data • Tikhonov regularization • ridge regression • LASSO • STL • TBATS • X-12-ARIMA • BSM

1. Time Series Decomposition

Decomposition of univariate time series data is a common tool used by government and business in order to understand time series variation. A time series decomposition method separates a time series into components, typically trend, seasonality, and “remainder.” These can be used in many ways for decision making:

1. It helps separate and understand trend and seasonal effects within a time series. This allows a greater understanding of peak times and an estimation of the relative effect of holidays, working days, and week-ends. These can then be used in resource planning.
2. It allows the estimation of seasonally adjusted data, which can then be used to assess the health of a business or industry after adjusting for periodic effects. This is done, for example, with labor market statistics as the seasonally adjusted unemployment rate is used by governments in assessing the state of the economy.
3. It can be used to identify anomalies by eliminating the predictable variation resulting from trend and seasonality, leaving the unpredictable variation (the remainder). Any anomalies in the data are revealed in this remainder component when they are often hidden in the original series. Anomaly detection helps organizations identify problems that may need attention. For

example, many organizations measure millions of time series at high frequency, and some anomaly detection is essential to identify toward what analysts should address their attention.

Existing time series decomposition methods are designed for monthly and quarterly data with few tools available for more frequent data. The seasonal-trend decomposition using Loess (STL) procedure of Cleveland et al. (1990) is the only widely available decomposition tool for data observed more frequently than monthly, but it assumes a very simple additive structure and allows no covariates. This makes it unsuitable for a large number of problems.

For example, electricity companies measure hourly electricity demand patterns for their customers. Hourly electricity demand data typically shows time-of-day, day-of-week, and time-of-year patterns. These seasonalities interact; the demand pattern on weekdays is different from the demand pattern on weekends, for example. A further complicating issue is that, on public holidays, demand tends to look more like weekend patterns than weekday patterns, thus perturbing the periodicity of the patterns. The demand is strongly driven by temperatures (because of the widespread use of electric heating and cooling). Any

time series decomposition for electricity demand must allow for all of these effects. However, no existing time series decomposition methods handle this problem despite this combination of features being extremely common in any human data recorded at a subdaily frequency.

1.1. A Brief History of Time Series Decomposition

The first attempts to decompose time series into various components can be dated to the 19th century, when Poynting (1884) proposed price averaging as a tool for eliminating trend and seasonal fluctuations. Later, his approach was extended by Hooker (1901), Spencer (1904), and Anderson and Nochmals (1914). Further research in this direction includes that by Copeland (1915), who was the first to attempt to extract the seasonal component until Macaulay (1931) proposed a method that gradually became “classical.” The work of Macaulay led to the Census II method (Shiskin 1957), which became popular among national statistics offices. The Census II method has continued to evolve with additional features, including robustness, calendar effects, covariates, ARIMA extensions, and extensive diagnostics. Widely used versions of this approach are X-11 (Shiskin et al. 1967), X-11-ARIMA (Dagum 1988, Ladiray and Quenneville 2001), X-12-ARIMA (Findley et al. 1998), and X-13-ARIMA-SEATS (Findley 2005, Dagum and Bianconcini 2016).

A different approach was followed by Cleveland et al. (1990), who developed STL, which has come to be widely used outside the national statistics agencies, partly because of its availability in R (R Core Team 2020). 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.

Several model-based methods for seasonal decomposition have been developed, including the TRAMO/SEATS procedure, which was developed at the Bank of Spain (Gómez and Maravall 2001), the TBATS model of De Livera et al. (2011), and various structural time series model approaches (Harvey 1990, Commandeur et al. 2011). One big advantage of using a model for seasonal decomposition and adjustment is that it provides a natural way to compute confidence and prediction intervals.

Despite this long history and the availability of many algorithms and models for time series decomposition, there are many time series characteristics that are not addressed in these approaches. We seek a time series decomposition method with the following attributes:

- Provides a meaningful and simple statistical model.
- Allows for computation of confidence intervals of components.
- Can take account of covariates.

- Allows for fractional seasonal periods (e.g., with weekly data).
- Allows for multiple seasonal periods (e.g., with daily or subdaily data).
- Allows for complex seasonal topology (e.g., because of public holidays).
- Can incorporate covariates that interact with the seasonality.

None of the existing methods satisfies all of these requirements. We aim to fill this gap with our new approach, which is clear, general, model-based, robust (if required), and simple. We show that the problem of seasonal decomposition can be recast in the framework of ordinary least squares or quantile regression. Moreover, our approach provides new features (such as covariates that affect data in a complex seasonal manner and the ability to model complex seasonality) that have not been developed before. Our new seasonal-trend decomposition using regression (STR) method is the most general framework for the decomposition of seasonal data that is available at present.

2. STR Models

We suppose our time series y_t can be decomposed additively as follows:

$$y_t = T_t + \sum_{i=1}^I S_t^{(i)} + \sum_{p=1}^P \phi_{p,t} z_{t,p} + R_t, \quad (1)$$

where

- T_t is a smoothly changing trend.
- $S_t^{(i)}$ are smoothly changing seasonal components with possibly complex topology.
- $z_{p,t}$ are covariates with coefficients $\phi_{p,t}$ that may be time-varying and even seasonal.
- R_t is the remainder.

The seasonal component is assumed to have a repeating pattern that changes very slowly or is constant over time. The trend component describes the smooth underlying mean of the data (after accounting for seasonality). The remainder component (which we assume to be uncorrelated) contains only noise and idiosyncratic patterns in the data.

When components behave multiplicatively, we can use a log or Box–Cox transformation (Box and Cox 1964) to made the additive assumption more reasonable.

The total number of coefficients to be estimated is usually much larger than the number of observations. Consequently, we impose some regularization on the coefficient estimates. In fact, without some regularization, the trend, seasonal components, and time-changing coefficients are unidentifiable.

The innovation in our approach is to provide an estimation method that allows these regularizations to be imposed in an efficient manner using matrix differencing operators, which convert the estimation

method to a linear model. In this way, the estimation becomes analogous to ridge regression. Although the matrices involved can become very large, they are also sparse, so sparse matrix algebra methods can be used to reduce the computational burden.

2.1. Smoothness Constraints via Difference Operators

Smoothness of a function implies that its second derivatives are small because, otherwise, large values of the second derivatives would lead to high curvature of the function making it “wiggly” (see Wood 2006). Thus, we consider a function smooth if its second derivatives are normally distributed with zero mean and when the variance controls the degree of smoothness.

Because time is discrete, we approximate derivatives with differences. Here, we use the notation Δ_2 for the double difference operator, $\Delta_2 = (1 - B)^2$, where B is the backshift operator.

So a smooth trend is obtained by requiring that $\Delta_2 T_t = T_t - 2T_{t-1} + T_{t-2}$ are independent and identically distributed (iid) $\mathcal{N}(0, \sigma_L^2)$, where σ_L controls the degree of smoothness. This can be expressed via the multivariate normal density

$$f(D_\ell \ell) = f_L(\sigma_L) \exp \left\{ -\frac{1}{2} \|D_\ell \ell / \sigma_L\|_{L_2}^2 \right\},$$

where $\ell = \langle T_t \rangle_{t=1}^n$ is an n -vector of the trend component, D_ℓ is the $(n-2) \times n$ difference operator matrix such that $D_\ell \ell = \langle \Delta^2 T_t \rangle_{t=3}^n$, and $f_L(\sigma_L)$ is a normalizing function.

Note that the model subsumes the case of a linear trend, obtained when $\sigma_L^2 = 0$.

2.2. Smooth Two-Dimensional Seasonal Surfaces

Let m_i denote the number of “seasons” in the seasonal component $S_i^{(i)}$. For example, if y_t is monthly data, there is only one seasonal component ($I=1$) with $m_1 = 12$; if y_t is daily data, there may be two seasonal components ($I=2$) with a weekly pattern ($m_1 = 7$) and an annual pattern ($m_2 = 365$ ignoring leap years). Let us also define the function $\kappa_i(t) = t \bmod m_i$, which transforms time t into the corresponding season $\kappa_i(t) \in \{1, \dots, m_i\}$.

At time t , we observe only one element of each seasonal component. It is reasonable to ask what the other elements of the component are at that moment t . For example, if we have daily data with weekly seasonality, it is valid to ask on a Wednesday what the current value of the seasonal component corresponding to Friday is? In other words, we define latent components that are responsible for seasons other than $\kappa_i(t)$. In this way, we treat the i th seasonal component as a two-dimensional matrix, $[S_{k,t}^{(i)}]$ ($k = 1, \dots, m_i; t = 1, \dots, n$). This formulation of each seasonal component as a two-dimensional structure is one of the novelties and

features of STR, and it allows for simple and natural smoothing constraints to be applied.

The seasonal two-dimensional surface actually has the topology of a cylinder, with which dimension t (time) is along the length of the cylinder, but dimension s (season) is “circular” around the cylinder. For example, in the case of a weekly seasonal component in daily data, the season dimension corresponds to days of the week. A time series corresponds to a spiral on the cylinder. The other locations on the cylinder (between the spirals) represent “imaginary” seasonality. These correspond, for example, to the current value of Friday’s seasonality when today is Wednesday.

To ensure they remain identifiable and interpretable, the seasonal terms must sum to zero at any time t , so they must satisfy the property $\sum_k S_{k,t}^{(i)} = 0$ for each t .

Smoothness of the weekly seasonal pattern can be described as smoothness of $S_{k,t}^{(i)}$ in the seasonal direction k , and the speed at which the seasonality changes over time is controlled by the smoothness of $S_{k,t}^{(i)}$ in the time direction t . Moreover, some other features are possible to express. For example, by restricting second partial derivatives of $S_{k,t}^{(i)}$ in both time and season directions, it is possible to describe the weekly seasonal pattern changing “synchronously” in the time–season direction.

The seasonal smoothness can also be expressed using difference operator matrices. We define three such matrices, $D_{tt,i}$, $D_{ss,i}$, and $D_{st,i}$, corresponding to smoothness in the time, season, and time–season directions, respectively.

For smoothness in the time direction, we require that the seasonal vectors $\langle \Delta_2 S_{k,t}^{(i)} \rangle_{k=1}^{m_i}$ are iid $\mathcal{N}(0, \sigma_i^2 \Sigma_i)$, where σ_i^2 controls the degree of smoothness and $\Sigma_i = I_{m_i} - \frac{1}{m_i} \mathbf{1}_{m_i} \mathbf{1}_{m_i}^T$ is the covariance matrix of the m_i random variables $\xi_k = \eta_k - m_i^{-1} \sum_{r=1}^{m_i} \eta_r$ obtained from the iid standard normal random variables $\eta_1, \dots, \eta_{m_i}$.

Let $S^{(i)} = [S_{k,t}^{(i)}]$ contain the i th seasonal surface, S_i^- be the matrix $S^{(i)}$ without the last row, and let $s_i = \text{vec}(S_i^-)$ be a vector of length $n(m_i - 1)$ representing the i th seasonal component in vector form. Thus, we can write the second differences of the seasonal component along the time dimension as $D_{tt,i} s_i = \langle \Delta_t^2 S_{k,t}^{(i)} \rangle_{k=1}^{m_i}$, $3 \leq t \leq n$ for a suitably defined $(n-2)m_i \times nm_i$ difference operator matrix $D_{tt,i}$.

Then, the density for the second derivatives of the i th seasonal component is given by

$$f(s_i) = f_{tt,i}(\sigma_i) \exp \left\{ -\frac{1}{2} \|D_{tt,i} s_i / \sigma_i\|_{L_2}^2 \right\},$$

where $f_{tt,i}(\sigma_i)$ is a normalizing function.

Smoothness in the season and time–season directions are defined analogously by restricting the derivatives $\frac{\partial^2}{\partial s^2}$ and $\frac{\partial^2}{\partial s \partial t}$. It should be noted that the matrices D_{st} and D_{ss} take the cylinder topology into account when calculating proper circular differences.

Again, by setting the smoothing parameters to zero, some simple special cases are obtained. Setting σ_i to zero eliminates the i th seasonal component (as it is equal to zero), setting the variance in the time direction to zero causes the seasonal component to be periodic. By setting both trend and time-seasonal variances to zero we effectively force our model to have a linear trend with seasonal dummies. Setting the variance of the second derivatives of the seasonal component in the time direction to zero alone makes changes in the seasonal component linear over time, but does not preclude the seasonal pattern from changing.

2.3. STR with Complex Seasonal Topology

A seasonal two-dimensional surface can have a topology other than a cylinder. Suppose that we are going to model some social behavior (e.g., electricity demand) during working days and holidays (including weekends). The topology modeling the human behavior is shown in Figure 1.

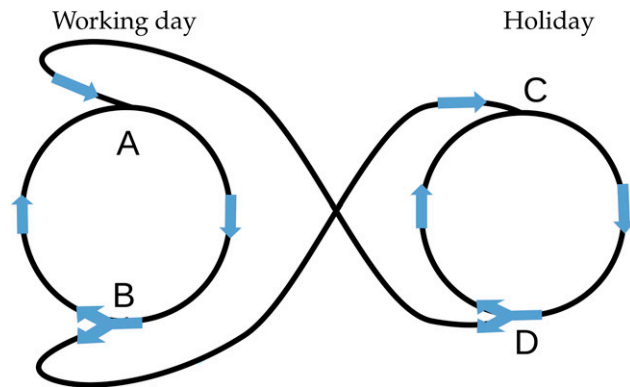
The left circle represents a working day; the right circle represents a holiday. They are connected by lines representing transition periods. Points A and C represent hour 0 of a day, and points B and D represent hour 12. Every day has 24 hours, and the transition periods take 12 hours.

According to the diagram, a working day can either be followed by another working day or flow until hour 12 and then undergo a 12-hour transition (line B–C) into a holiday. Similarly, a holiday can either be followed by another holiday or undergo a 12-hour transition (line D–A) into a working day.

Equivalently, the topology shown in Figure 1 can be described as two connected cylinders.

The differencing matrices D_{ss} and D_{st} are defined as before for all data points except A, B, C and D. At points A, B, C and D, the second derivatives can be restricted in various ways. One possibility is to define

Figure 1. (Color online) Topology of the Complex Seasonality Modeling Human Behavior During Working Days, Holidays, and Transition Periods



D_{ss} and D_{st} to regularize the derivatives twice, once for each path in the diagram.

An example of a decomposition with complex topology can be found in Section 4.2.

2.4. Gaussian Remainders

The remainder terms R_t are assumed to be iid $\mathcal{N}(0, \sigma_R^2)$. Let $\mathbf{y} = [y_1, \dots, y_n]'$ be an n -vector of observations, $\mathbf{Z} = [z_{t,p}]$ be a matrix of covariates with coefficients given by $\Phi = [\phi_{p,t}]$, and let \mathbf{Q}_i be an $n \times n(m_i - 1)$ matrix that extracts $\langle S_{\kappa(t),t}^{(i)} \rangle_{t=1}^n$ from the vector \mathbf{s}_i .

Then, the residuals $\mathbf{r} = \mathbf{y} - \sum_i \mathbf{Q}_i \mathbf{s}_i - \ell - \mathbf{Z}\Phi$ have probability density function

$$f(\mathbf{r}) = f(\sigma_R) \exp \left\{ -\frac{1}{2} \|\mathbf{r}/\sigma_R\|_{L_2}^2 \right\},$$

where $f(\sigma_R)$ is a normalizing function.

2.5. Covariates with Time-Varying Coefficients

Covariates are often available to help explain some of the variation seen in a time series. For example, Findley et al. (2012) consider the effects of various moving holidays on human activities, and Bell and Martin (2004) consider covariates with time-varying coefficients.

We consider time series that are affected by covariates of three types: (1) static covariates for which we assume that the associated coefficients are constant over time, (2) flexible covariates with coefficients whose magnitudes change smoothly over time but for which the coefficients do not exhibit any seasonal pattern, and (3) seasonal covariates 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.

To distinguish the three types of covariates, we write

$$\begin{aligned} \mathbf{Z}\Phi &= \sum_{p=1}^P \phi_{p,t} z_{t,p} = \sum_{h=1}^H \alpha_h a_{t,h} + \sum_{j=1}^J \beta_j b_{t,j} + \sum_{k=1}^K \gamma_k c_{t,k} \\ &= \mathbf{A}\alpha + \sum_{j=1}^J \mathbf{B}_j \beta_j + \sum_{k=1}^K \mathbf{C}_k \gamma_k, \end{aligned}$$

where

- $a_{t,h}$ are static covariates with constant coefficients.
- $b_{t,j}$ are flexible covariates with time-varying but nonseasonal coefficients.
- $c_{t,k}$ are seasonal covariates with time-varying coefficients for which the coefficients have seasonal patterns with corresponding seasonal periods $p_k \in \{m_1, \dots, m_J\}$.
- $\mathbf{A} = [a_{t,h}]$ is an $n \times H$ matrix of static covariates in which every covariate occupies a single column.
- α is an H -vector of coefficients of the static covariates.

- $B_j = \text{diag}(b_j)$ and $b_j = [b_{1,j}, \dots, b_{n,j}]'$.
- β_j is the j th n -vector of changing coefficients for the j th flexible covariate.
- $C_k = \text{diag}(c_k)$ and $c_k = [c_{1,k}, \dots, c_{n,k}]'$.
- γ_k is the k th n -vector of changing coefficients for the k th seasonal covariate.

To ensure the time-varying coefficients change slowly over time, we impose smoothness via differencing operators, much as we did for the trend and seasonal components. For this purpose, we define matrices $\Delta_{tt,k}$, $\Delta_{st,k}$, and $\Delta_{ss,k}$ that take second differences of the k th seasonal coefficients in the time, time–season, and season dimensions, respectively.

2.6. Maximum Likelihood Estimation

The parameters of the model are given by $\alpha, \beta_1, \dots, \beta_J, \gamma_1, \dots, \gamma_K, \ell, s_1, \dots, s_I$ and the standard deviations $\sigma_R, \sigma_L, \sigma_1, \dots, \sigma_I$ along with those corresponding to the other smoothness constraints.

The model is overparameterized with more parameters than observations. However, the regularization via the smoothness constraints allow it to be estimated.

Combining the preceding results and assuming that all terms are independent of each other and that the standard deviations are known, we find that the minus log likelihood function for this model (up to a constant) is

$$\begin{aligned}
 -\log \mathcal{L} = & \frac{1}{2\sigma_R^2} \left\{ \left\| y - \sum_{i=1}^I Q_i s_i - \ell - A\alpha - \sum_{j=1}^J B_j \beta_j \right. \right. \\
 & - \sum_{k=1}^K C_k \gamma_k \left. \right\|_{L_2}^2 + \lambda_\ell \|D_\ell \ell\|_{L_2}^2 \\
 & + \sum_{i=1}^I \left(\|\lambda_{tt,i} D_{tt,i} s_i\|_{L_2}^2 + \|\lambda_{st,i} D_{st,i} s_i\|_{L_2}^2 \right. \\
 & \left. + \|\lambda_{ss,i} D_{ss,i} s_i\|_{L_2}^2 \right) \\
 & + \sum_{j=1}^J \|\theta_j D_{tt} \beta_j\|_{L_2}^2 \\
 & + \sum_{k=1}^K \left(\|\theta_{tt,k} \Delta_{tt,k} \gamma_k\|_{L_2}^2 + \|\theta_{st,k} \Delta_{st,k} \gamma_k\|_{L_2}^2 \right. \\
 & \left. \left. + \|\theta_{ss,k} \Delta_{ss,k} \gamma_k\|_{L_2}^2 \right) \right\}, \quad (2)
 \end{aligned}$$

where the λ and θ coefficients are ratios of standard deviations. Then maximum likelihood estimates can be obtained by minimizing (2) over

$$\eta = [s'_1, \dots, s'_I, \ell', \alpha', \beta'_1, \dots, \beta'_J, \gamma'_1, \dots, \gamma'_K]'. \quad (3)$$

Note that (2) corresponds to the minus log likelihood function for the linear model

$$y_+ = X\eta + \varepsilon, \quad (4)$$

where $y_+ = [y', 0']'$ is the vector y padded with zeros, η is a vector of unknown coefficients (with seasonal

components adjusted by removing the last row of seasonal observations), $\varepsilon \sim N(0, \sigma_R^2 I)$, and

$$X = \begin{bmatrix}
 Q_1 & \dots & Q_I & I_n & A & B_1 & \dots & B_J & C_1 & \dots & C_K \\
 \lambda_{tt,1} & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\
 D_{tt,1} & & & & & & & & & & \\
 \lambda_{st,1} & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\
 D_{st,1} & & & & & & & & & & \\
 \lambda_{ss,1} & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\
 D_{ss,1} & & & & & & & & & & \\
 0 & \ddots & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\
 0 & \dots & \lambda_{tt,I} & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\
 & & D_{tt,I} & & & & & & & & \\
 0 & \dots & \lambda_{st,I} & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\
 & & D_{st,I} & & & & & & & & \\
 0 & \dots & \lambda_{ss,I} & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\
 & & D_{ss,I} & & & & & & & & \\
 0 & \dots & 0 & \lambda_\ell & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\
 & & & D_{tt} & & & & & & & \\
 0 & \dots & 0 & 0 & 0 & \theta_1 & \dots & 0 & 0 & \dots & 0 \\
 & & & & & D_{tt} & & & & & \\
 0 & \dots & 0 & 0 & 0 & 0 & \ddots & 0 & 0 & \dots & 0 \\
 0 & \dots & 0 & 0 & 0 & 0 & \dots & \theta_m & 0 & \dots & 0 \\
 & & & & & & & D_{tt} & & & \\
 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & \theta_{tt,1} & \dots & 0 \\
 & & & & & & & & \Delta_{tt,1} & & \\
 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & \theta_{st,1} & \dots & 0 \\
 & & & & & & & & \Delta_{st,1} & & \\
 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & \theta_{ss,1} & \dots & 0 \\
 & & & & & & & & \Delta_{ss,1} & & \\
 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \ddots & 0 \\
 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & \theta_{tt,K} \\
 & & & & & & & & & & \Delta_{tt,K} \\
 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & \theta_{st,K} \\
 & & & & & & & & & & \Delta_{st,K} \\
 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & \theta_{ss,K} \\
 & & & & & & & & & & \Delta_{ss,K}
 \end{bmatrix} \quad (5)$$

with fixed parameters

$$\Lambda = \{\lambda_{tt,1}, \lambda_{st,1}, \lambda_{ss,1}, \dots, \lambda_{tt,r}, \lambda_{st,r}, \lambda_{ss,r}, \lambda_\ell, \theta_1, \dots, \theta_J, \theta_{tt,1}, \theta_{st,1}, \theta_{ss,1}, \dots, \theta_{tt,K}, \theta_{st,K}, \theta_{ss,K}\}.$$

If some values of Λ are zeros, the corresponding rows of matrix X can (and should) be removed as they have no effect, and removing them improves the computation time.

The total number of coefficients (the length of η) that we need to estimate is usually much larger than the number of observations (the length of y). This does not cause computational problems because the coefficients are regularized via smoothness constraints, and

the estimation is performed using (4), where y_+ is longer than η .

Using standard linear regression results, the maximum likelihood solution is given by

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

This has covariance matrix

$$\text{Cov}(\hat{\eta}) = \sigma_R^2(X'X)^{-1}. \quad (7)$$

The trend component \hat{T}_t and seasonal components $\hat{S}_t^{(1)}, \dots, \hat{S}_t^{(l)}$ and their corresponding confidence intervals can be obtained directly from $\hat{\eta}$ and $\text{Cov}(\hat{\eta})$.

2.7. Smoothing Parameter Estimation

We use cross-validation for estimating the λ and θ parameters of (5). Because the model in (4) is a linear model, the leave-one-out cross-validation residuals can be calculated (see Seber and Lee 2003) using

$$e_{(i)} = \frac{y_i - \hat{y}_i}{1 - h_{ii}}, \quad (8)$$

where y_i is the i th element of the vector y , \hat{y}_i is the i th element of the 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 in linear regression (see, for example, Ruppert et al. 2003):

$$\text{CV} = \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{1 - h_{ii}} \right)^2. \quad (9)$$

STR finds the optimal λ and θ smoothing parameters by minimizing CV. The problem of minimizing CV can be complex, for example, if there are many local minima, and we have no method that guarantees the finding of the global minimum. However, rather simple methods often work well in practice. We perform such optimization using the Nelder–Mead method as implemented in the `optim()` function from the `stats` package in R (R Core Team 2020).

In cases in which numerical difficulties may arise in using leave-one-out cross-validation, we resort to K -fold cross-validation in which we split the data set into K subsets such that the observation at time t belongs to subset i if and only if $(t-1) \bmod (Kg) \in [ig, \dots, (i+1)g-1]$.

When $g = 1$, no consecutive observations lie in the same subset of data. This gives a reasonable sparsity of the subsets, and we speculate that the resulting K -fold cross-validation does not differ much from the result of the leave-one-out cross-validation. However, in this case, the trend component usually absorbs any serial correlation that may be present in the data.

For $g > 1$, blocks of g consecutive observations are selected within each subset, similar to a block-bootstrap method but with a greater separation of neighboring

blocks. The value of g determines the amount of correlation that is allowed to remain in the residuals and, therefore, controls the trade-off between correlated residuals and trend flexibility.

In some cases, we observe decompositions in which the seasonal component “leaks” into the trend component. This can be prevented by selecting an appropriate value for g . Setting g equal to the number of observations in seasonal patterns is an effective choice as it prevents the trend component from absorbing seasonal patterns with periods less than g .

3. Simulations

To demonstrate that our method estimates the appropriate time series components, we use simulations in which we know the true underlying components. We compare our results against two other decomposition methods that handle multiple seasonal components, STL and TBATS. STL (Cleveland et al. 1990) is a well-known decomposition method and, therefore, a natural choice to use as a method to compare. The original methodology and implementation were designed for only one seasonal period. However, the implementation in the forecast package for R (Hyndman et al. 2021) computes an STL decomposition iteratively with multiple seasonal periods. TBATS (De Livera et al. 2011) is a forecasting method designed for multiple seasonalities; it provides a decomposition of the time series into components and so can provide an alternative comparison.

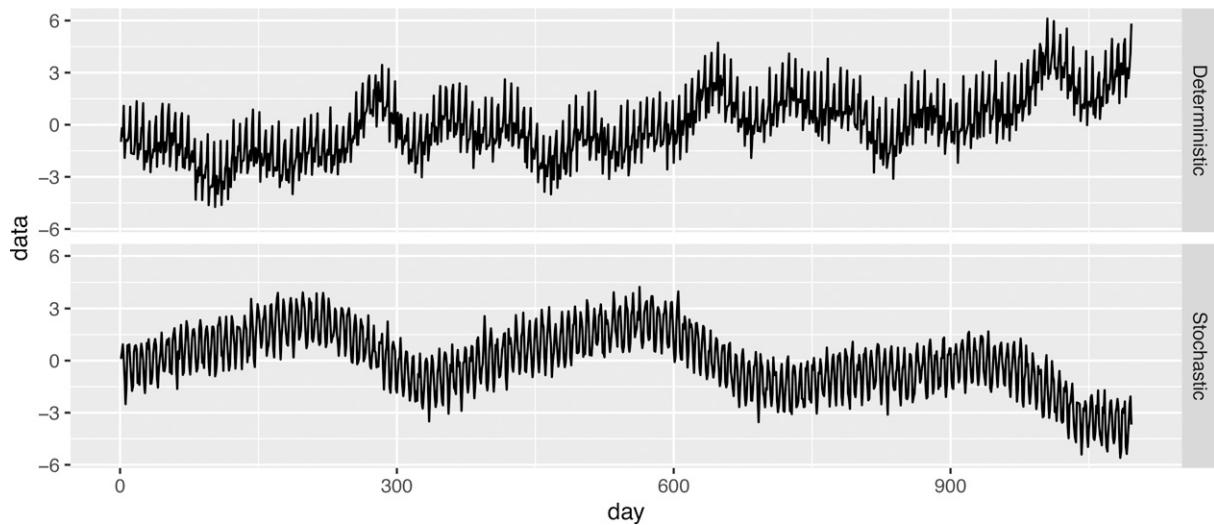
Our simulated daily data consists of four components:

$$y_t = T_t + \alpha S_t^W + \beta S_t^Y + \gamma R_t, \quad t = 1, \dots, n, \quad (10)$$

where T_t is a trend, S_t^W is a weekly seasonal component; S_t^Y is a yearly seasonal component; R_t is the remainder; and α , β , and γ are parameters that control the contribution of the components to y_t .

We use two data-generating processes (DGPs) to simulate data. For the first DGP, we use a quadratic trend function with random coefficients such that $T_t^* = N_1(t + n/2(N_2 - 1))^2$, where N_1 and N_2 are independent $N(0,1)$ random variables. Then, T_t^* is normalized to give T_t with mean zero and unit variance. The seasonal components consist of five pairs of Fourier terms with random $N(0,1)$ coefficients, which are then also normalized. Because this DGP has deterministic components, we refer to it as the “deterministic” DGP.

For the second “stochastic” DGP, the trend is given by an ARIMA(0,2,0) model with standard normal errors, which is then normalized. The seasonal components are also smooth and obtained using a similar process constrained to be periodic. Specifically, for each seasonal component, we start with a vector of iid $i(0,1)$ values of the length of one season, normalized

Figure 2. Examples of Time Series Generated by the Two DGPs

and replicated to be of length n . This is then integrated and normalized twice to make it smooth. So each seasonal component is as smooth as the trend, is centered on zero, and is periodic.

Figure 2 shows examples of simulated time series from each of the two DGPs with $\alpha = \beta = 1$ and $\gamma = 0.25$. We generated 1,096 days in each case (one day more than three nonleap years of daily data).

For comparison of the three decomposition methods, we use parameter sets $(\alpha, \beta, \gamma) = \{(1, 1, 0.2), (1, 1, 0.4), (1, 1, 0.6)\}$ and create 20 time series for each DGP and parameter set. We apply STR, STL, and TBATS to each series and compute the root-mean-square error (RMSE) for each component (averaging across all data sets and

days). We measure the significance of the differences between the methods using a linear model applied to the squared errors. The results are shown in Table 1. It is clear that STR outperforms STL in estimating all components other than the trend and that STR outperforms TBATS in estimating all components other than the weekly seasonal pattern. In no case do the other methods outperform STR.

4. Applications

We illustrate our proposed method using two data sets. The first is a simple application with monthly data that allows us to compare our results with existing

Table 1. RMSE over 100 Simulations for Each DGP and Value of γ

| γ | Method | Trend RMSE | Weekly RMSE | Yearly RMSE | Remainder RMSE |
|--------------------------|--------|---------------|---------------|---------------|----------------|
| Stochastic DGP | | | | | |
| 0.2 | STL | 0.0735 | 0.0564 | 0.1272 | 0.1495 |
| 0.2 | TBATS | 0.1686 | 0.0169 | 0.1598 | 0.0579 |
| 0.2 | STR | 0.0176 | 0.0150 | 0.0510 | 0.0544 |
| 0.4 | STL | 0.0798 | 0.1149 | 0.2295 | 0.2653 |
| 0.4 | TBATS | 0.3172 | 0.0344 | 0.3044 | 0.1182 |
| 0.4 | STR | 0.0315 | 0.0306 | 0.0840 | 0.0929 |
| 0.6 | STL | 0.0872 | 0.1713 | 0.3385 | 0.3887 |
| 0.6 | TBATS | 0.5194 | 0.0516 | 0.4929 | 0.1852 |
| 0.6 | STR | 0.0505 | 0.0447 | 0.1347 | 0.1470 |
| Deterministic GDP | | | | | |
| 0.2 | STL | 0.1279 | 0.0576 | 0.1387 | 0.1760 |
| 0.2 | TBATS | 0.1849 | 0.0170 | 0.1803 | 0.0675 |
| 0.2 | STR | 0.0552 | 0.0143 | 0.0700 | 0.0535 |
| 0.4 | STL | 0.1129 | 0.1147 | 0.2366 | 0.2763 |
| 0.4 | TBATS | 0.3232 | 0.0338 | 0.3146 | 0.1157 |
| 0.4 | STR | 0.0541 | 0.0311 | 0.0930 | 0.0945 |
| 0.6 | STL | 0.1362 | 0.1716 | 0.3428 | 0.3995 |
| 0.6 | TBATS | 0.4834 | 0.0509 | 0.4698 | 0.1692 |
| 0.6 | STR | 0.0793 | 0.0426 | 0.1304 | 0.1329 |

Note. Bold values indicate results that are significantly different from the STR values.

methods. The second is more complicated and demonstrates the full range of capabilities of our method, including multiple seasonal patterns, complex seasonal topology because of public holidays, and interacting seasonalities with a nonlinear covariate.

4.1. Monthly Supermarket Revenue

The time series shown in the top panel of Figure 3 concerns logarithms of monthly supermarket and grocery store revenue in New South Wales, Australia, from 2000 to 2009.

We apply the model in (4) to the logged data with no covariates and one seasonal component ($m_1 = 12$) with the smoothing parameters selected using leave-one-out cross-validation. The resulting decomposition is shown in Figure 3 along with 95% confidence intervals shown as shaded bands. Here, the trend

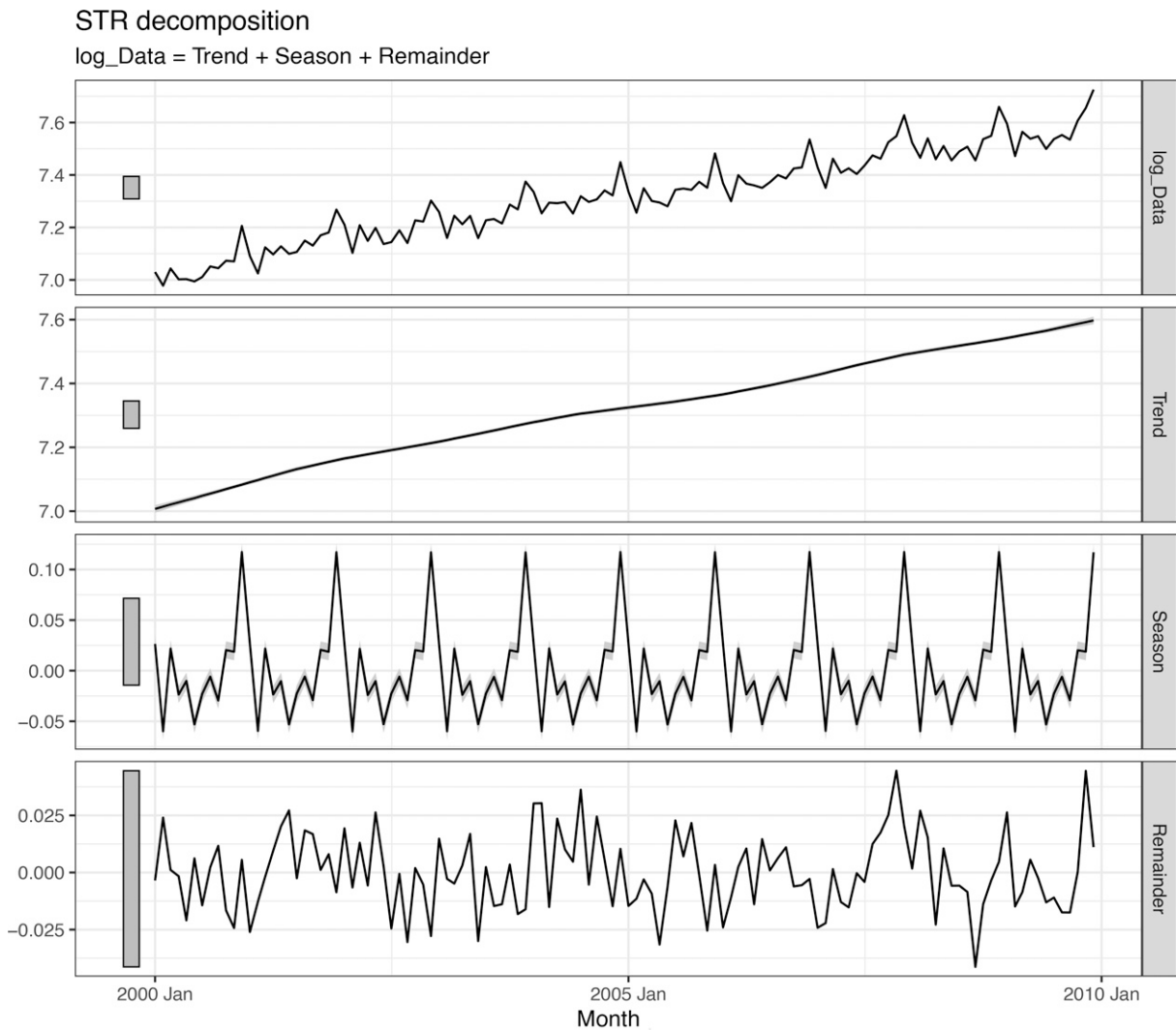
component measures the overall health and size of the economy, the seasonal component allows the study of how human behavior changes over time, and the remainder component allows us to identify unusual features of the data.

We obtain similar results using other decomposition methods applied to the same data, including STL (Cleveland et al. 1990), TBATS (De Livera et al. 2011), and X-13-ARIMA-SEATS (Findley 2005, Dagum and Bianconcini 2016). These comparisons are provided in the online appendix.

4.2. Half-Hourly Electricity Consumption with Temperature Predictors

We illustrate the power of the STR approach by presenting a more complicated example of time series decomposition using half-hourly electricity consumption

Figure 3. Supermarket and Grocery Store Log Revenue in New South Wales Decomposed with STR



Note. The original data (logged) are shown in the top panel, and the components are shown in the other panels with 95% confidence intervals for the trend and season components as shaded regions.

in the state of Victoria, Australia, during the 115 days starting on January 10, 2000.

For each 30-minute period, we also have the air temperature at the Melbourne weather station (near the center of the largest city in Victoria). We use concurrent temperatures and their squared values as predictors. Because the effect of temperature can change over time (a mild day in summer has a different effect from a day of the same temperature in winter), we allow for changing coefficients over time.

This data set has two seasonal patterns. The first pattern is a weekly seasonal pattern ($m_1 = 48 \times 7 = 336$) that represents the specific demand features that are attributable to a particular day of the week. The second pattern is a daily seasonal pattern ($m_2 = 48$) with the topology in Figure 1, which allows the model to distinguish between working days and holidays/weekends and to make transitions between them. The pattern reflects the tendency to have a higher electricity demand during standard working hours and a lower demand at night. It also reflects the tendency to have different demand patterns on working days and holidays/weekends. A longer series would also have an annual seasonal pattern, but with only 115 days, this cannot be distinguished from the trend component.

Figure 4 shows the time series decomposed using STR. The λ coefficients were chosen semiautomatically (the starting point for the minimization procedure was chosen according to experiments involving the minimization of the same problem with fewer predictors). Fivefold cross-validation with a gap length of $g = 336$ (one week) for the optimal smoothing parameters yielded the optimal mean squared error.

Two seasonal patterns and two regressors are used for the decomposition. Thus, the data are represented as the sum of six components: trend, weekly seasonality, daily seasonality with a complex topology (workday and non-workday seasonality, including transition periods), temperature, squared temperature (which is time-varying but nonseasonal), and the remainder. It has to be emphasized that the second seasonal component represents workday and non-workday seasonal patterns and the patterns of the transition periods between them as a single structure. To the best of our knowledge, no other seasonal decomposition method has this powerful ability. Another unique STR feature, illustrated in this decomposition, is the use of flexible regressors (components four and five). We could also have used a more complex approach with seasonal regressors (which is another unique feature of STR).

It is interesting to look carefully at the remainder term as it contains evidence of events that affect (or are correlated with) electricity consumption but that have not been captured in the model. The 10 residuals that are largest in absolute value are shown in Table 2, corresponding to the gray vertical lines in Figure 4.

Melbourne is famous for its unstable weather, and the outliers can be explained by unusual weather on those days. Five positive outliers occurred, starting from 2:30 p.m. on Thursday, February 3, 2000, probably because it was one of the hottest days in Melbourne (40°C) but with a sudden drop in temperature (see the first five red lines in Figure 5). We can speculate that, although the temperature dropped at around 2:30 p.m., the buildings stayed hot because they had heated up during the very hot previous three days and two nights. Thus, the electricity consumption was higher than that expected by the model, which only accounts for the contemporaneous temperature.

A negative outlier appeared at 4:00 p.m. on the following day, which was also very hot (39.8°C) (see the separate single red line in Figure 5). It is very unusual to have a hot day following a cool night in Melbourne. In this case, the buildings did not require as much air-conditioning as usual because of the cooling effect of the previous night. Consequently, the model overestimated the electricity consumption at that time.

The last four negative outliers started at 2:30 p.m. on Thursday, February 10, 2000, exactly one week later than the first four outliers. The explanation is similar to the previous case: it was a very hot day (38.1°C) after a series of rather cool days and nights, and the model overestimated electricity consumption.

The model can be modified to better handle these situations by using lagged as well as concurrent temperatures along with the change in temperatures over recent periods.

5. Extensions and Discussion

Because we have used a linear modeling framework, it is very easy to extend our approach in many directions, exploiting the vast literature on linear regression models. We highlight a few possible extensions here.

5.1. Functional Coefficients

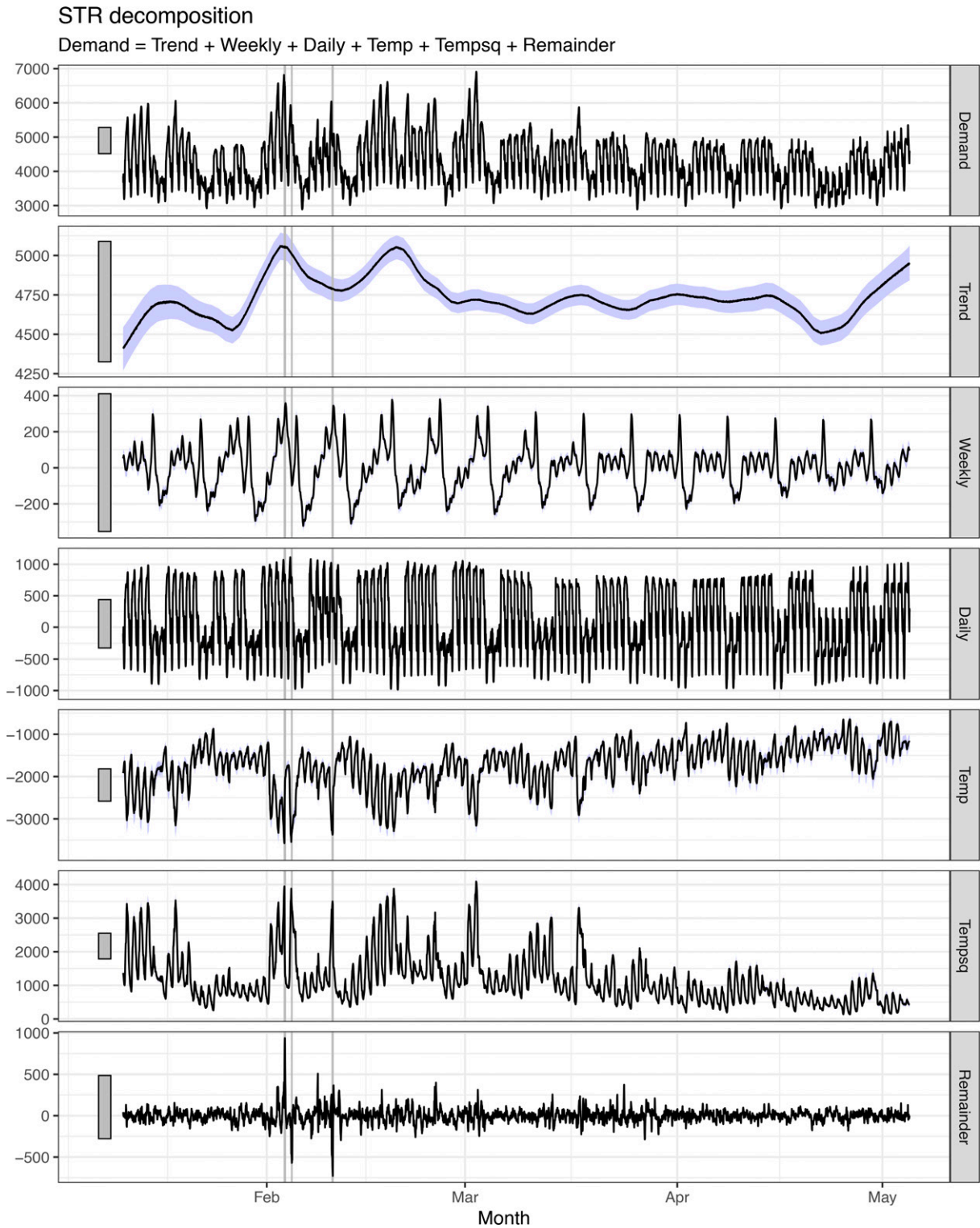
We can reduce the dimension of the model by using a linear combination of smooth basis functions (such as splines) for the trend:

$$T_t = \sum_{j=1}^q v_j \psi_j(t), \quad (11)$$

so that $\ell = \Psi \nu$, where ν contains coefficients to be estimated. Then, no difference operators are needed to ensure smoothness. Instead, the value of q and shape of the basis functions ψ_j determine the smoothness.

Similarly, seasonal functional components can be obtained, using Fourier terms, for example (reminiscent of De Livera et al. 2011). This approach leads, with no additional effort, to handling seasonality with a fractional or varying period.

Figure 4. (Color online) Electricity Demand in Victoria Decomposed Using STR



Notes. The original data are plotted in the first panel with the decomposed trend plotted in the second panel. The subsequent panels contain the weekly seasonal pattern; the 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. Gray vertical lines represent the residuals with the largest absolute values.

Table 2. The 10 Residuals That Are Largest in Absolute Value After an STR Decomposition

| Date | Day of week | Time period | Residual |
|----------------------|-------------|---------------|----------|
| February 3, 2000 | Thursday | 14:30 – 15:00 | 687.3 |
| 10 February 3, 2000 | Thursday | 15:00 – 15:30 | 871.3 |
| 10 February 3, 2000 | Thursday | 15:30 – 16:00 | 937.0 |
| 10 February 3, 2000 | Thursday | 16:00 – 16:30 | 767.2 |
| 10 February 3, 2000 | Thursday | 16:30 – 17:00 | 580.4 |
| 10 February 4, 2000 | Friday | 16:00 – 16:30 | –569.7 |
| 10 February 10, 2000 | Thursday | 14:30 – 15:00 | –611.4 |
| 10 February 10, 2000 | Thursday | 15:00 – 15:30 | –658.9 |
| 10 February 10, 2000 | Thursday | 15:30 – 16:00 | –695.7 |
| 10 February 10, 2000 | Thursday | 16:00 – 16:30 | –728.5 |

5.2. Robust Regularization

If we replace the L_2 norm used in (2) with the L_1 norm, we obtain a robust variant of STR. Equivalently, we replace the normality assumption with a double exponential distribution in the expression for the log-likelihood.

This can be written as a quantile regression (Koenker 2005):

$$\hat{\eta} = \arg \min_{\eta} \|y^+ - X\eta\|_{L_1}, \quad (12)$$

where y^+ , X , and η are defined as in (4). This can be solved numerically using a quantile regression algorithm (Koenker 2020).

The L_1 and L_2 norms can also be mixed according to different assumptions regarding the distributions of the model components. 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 in which the trend and seasonal patterns are smooth but 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).

5.3. Correlated Residuals

A variation of (4) allows the errors to be correlated with covariance matrix Σ_{ϵ} . Then, the maximum likelihood estimates are given by the generalized least squares solution

$$\hat{\eta} = (X' \Sigma_{\epsilon}^{-1} X)^{-1} X' \Sigma_{\epsilon}^{-1} y^+, \quad (13)$$

with

$$\text{Cov}(\hat{\eta}) = (X' \Sigma_{\epsilon}^{-1} X)^{-1}. \quad (14)$$

5.4. Level Shifts and Interventions

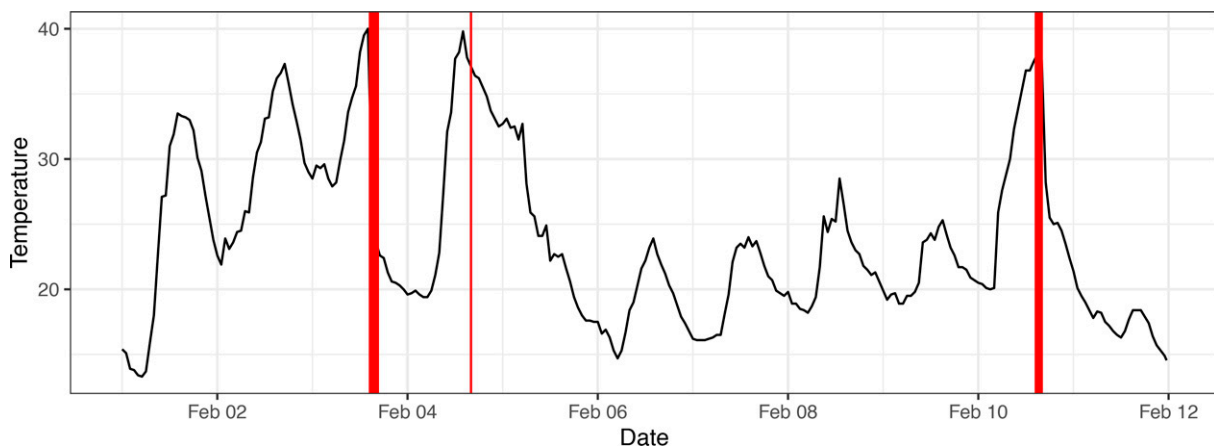
The STR approach allows us to deal easily with level shifts, spikes, and other intervention patterns as defined by Box and Tiao (1975). When these are linear terms, they can simply be added to the X matrix in the usual way. For nonlinear intervention effects, such as “shift and decay” patterns, nonlinear least squares need to be used.

5.5. Forecasting

The model can be used for forecasting by simply treating future observations as missing and then estimating them. The smooth trend T_t is continued linearly, much like a natural spline.

If the remainder term R_t is autocorrelated, the forecasts can be adjusted to allow for the short-term dynamics represented by the autocorrelations (e.g., by fitting an ARMA model to R_t). Prediction intervals can be obtained in the standard way for linear regression models.

When the STR model is being used for forecasting, better results are probably obtained by selecting the smoothing parameters using K -fold cross-validation

Figure 5. (Color online) The Temperature in Melbourne on February 1, 2000, and the Following 10 Days

Notes. Ten vertical lines (two groups of five lines each) mark the times of the outliers. The first group spans 14:30–17:00 on February 3, and the second group spans 15:30–18:00 on the following day.

in which the value of g should be comparable to the forecasting horizon h in order to avoid the trend being too flexible.

5.6. Feature Engineering

The STR model provides a way to identify new features in time series data, such as weekly or daily patterns. These can then be used within other algorithms. For example, they could be used to classify time series (Lubba et al. 2019), to find anomalous time series (as distinct from anomalous observations) (Hyndman et al. 2015), to select a forecasting model (Talagala et al. 2018), or to create a weighted forecast combination (Montero-Manso et al. 2020).

5.7. Other Applications

We look at two different business applications in monthly supermarket revenue and half-hourly electricity demand. It is easy to imagine many others, and we conclude with just three to illustrate the range of possibilities that could be used with this method:

1. A retail outlet wishes to measure the effect of a promotion by studying the increased number of customers in its stores. The promotion took place during a week that includes a public holiday. It can use STR with seasonal, trend, and holiday effects along with a categorical variable indicating the periods before, during, and after the promotion. This gives a direct measure of the number of additional customers during the promotion and the number of additional customers in the weeks following the promotion.

2. A web services company wants to identify potential hostile attacks by measuring unusual traffic on its servers. The usual traffic varies by time of day, day of week, time of year, whether there is a public holiday or not, and what the weather is like in the area. It can use STR with all of these seasonal variables and some weather variables and look for unusually large values in the remainder series. These indicate traffic that is unusual after taking account of the regular causes of variation.

3. An energy company needs to forecast gas usage in a region for the next month, and it has 10 years of hourly data. However, COVID-19 has caused changes in the usage profile as many people are working from home. STR can be used to estimate the trend, regular seasonal patterns with public holiday effects, and new seasonal patterns since people started working from home (including possibly new public holiday effects). The model can include temperature variables to allow for heating effects, and these can interact with the COVID-19 variables.

5.8. Concluding Comments

This article introduces a new flexible approach to seasonal-trend decomposition using a linear regression

model that can be used for a wide range of time series. The STR method allows for multiple seasonal periods and complex seasonality and missing values in input data, provides confidence intervals, finds smoothing parameters, and allows regressors to be taken into account with coefficients that may be time-varying and seasonal.

The main novelties of the STR approach are as follows. First, the seasonal components are represented as two-dimensional structures of optionally complex topology. This is a completely new way of viewing seasonal components in the field of seasonal-trend decomposition. Second, the influence of covariates is allowed to be time-varying and can be seasonal, cyclic, etc. In all cases, the seasonal, cyclic, or other components can have complex constraints that are difficult (if not impossible) to represent using other methods. Third, the approach proposes a regression perspective on time series decomposition rather than the filtering or stochastic process approach that underpins most other methods. Fourth, beyond the methodological contributions mentioned, the main theoretical contribution of the new approach is that it allows us to build a bridge between some stochastic processes and linear regression problems, which can give new mathematical interpretations and new computational methods for solving stochastic process problems. Finally, the new approach provides a unified framework for handling a very wide variety of seasonal-trend decomposition problems, which is easy to adapt in order to take into account further time series features.

Acknowledgments

An R package, **str** (Dokumentov and Hyndman 2018), implementing the authors' model, is available on CRAN. The package also contains a vignette, which describes various practical aspects of the implementation in detail. This paper was produced using Rmarkdown (Allaire et al. 2020). All data included are publicly available with no data ethics issues. The source files to reproduce the paper, including all the examples, are available at <https://doi.org/10.24433/CO.2481805.v1> and https://github.com/robjhyndman/STR_paper.

References

- Allaire J, Xie Y, McPherson J, Luraschi J, Ushey K, Atkins A, Wickham H, Cheng J, Chang W, Iannone R (2020) rmarkdown: Dynamic documents for R. R package version 2.5.3. <https://github.com/rstudio/rmarkdown>.
- Anderson VO, Nachmals U (1914) The elimination of spurious correlation due to position in time or space. *Biometrika* 10(2/3):269–279.
- Bell WR, Martin DEK (2004) Modeling time-varying trading-day effects in monthly time series. *Proc. Joint Statist. Meetings*, 8–12.
- Box GEP, Cox DR (1964) An analysis of transformations. *J. Roy. Statist. Soc. Ser. B. Statist. Methodology* 26(2):211–252.
- Box GEP, Tiao GC (1975) Intervention analysis with applications to economic and environmental problems. *J. Amer. Statist. Assoc.* 70(349):70–79.

- Cleveland RB, Cleveland WS, McRae JE, Terpenning I (1990) STL: A seasonal-trend decomposition procedure based on Loess. *J. Official Statist.* 6(1):3–73.
- Commandeur JFF, Koopman SJ, Ooms M (2011) Statistical software for state space methods. *J. Statist. Software* 41(1):1–18.
- Copeland MT (1915) Statistical indices of business conditions. *Quart. J. Econom.* 29(3):522–562.
- Dagum EB (1988) *The X11ARIMA/88 Seasonal Adjustment Method: Foundations and User's Manual* (Statistics Canada, Time Series Research and Analysis Division).
- Dagum EB, Bianconcini S (2016) *Seasonal Adjustment Methods and Real Time Trend-Cycle Estimation* (Springer, Switzerland).
- De Livera AM, Hyndman RJ, Snyder RD (2011) Forecasting time series with complex seasonal patterns using exponential smoothing. *J. Amer. Statist. Assoc.* 106(496):1513–1527.
- Dokumentov A, Hyndman RJ (2018) stR: STR decomposition. R package version 0.4. <https://CRAN.R-project.org/package=stR>.
- Findley DF (2005) Some recent developments and directions in seasonal adjustment. *J. Official Statist.* 21(2):343–365.
- Findley DF, Monsell BC, Hou CT (2012) Stock series holiday regressors generated by flow series holiday regressors. *Taiwan Econom. Forecast Policy* 43(1):75–122.
- Findley DF, Monsell BC, Bell WR, Otto MC, Chen BC (1998) New capabilities and methods of the X-12-ARIMA seasonal-adjustment program. *J. Bus. Econom. Statist.* 16(2):127–152.
- Gómez V, Maravall A (2001) Seasonal adjustment and signal extraction in economic time series. Peña D, Tiao GC, Tsay RS, eds. *A Course in Time Series Analysis* (John Wiley & Sons, New York), 202–246.
- Harvey AC (1990) *Forecasting, Structural Time Series Models and the Kalman Filter* (Cambridge University Press, Cambridge, UK).
- Hooker RH (1901) The suspension of the Berlin produce exchange and its effect upon corn prices. *J. Roy. Statist. Soc.* 64(4):574–613.
- Hyndman RJ, Wang E, Laptev N (2015) Large-scale unusual time series detection. *Proc. IEEE Internat. Conf. Data Mining*, 1616–1619.
- Hyndman RJ, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, et al. (2021) forecast: Forecasting functions for time series and linear models. Version 8.14. <https://CRAN.R-project.org/package=forecast>.
- Koenker R (2005) *Quantile Regression* (Cambridge University Press, Cambridge, UK).
- Koenker R (2020) quantreg: Quantile regression. R package version 5.75. <https://CRAN.R-project.org/package=quantreg>.
- Ladiray D, Quenneville B (2001) *Seasonal Adjustment with the X-11 Method*. Lecture Notes in Statistics, vol. 158 (Springer, New York).
- Lubba CH, Sethi SS, Knaute P, Schultz SR, Fulcher BD, Jones NS (2019) catch22: Canonical time-series characteristics. *Data Mining Knowledge Discovery* 33(6):1821–1852.
- Macaulay FR (1931) *The Smoothing of Time Series* (NBER Books, New York).
- Montero-Manso P, Athanasopoulos G, Hyndman RJ, Talagala TS (2020) FFORMA: Feature-based forecast model averaging. *Internat. J. Forecasting* 36(1):86–92.
- Poynting JH (1884) A comparison of the fluctuations in the price of wheat and in the cotton and silk imports into Great Britain. *J. Statist. Soc. London* 47(1):34–74.
- R Core Team (2020) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).
- Ruppert D, Wand MP, Carroll RJ (2003) *Semiparametric Regression* (Cambridge University Press, New York).
- Seber GAF, Lee AJ (2003) *Linear Regression Analysis*, 2nd ed. (John Wiley & Sons, New York).
- Shiskin J (1957) Electronic computers and business indicators. Occasional Paper 57, National Bureau of Economic Research, Washington, DC.
- Shiskin J, Young AH, Musgrave JC (1967) The X-11 variant of the Census II method seasonal adjustment program. Technical Report 15, Bureau of the Census, U.S. Department of Commerce, Washington, DC.
- 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. *J. Inst. Actuaries* 38(4):334–343.
- Talagala TS, Hyndman RJ, Athanasopoulos G (2018) Meta-learning how to forecast time series. Working Paper 6/18, Department of Econometrics & Business Statistics, Monash University, Clayton, Victoria, Australia.
- Wood SN (2006) *Generalized Additive Models: An Introduction with R* (CRC Press, Boca Raton, FL).