DEUTSCHE
BUNDESBANK
EUROSYSTEM

Discussion Paper

Deutsche Bundesbank
No 24/2023

Towards seasonal adjustment of infra-monthly time series with JDemetra+

Karsten Webel
(Deutsche Bundesbank)

Anna Smyk
(Insee)

Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main,
Postfach 10 06 02, 60006 Frankfurt am Main

Tel +49 69 9566-0

Please address all orders in writing to: Deutsche Bundesbank,
Press and Public Relations Division, at the above address or via fax +49 69 9566-3077

Internet http://www.bundesbank.de

ISBN 978–3–95729–955–0
ISSN 2941–7503

Non-technical summary

Research Question

Official statistics has paid more and more attention to infra-monthly time series in recent years, mainly due to its ongoing digital transformation and especially as a reaction to the soaring demand for timely economic data in the wake of the COVID-19 pandemic outbreak in 2020. Infra-monthly time series, however, often display complex seasonal dynamics and other peculiarities not commonly seen in data sampled at a monthly or quarterly rate. As a consequence, traditional methods for modeling and seasonally adjusting economic time series in official statistics cannot be applied without appropriate modifications.

Contribution

We give a thorough methodological description of the modifications to the three traditional methods implemented in the official time series and seasonal adjustment software used to produce harmonized statistics within the European Statistical System and the European System of Central Banks. We also discuss key statistical properties of those modifications from a theoretical perspective and consider three examples to highlight their main capabilities.

Results

Our illustrations suggest that the implemented modifications to the traditional modeling and seasonal adjustment approaches provide solid results for various types of infra-monthly economic time series and hence sound alternatives to those stand-alone solutions that are currently not integrated into official statistics' workhorse software. However, more research is needed to achieve permanent acceptance of those modifications among practitioners, and we finally share some thoughts on potential areas for future developments.

Nichttechnische Zusammenfassung

Fragestellung

In den letzten Jahren hat die amtliche Statistik untermonatlichen Zeitreihen zunehmend Aufmerksamkeit geschenkt, hauptsächlich im Rahmen ihrer anhaltenden Digitalisierung und insbesondere als Reaktion auf die gestiegene Nachfrage nach zeitnah verfügbaren ökonomischen Daten nach Ausbruch der COVID-19-Pandemie im Jahr 2020. Untermonatliche Zeitreihen weisen jedoch häufig komplexe saisonale Muster und andere Charakteristika auf, die in dieser Form in monatlich und vierteljährlich erhobenen Daten für gewöhnlich nicht zu beobachten sind. Traditionelle Modellierungs- und Saisonbereinigungsansätze der amtlichen Statistik können somit ohne weitere Anpassungen nicht auf solche ökonomischen Zeitreihen angewendet werden.

Beitrag

Wir beschreiben ausführlich die methodischen Anpassungen an den drei traditionellen Verfahren, die in der offiziellen Software zur harmonisierten Zeitreihenmodellierung und Saisonbereinigung innerhalb des Europäischen Statistischen Systems und des Europäischen Systems der Zentralbanken implementiert sind. Wir diskutieren ebenfalls die wichtigsten statistischen Eigenschaften dieser Anpassungen aus theoretischer Sicht und betrachten drei Beispiele zur Illustration ihrer wesentlichen Fähigkeiten.

Ergebnisse

Unsere Illustrationen lassen den Schluss zu, dass die implementierten Anpassungen an den traditionellen Modellierungs- und Saisonbereinigungsverfahren zufriedenstellende Ergebnisse für diverse Arten untermonatlicher ökonomischer Zeitreihen liefern und damit attraktive Alternativen zu denjenigen Einzellösungen darstellen, die aktuell noch nicht in die offizielle Software der amtlichen Statistik integriert wurden. Nichtsdestotrotz ist weitere Entwicklungsarbeit erforderlich, um für diese Anpassungen eine breite Akzeptanz in der amtlichen Statistik zu schaffen, und wir schließen mit einigen Gedanken zu möglichen zukünftigen Forschungsfeldern.

Bundesbank Discussion Paper No 24/2023

Towards seasonal adjustment of infra-monthly time series with JDemetra+*

Karsten Webel
Deutsche Bundesbank

Anna Smyk
Insee

Abstract

Infra-monthly economic time series have become increasingly popular in official statistics in recent years. This evolution has been largely fostered by official statistics' digital transformation during the last decade. The COVID-19 pandemic outbreak in 2020 has added fuel to the fire as many data users immediately asked for timely weekly and even daily data on economic developments. Such infra-monthly data often display seasonal behavior that calls for adjustment. For that reason, JDemetra+, the official software for harmonized seasonal adjustment of monthly and quarterly data in the European Statistical System and the European System of Central Banks, has been augmented recently with a regARIMA-esque pretreatment model and extended versions of the ARIMA model-based, STL and X-11 seasonal adjustment approaches that are tailored to the specifics of infra-monthly data and accessible through an ecosystem of R packages. This ecosystem also provides easy access to structural time series modeling. We give a comprehensive overview of the packages' current developmental stage and illustrate selected capabilities, including code snippets, using daily births in France, hourly electricity consumption in Germany, and weekly initial claims for unemployment insurance in the United States.

Keywords: extended Airline model, high-frequency data, official statistics, signal extraction, unobserved-components decomposition

JEL classification: C01, C02, C14, C18, C22, C40, C50.

*Contact address: Karsten Webel, Deutsche Bundesbank, Central Office, Statistics Department and Research Centre, Wilhelm-Epstein-Strasse 14, 60431 Frankfurt am Main, Germany. Phone: +49 69 9566 32702. E-mail: karsten.webel@bundesbank.de. The authors thank Christiane Hofer, James Livsey and Jean Palate for valuable comments.

1 Motivation

JDemetra+ (JD+) is the software recommended for the harmonized production of official seasonally adjusted data in the European Statistical System and the European System of Central Banks. It implements the popular X-11 and ARIMA model-based seasonal adjustment approaches for monthly and quarterly data, as well as respective time series regression models for data pretreatment. Version 3.0 of JD+ is currently under development mainly due to the addition of seasonal adjustment methods for infra-monthly economic data. Such time series have become increasingly popular in official statistics over the past decade. A key driver has been the emergence of new digital data sources that provide easy and sometimes almost real-time access to such data. In 2020, the demand sharply accelerated immediately after the outbreak of the COVID-19 pandemic when many institutional data users asked for more timely indicators in order to track economic developments on a daily or at least weekly basis. Although being strongly seasonal in many cases, infra-monthly economic data do not lend themselves to being modeled with classic Box-Jenkins approaches and seasonally adjusted with traditional methods due to a variety of peculiarities not seen in monthly and quarterly data. Prime examples include the superimposition of multiple seasonal patterns with potentially fractional periodicities that result in a dense set of seasonal frequencies, direct measurability of granular calendar variation and a high susceptibility to irregular spacing, outliers and missing values (see the discussions in Ollech, 2023; Proietti and Pedregal, 2022; Webel, 2022). Thus, possible strategies for modeling and seasonally adjusting infra-monthly time series are (1) proper data regularization so that traditional approaches become applicable to the regularized data, (2) proper modifications to traditional approaches so that they become applicable to infra-monthly data, and (3) the design and application of new approaches not considered in official statistics so far.

The current solutions offered by JD+ 3.0 follow the second branch of strategies as they are founded on modified established methods, such as TRAMO for data pretreatment and extended versions of the ARIMA model-based (AMB), X-11 and STL methods for seasonal adjustment. The corresponding routines have been implemented in Java and can be accessed easily through an ecosystem of R packages (see the Appendix for further details). Although the Java code is still experimental, some methodological overviews and applications are already available (Ladiray, Palate, Mazzi, and Proietti, 2018; Webel, 2022). We now elaborate on the packages' main modifications and give additional theoretical details, derive and discuss selected properties of the key models and filtering methods and provide code snippets plus examples for illustration.

We start with a discussion about data pretreatment through a time series regression model in which the residuals follow an Airline-type ARIMA model that allows for fractional powers of the backshift operator (Section 2). The latter model is then used to introduce proper modifications to the traditional AMB approach (Section 3). Key extensions to the traditional X-11 and STL approaches are presented in Section 4 and Section 5, respectively, followed by a brief description of structural time series models (Section 6). We illustrate selected capabilities of the discussed methods using daily births in France, hourly realized electricity consumption in Germany, and weekly initial claims for unemployment insurance in the United States (Section 7) before we conclude with some final remarks and suggestions for future developments (Section 8).

2 Pretreatment

2.1 Model and properties

Let $\{y_t\}$ denote an observed infra-monthly time series and assume that it can be decomposed additively, maybe after taking logs, into a trend-cyclical component $\{t_t\}$, a seasonal component $\{s_t\}$, a calendar component $\{c_t\}$ and an irregular component $\{i_t\}$, resulting in the standard unobserved components (UC) decomposition

$$y_t = t_t + s_t + c_t + i_t. \tag{1}$$

As for monthly and quarterly data, calendar variation and outliers should be removed from $\{y_t\}$ prior to the extraction of seasonal movements. This linearization is achieved through the time series regression

$$y_t = \mathbf{x}_t^\top \boldsymbol{\beta} + \eta_t, \tag{2}$$

where $\mathbf{x}_t$ is a vector of regression variables associated with calendar events and outliers, $\boldsymbol{\beta}$ is a vector of unknown calendar and outlier effects and $\{\eta_t\}$ is a residual process.

Infra-monthly time series often contain multiple seasonal patterns, so that the seasonal variation in (1) can be represented as

$$s_t = \sum_{\tau \in \mathbb{S}} s_t^{(\tau)}, \tag{3}$$

where $\mathbb{S} = \{\tau_1, \ldots, \tau_{|\mathbb{S}|}\} \subset \mathbb{R}$ is a set of seasonal periodicities and $\{s_t^{(\tau)}\}$ is the seasonal pattern with potentially fractional periodicity τ. The residuals in (2) naturally carry the same superimposed seasonal patterns and, therefore, cannot be modeled as a standard ARIMA process. An alternative distributional assumption is provided by the extended Airline model, which is essentially an Airline-type model that allows for multiple first-order seasonal integration including fractional powers of the backshift operator. It is given by

$$\delta_1(B) \prod_{\tau \in \mathbb{S}} \delta_\tau(B)\, \eta_t = \theta_1(B) \prod_{\tau \in \mathbb{S}} \theta_\tau(B)\, \varepsilon_t, \tag{4}$$

where B is the backshift operator, $\delta_1(B) = 1 - B$ is the non-seasonal differencing operator, $\delta_\tau(B) = 1 - B^\tau$ is the seasonal differencing operator at lag τ, $\theta_1(B) = 1 - \theta_1 B$ is a non-seasonal moving average (MA) operator, $\theta_\tau(B) = 1 - \theta_\tau B^\tau$ is a seasonal MA operator at lag τ and $\{\varepsilon_t\}$ is white noise with finite variance $\sigma_\varepsilon^2 > 0$. Fractional powers of B are defined through the first-order Taylor approximation at unity, using the fact that $B^k = B^{\lfloor k \rfloor} B^{\alpha_k}$ holds for any $k \in \mathbb{R}$, where $\lfloor k \rfloor \in \mathbb{Z}$ is the largest integer not exceeding k and $\alpha_k = k - \lfloor k \rfloor \in [0, 1)$ is the fractional remainder of k. Now, $B^{\alpha_k} \approx (1 - \alpha_k) + \alpha_k B$ and, hence, B^k is essentially approximated with a weighted average of $B^{\lfloor k \rfloor}$ and $B^{\lfloor k \rfloor + 1}$:

$$B^k \approx (1 - \alpha_k) B^{\lfloor k \rfloor} + \alpha_k B^{\lfloor k \rfloor + 1}. \tag{5}$$

The approximation (5) can be plugged into the seasonal differencing and MA operators in (4). For weekly data, for instance, we have $|\mathbb{S}| = 1$ with $\tau_1 = 52.18$ and seasonal

differencing becomes

$$\delta_{52.18}(B)\, y_t = \left(1 - B^{52.18}\right) y_t = y_t - (0.82\, y_{t-52} + 0.18\, y_{t-53}). \tag{6}$$

In general, the squared gain of any seasonal differencing operator $\delta_\tau(B)$ is given by

$$\left|\delta_\tau\left(e^{-i\omega}\right)\right|^2 = 1 - 2c_\tau(\omega) + \left[c_\tau^2(\omega) + s_\tau^2(\omega)\right], \tag{7}$$

where

$$c_\tau(\omega) = (1-\alpha_\tau)\cos\left(\lfloor\tau\rfloor\,\omega\right) + \alpha_\tau \cos\left[(\lfloor\tau\rfloor + 1)\,\omega\right]$$

and $s_\tau(\omega)$ is defined analogously based on the sine instead of cosine function. Note that $c_\tau(\omega) = \cos(\tau\omega)$ and $s_\tau(\omega) = \sin(\tau\omega)$ if $\tau \in \mathbb{N}$, so that (7) simplifies to a well-known standard case. Figure 1 shows the squared gain (7) for the seasonal differencing operator applied in (6), revealing that this operator does not completely annihilate week-of-the-year (WOY) dynamics. Although the squared gain dips at the WOY harmonics, it does not reach zero (in fact, those dips remain even larger than 0.1 at higher WOY harmonics). This flaw is common to all operators $\delta_\tau(B)$ with a fractional τ and essentially a consequence of the (first-order) truncation in the Taylor series expansion of B^{α_τ}, which results in $\delta_\tau(B)$ having only a single unit root at unity. To see this, note that this operator factorizes as $\delta_\tau(B) = \delta_1(B)S_\tau(B)$, where

$$S_\tau(B) = 1 + B + \cdots + B^{\lfloor\tau\rfloor - 1} + \alpha_\tau B^{\lfloor\tau\rfloor} \tag{8}$$

is the fractional aggregation operator associated with τ. When $\tau \in \mathbb{N}$ and hence $\alpha_\tau = 0$, then $S_\tau(B)$ carries $\lfloor(\tau-1)/2\rfloor$ pairs of complex conjugate unit roots that are equally spaced around the unit circle at the seasonal frequencies and accompanied by an additional single real unit root at -1 if τ is even. However, when $\tau \notin \mathbb{N}$ and hence $\alpha_\tau > 0$, then $S_\tau(B)$ must contain one additional root since the polynomial order of (8) has been increased by one (compared to the case of $\tau \in \mathbb{N}$). Although the roots of $S_\tau(B)$ still reflect seasonal dynamics, they are no longer exactly associated with the respective seasonal frequencies and also lie outside the unit circle.[1]

The weighted averaging principle (5) also facilitates the derivation of the pseudo-spectral density of the regression residuals $\{\eta_t\}$ in (2). It follows directly from (4) and (5)

[1]In fact, the real root at -1 that is present in (8) when $\tau \in \mathbb{N}$ is even turns into a complex root with an associated frequency close to π when $\tau \notin \mathbb{N}$ and $\lfloor\tau\rfloor$ is even. Similarly, an additional negative real root is introduced when $\tau \notin \mathbb{N}$ and $\lfloor\tau\rfloor$ is odd (compared to the case in which $\tau \in \mathbb{N}$ is odd). Numerical solutions for fractional periodicities based upon the Jenkins-Traub algorithm (Jenkins and Traub, 1972) suggest that the roots of $S_\tau(B)$ are associated with frequencies slightly larger than the seasonal ones and have moduli slightly larger than 1. For example, when $\tau = 52.18$, the differences in radians between the frequencies of the complex roots in (8) and the respective WOY frequencies are found to increase from 5.2×10^{-7} for the chief WOY frequency to 9.8×10^{-3} for the 25-th WOY harmonic, while the corresponding moduli gradually increase from 1.000021 to 1.008634. Conceptually, when $\tau \notin \mathbb{N}$, principle (5) sacrifices the seasonal unit root properties of the fractional aggregation operator $S_\tau(B)$ that appears in the top-down factorization of $\delta_\tau(B)$. An alternative seasonal differencing operator that retains those seasonal unit root properties can be constructed via an a priori selection of the relevant seasonal frequencies and an atomic bottom-up multiplication of the respective minimal degree polynomials (see the discussion in McElroy and Livsey, 2022, Appendix B).

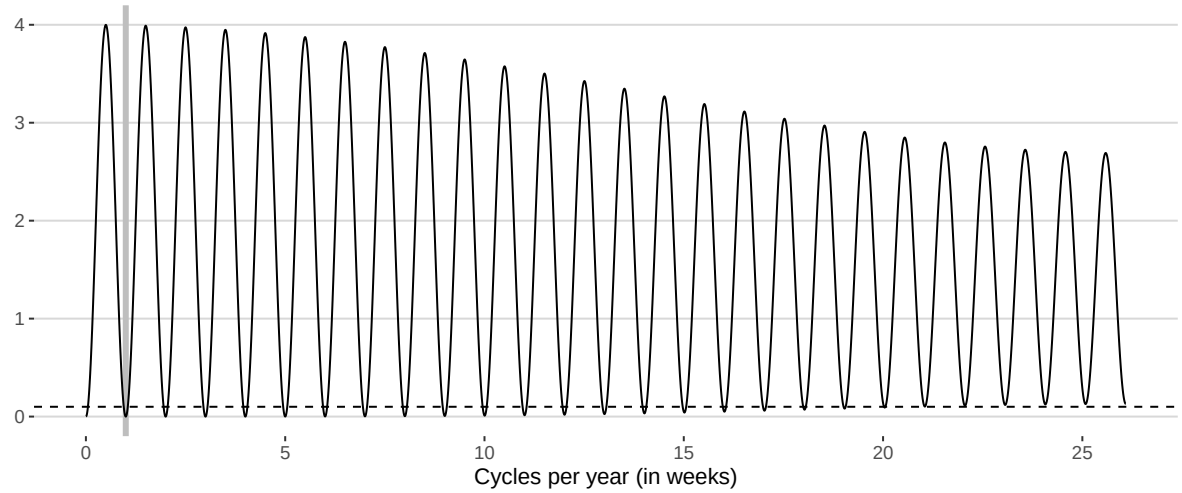


Figure 1: Squared gain (7) for $\tau = 52.18$. The gray vertical marks the chief week-of-the-year frequency. The dashed horizontal marks 0.1.

that

$$f_\eta(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \times \frac{\left|\theta_1\left(e^{-i\omega}\right)\right|^2 \prod_{\tau\in\mathbb{S}} \left|\theta_\tau\left(e^{-i\omega}\right)\right|^2}{\left|\delta_1\left(e^{-i\omega}\right)\right|^2 \prod_{\tau\in\mathbb{S}} \left|\delta_\tau\left(e^{-i\omega}\right)\right|^2}, \quad \omega \in (0,\pi], \tag{9}$$

where the squared gains of the seasonal differencing filters are given by (7) and those of the other involved filters read

$$\begin{aligned}
\left|\theta_1\left(e^{-i\omega}\right)\right|^2 &= 1 - 2\theta_1\cos(\omega) + \theta_1^2, \\
\left|\theta_\tau\left(e^{-i\omega}\right)\right|^2 &= 1 - 2\theta_\tau c_\tau(\omega) + \theta_\tau^2\left[c_\tau^2(\omega) + s_\tau^2(\omega)\right], \\
\left|\delta_1\left(e^{-i\omega}\right)\right|^2 &= 2\left[1 - \cos(\omega)\right].
\end{aligned}$$

Figure 2 shows the pseudo-spectral density (9) for the weekly extended Airline model and four combinations of its MA parameters. It is seen that, as for the classic Airline model, the non-seasonal MA parameter governs the shape of (9) as $\omega \to 0$ (higher/lower values of θ_1 correspond to steeper/flatter increases), whereas the seasonal MA parameter governs the shape of (9) in the vicinity of the WOY frequencies $\omega_k = 2\pi k/52.18$ with $k \in \{1,\ldots,26\}$ (higher/lower values of $\theta_{52.18}$ correspond to narrower/wider peaks). However, in contrast to the classic Airline model, the magnitude of the seasonal peaks decreases dramatically at the higher WOY harmonics due to the aforementioned fact that $\delta_{52.18}(B)$ does not entirely annihilate WOY dynamics (recall that its squared gain appears in the denominator of (9)).

2.2 Model estimation

The unknown parameters of pretreatment model (2) and (4) are estimated through the iterative procedure schematized in Algorithm 1. In essence, in each iteration, the regression parameters in (2) are estimated via ordinary least squares (OLS), whereas the ARMA parameters in (4) are estimated via maximum likelihood (ML) techniques. Running through

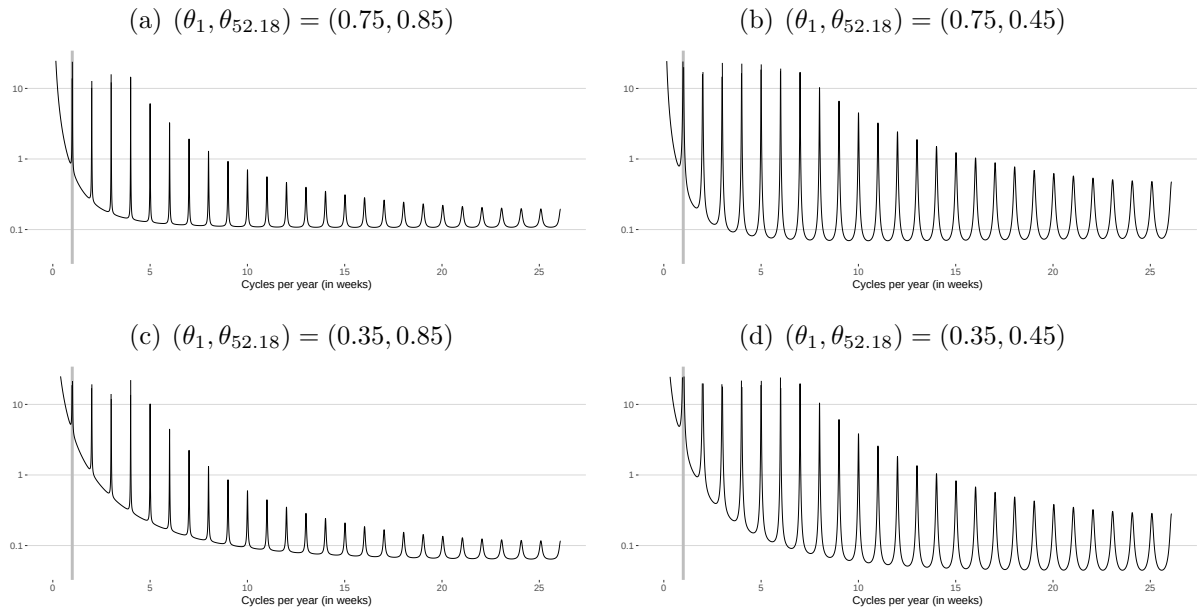


Figure 2: Pseudo-spectral density (9) with $\mathbb{S} = \{52.18\}$, $(\theta_1, \theta_{52.18}) \in \{0.35, 0.75\} \times \{0.45, 0.85\}$ and $\sigma_\varepsilon^2 = 1$ (in decibel). The gray vertical marks the chief week-of-the-year frequency.

the iterations, optimal regression and ARMA parameter estimates are obtained through the Levenberg-Marquardt algorithm (Levenberg, 1944; Marquardt, 1963). The following list contains additional details, with item numbers referring to lines in Algorithm 1:

3: To avoid time-consuming matrix inversions, the OLS solution utilizes the QR factorization of the design matrix, which is computed via Householder reflections.

4–7: The automatic forward-addition-backward-deletion search for outliers is optional.

Algorithm 1 Iterative estimation of pretreatment model (2) and (4)

Let $\boldsymbol{\psi} = (\theta_1, \theta_{\tau_1}, \ldots, \theta_{\tau_{|\mathbb{S}|}}, \sigma_\varepsilon^2)^\top$ be the vector of the ARMA parameters from (4), including the white-noise variance.

1: Set $\boldsymbol{\psi}^{(0)} = (0.2, \ldots, 0.2, 1)^\top$.
2: **repeat**
3: Given $\hat{\boldsymbol{\psi}}^{(i-1)}$ in (4), estimate $\hat{\boldsymbol{\beta}}^{(i)}$ in (2) via OLS after applying the Kalman filter and smoother (KFS) to $\{y_t\}$ and $\{\mathbf{x}_t\}$.
4: **repeat**
5: Run a sequence of OLS regressions of the KFS residuals on each potential outlier type and date and add the most significant outlier to (2).
6: Remove insignificant outliers sequentially from (2) without model re-estimation.
7: **until** a stable tentative version of model (2) is found, or a maximum of 100 search rounds is reached.
8: Given $\hat{\boldsymbol{\beta}}^{(i)}$ in (2), establish the state space representation of model (4), run the KFS and estimate $\hat{\boldsymbol{\psi}}^{(i)}$ via ML, utilizing the KFS output.
9: **until** convergence, or a maximum of 200 iterations is reached.

Valid types are additive outliers, level shifts and lag-1 switch interventions, which are a special case of reallocation outliers (Wu, Hosking, and Ravishanker, 1993). Given a prespecified critical value, outlier significance is determined through a diagnostic statistic akin to the one derived in de Jong and Penzer (1998).

8: Model (4) is put into Akaike's Markovian (minimal state space) representation (Gómez and Maravall, 1994). The KFS implementation utilizes fast Chandrasekhar recursions (Morf, Sidhu, and Kailath, 1974; Morf and Kailath, 1975) and a proper calculation of the initial state covariance matrix (Jones, 1980). ML estimates are found through maximizing the concentrated profile log-likelihood function (Francke, Koopman, and de Vos, 2010).

9: Convergence is measured by the incremental changes of (1) the estimated regression parameters in (2), (2) the gradient of the likelihood function used to obtain the ML estimates of the ARMA parameters in (4) and (3) the residual sum of squares of the entire pretreatment model.

3 Extended ARIMA model-based approach

The extension of classic AMB seasonal adjustment for monthly and quarterly time series also rests on the extended Airline model (4). In the first step, this model is estimated (from the linearized observations) in the same way as described in Section 2. In a second step, the canonical decomposition of the estimated extended Airline model into the requested UC models is found according to the classic approach (Burman (1980); see also Gómez and Maravall (2001) and Maravall (1995) for extensive overviews). To this end, the non-seasonal and seasonal differencing operators in (4) are factorized by means of polynomial division, which after substituting (8) into (4) results in

$$\delta_1^{1+|\mathbb{S}|}(B)\prod_{\tau\in\mathbb{S}} S_\tau(B),$$

so that the unit roots at unity are assigned to the trend-cycle and the non-stationary ($\alpha_\tau = 0$) or stationary ($\alpha_\tau > 0$) roots of the fractional aggregation operators are assigned to the corresponding seasonal patterns. Given such a factorization, the decomposition of the model's MA part is achieved through a partial fraction expansion. Combining the UC models' differencing and MA polynomials according to classic Wiener-Kolmogorov (WK) theory then gives the minimum mean squared error (MMSE) estimators for the requested UC, including the seasonally adjusted time series. In some cases, this decomposition can be inadmissible in the sense that the spectrum of the estimated irregular component is negative for at least one frequency. Variance inflation is used to solve this issue, that is artificial white noise is added to the model until the spectrum of the irregular component becomes non-negative. Afterwards, the WK filters and MMSE estimators are recomputed from the admissible decomposition of the noise-inflated extended Airline model. Note, however, that the MMSE estimator for any seasonal pattern with a fractional periodicity will technically be stationary since it inherits the stationarity-inducing transformation from the corresponding UC model.

Although the WK filters provide MSE-optimal UC estimators in theory, preliminary empirical studies conducted at the National Bank of Belgium suggest that the classic decomposition can become numerically unstable quite quickly for infra-monthly data and hence those optimal estimates may be unreliable. Therefore, refined UC estimates are calculated in a third step via an application of the KFS to the state space representation of the decomposed extended Airline model. The diffuse square-root KFS is applied if standard deviations of the UC estimates are opted for; otherwise, the fast disturbance smoother (Koopman, 1993) is applied (also with a diffuse initialization). Either way, the AMB decomposition can produce back- and forecasts for the UC estimates along the way. Those can then be used to construct model-based back- and forecasts for the observed time series through simple aggregation.

4 Extended X-11 approach

The classic X-11 seasonal adjustment method derived in Shiskin, Young, and Musgrave (1967) and reviewed in great detail in Ladiray and Quenneville (2001) suits monthly and quarterly time series. It is based upon an iterative application of the following 4-step principle. First, the trend-cycle is extracted from the input series through the application of $2 \times \tau$ moving averages or Henderson filters, depending on the position in the entire filtering process. Second, the estimated trend-cycle is removed from the input series. Third, the detrended input series is separated into normalized seasonal and irregular components through the application of $3 \times k$ seasonal moving averages. At some positions, this steps also includes an automatic detection and down-weighting of extreme values in the irregular component based upon the famous σ-limits. Fourth, the estimated seasonal component is removed from the input series. This key principle is applied twice in each of the X-11 iterations B through D. Afterwards, the final trend-cycle and irregular estimates are obtained from splitting the final seasonally adjusted series with the Henderson filter, completing a sequence of preliminary, refined and final UC estimates.

The extended X-11 approach essentially adopts this iterative smoothing process but incorporates some generalizations and modifications tailored to the peculiarities of infra-monthly data. This section describes the key changes to the classic X-11 method. The current implementation facilitates a sequential extraction of the seasonal patterns in (3) without automatic selection of trend-cycle and seasonal filters based on I/C and I/S ratios and without any forecasts, even naive ones, of the seasonal patterns.

4.1 Trend-cycle filters

4.1.1 Preliminary trend-cycle estimation

Preliminary trend-cycle extraction in the classic X-11 method is carried out with centered symmetric $2 \times \tau$ moving averages for $\tau \in \{4, 12\}$. The extended X-11 approach utilizes a generalized version of those crude filters that allows for any seasonal periodicity τ. Let $\mathbb{N}_{\text{odd}}$ be the set of odd integers and $\lceil k \rceil^{\text{odd}}$ denote the smallest odd integer not smaller than k, i.e. $\lceil k \rceil^{\text{odd}} = \min \{l \geq k \,|\, l \in \mathbb{N}_{\text{odd}}\}$. Then, the generalized $2 \times \tau$ moving average

has length $l_\tau = \lceil \tau \rceil^{\text{odd}}$ and its weights are given by

$$w_i = \begin{cases} 1/\tau, & i \in \{1, l_\tau\} \text{ with } \tau \in \mathbb{N}_{\text{odd}} \\ (\mathbb{I}\{\lfloor \tau \rfloor \text{ even}\} + \alpha_\tau)/(2\tau), & i \in \{1, l_\tau\} \text{ with } \tau \notin \mathbb{N}_{\text{odd}} \\ 1/\tau, & i \in \{2, \ldots, l_\tau - 1\} \end{cases}, \tag{10}$$

where $\mathbb{I}\{\cdot\}$ is the indicator function. It is easily seen that $\sum_{i=1}^{l_\tau} w_i = 1$ holds for any choice of τ. Asymmetric variants are not used during preliminary trend-cycle extraction so that $(l_\tau - 1)/2$ observations are temporarily lost at the beginning and end of the filter output.

4.1.2 Refined and final trend-cycle estimation

The refined and final trend-cycle extraction filters of the classic X-11 method are essentially a set of prespecified weights for symmetric m-term Henderson filters with $m \in \{3, \ldots, 101\} \cap \mathbb{N}_{\text{odd}}$ and their asymmetric Musgrave surrogates. The extended X-11 approach implements the generalized approach to trend-cycle estimation derived in Proietti and Luati (2008), which is akin to the approach of Gray and Thomson (2002), alongside three very different ways of calculating asymmetric variants.

Trend-cycle estimation according to Proietti and Luati (2008) is founded on applying local polynomial regressions to the input series. Let h denote the local bandwidth, or half-length, of the trend-cycle filter and d the polynomial degree which the filter should be able to preserve. Also, let $q \in \{0, \ldots, h\}$ be the number of future observations available after the current time point t. The local polynomial regression model is then given by

$$\mathbf{z}_t = \mathbf{X}_q \boldsymbol{\zeta} + \boldsymbol{\varepsilon}_t, \tag{11}$$

where $\mathbf{z}_t = (z_{t-h}, \ldots, z_{t+q})^\top$ is the local span of the input series,

$$\mathbf{X}_q = \begin{pmatrix} 1 & -h & h^2 & \cdots & (-h)^d \\ 1 & -(h-1) & (h-1)^2 & \cdots & [-(h-1)]^d \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & q & q^2 & \cdots & q^d \end{pmatrix}$$

is the time-constant design matrix, $\boldsymbol{\zeta} = (\zeta_0, \ldots, \zeta_d)^\top$ is a $(d+1)$-dimensional vector of unknown regression coefficients, and $\boldsymbol{\varepsilon}_t = (\varepsilon_{t-h}, \ldots, \varepsilon_{t+q})^\top$ is a local vector of zero-mean Gaussian white noise processes that are both mutually and serially uncorrelated. A weighted least squares (WLS) estimator of $\boldsymbol{\zeta}$ is obtained through minimization of the objective function

$$\sum_{j=-h}^{q} \kappa_j \left(z_{t+j} - \zeta_0 - \zeta_1 j - \cdots - \zeta_d j^d \right)^2, \tag{12}$$

where $\{\kappa_j\}$ is a set of non-negative weights defined through a prespecified kernel.[2] Using

[2]The R package currently implements the biweight, Epanechnikov, Gaussian, Henderson, trapezoidal, triangular, tricube, triweight and uniform kernels.

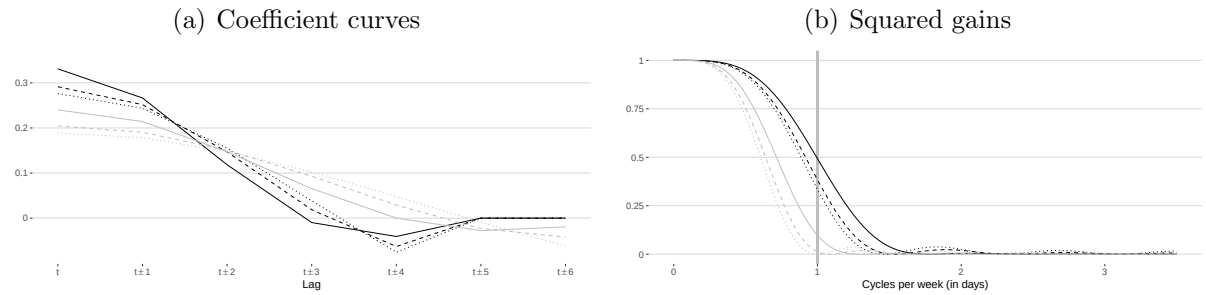


Figure 3: Symmetric kernel-based Henderson (*solid*), Epanechnikov (*dashed*) and unit-variance Gaussian (*dotted*) trend-cycle extraction filters (13) with $q = h = 4$ (*black*) and $q = h = 6$ (*gray*) and $d = 3$. The gray vertical in Panel (b) marks the chief day-of-the-week frequency.

standard least-squares theory, the WLS minimizer of (12) is found as

$$\hat{\boldsymbol{\zeta}} = (\mathbf{X}_q^\top \mathbf{K}_q \mathbf{X}_q)^{-1} \mathbf{X}_q^\top \mathbf{K}_q \mathbf{z}_t$$

with $\mathbf{K}_q = \text{diag}(\kappa_{-h}, \ldots, \kappa_q)$. Since the local trend-cycle estimator is given by $\hat{\zeta}_0 = \mathbf{e}_1^\top \hat{\boldsymbol{\zeta}}$, where $\mathbf{e}_1$ is the first $(d+1)$-dimensional unit vector, we finally have $\hat{\zeta}_0 = \mathbf{w}_q^\top \mathbf{z}_t$ with

$$\mathbf{w}_q = \mathbf{K}_q \mathbf{X}_q (\mathbf{X}_q^\top \mathbf{K}_q \mathbf{X}_q)^{-1} \mathbf{e}_1. \tag{13}$$

Assuming $2h \geq d$, the weights of the symmetric kernel-based trend-cycle extraction filters result from setting $q = h$ in (13). Figure 3 shows the squared gains of the cubic filters derived through the Henderson, Epanechnikov and unit-variance Gaussian kernels for bandwidths $h \in \{4, 6\}$ (which correspond to the lengths of classic 9-term and 13-term trend-cycle extraction filters). Any of these filters qualifies for trend-cycle extraction during the seasonal adjustment of daily data with pronounced day-of-the-week dynamics. For either bandwidth, the squared gains of the three filters are shaped very similarly, although the pass-band becomes slightly narrower when the Epanechnikov and Gaussian kernels are utilized in place of the Henderson kernel.

The weights of the corresponding asymmetric variants can be obtained for any choice of $q \in \{0, \ldots, h-1\}$ in (13). For example, the concurrent kernel-based trend-cycle extraction filters are given by $\mathbf{w}_0$. Those variants are conceptually consistent with the symmetric filters $\mathbf{w}_h$—and therefore called "direct" asymmetric filters—but often yield strongly localized and hence volatile trend-cycle estimates. The "cut-and-normalize" approach (Gasser and Müller, 1979) is a rather simple alternative according to which the asymmetric variants are found through dropping the unneeded weights from the symmetric filter $\mathbf{w}_h$ and dividing the remaining weights by their sum, that is

$$\mathbf{w}_q = \mathbf{w}_h^{(q)} \times \left[\mathbf{1}^\top \mathbf{w}_h^{(q)} \right]^{-1},$$

where $\mathbf{w}_h^{(q)}$ contains only the first $h+q+1$ elements of $\mathbf{w}_h$ and $\mathbf{1}$ is a $(h+q+1)$-dimensional vector of ones. A more sophisticated alternative is the "minimum mean squared revision error" (MMSRE) approach also discussed in Proietti and Luati (2008). Aiming at a bias-variance trade-off, its key idea is to purposely introduce a slight bias into the trend-

cycle estimators by assuming that the polynomial degree is of lower order $\overline{d} < d$ at the boundary and by seeking asymmetric trend-cycle filters that preserve polynomials of even lower degree $\underline{d} < \overline{d}$ subject to the reproduction constraints dictated by the symmetric filter $\mathbf{w}_h$. The Musgrave surrogates, for example, can be replicated in the linear-constant case $(\overline{d}, \underline{d}) = (1, 0)$ if the Henderson kernel is used to calculate $\mathbf{w}_h$ according to (13). The corresponding objective function, that is the analog to the matrix representation of (12), is then composed of the variance and squared bias of the trend-cycle revision error plus a vector of Lagrange multipliers linked to the reproduction constraints. The current implementation of the MMSRE approach even utilizes a generalized minimization criterion derived in Grun-Rehomme, Guggemos, and Ladiray (2018), which essentially translates the aforementioned objective function to the frequency domain where it is decomposable into effects related to the gain and phase shift properties of the involved trend-cycle filters and hence capable of balancing certain fidelity, smoothness and timeliness criteria.

Note that the framework of Proietti and Luati (2008) also enables the identification of the optimal bandwidth h through cross-validation for a given time series and a prespecified polynomial degree d and set of kernel weights $\{\kappa_j\}$. However, this procedure has not been implemented in the R package so far.

4.2 Seasonal filters

The classic X-11 method implements a set of symmetric $3 \times k$ seasonal moving averages, where $k \in \{1, 3, 5, 9, 15\}$, as well as asymmetric variants in order to extract estimates of the seasonal component from the detrended data (alias the seasonal-irregular component). The extended X-11 approach adopts the exact same seasonal filters but allows for more combinations of different filters when extracting the initial and final seasonal factors in iterations B through D. Seasonal estimates are normalized with the generalized symmetric $2 \times \tau$ moving average with weights given in (10) and its asymmetric variants defined implicitly through padding the beginning (end) of a seasonal estimate with $(l_\tau - 1)/2$ copies of its first (last) element. Principle (5) is used additionally in the case of fractional periodicities, that is weighted averages of the two corresponding adjacent integer-valued lagged values of the seasonal-irregular component constitute the input to the seasonal filter. Considering the symmetric 3×3 seasonal filter, for example, the estimate of the seasonal pattern is obtained from the detrended infra-monthly data according to

$$
\begin{aligned}
\hat{s}_t &= \frac{1}{9}\left[\alpha_{2\tau}(\widehat{si})_{t-\lfloor 2\tau \rfloor - 1} + (1-\alpha_{2\tau})(\widehat{si})_{t-\lfloor 2\tau \rfloor}\right] \\
&+ \frac{2}{9}\left[\alpha_{\tau}(\widehat{si})_{t-\lfloor \tau \rfloor - 1} + (1-\alpha_{\tau})(\widehat{si})_{t-\lfloor \tau \rfloor}\right] \\
&+ \frac{3}{9}(\widehat{si})_t \\
&+ \frac{2}{9}\left[(1-\alpha_{\tau})(\widehat{si})_{t+\lfloor \tau \rfloor} + \alpha_{\tau}(\widehat{si})_{t+\lfloor \tau \rfloor + 1}\right] \\
&+ \frac{1}{9}\left[(1-\alpha_{2\tau})(\widehat{si})_{t+\lfloor 2\tau \rfloor} + \alpha_{2\tau}(\widehat{si})_{t+\lfloor 2\tau \rfloor + 1}\right].
\end{aligned} \tag{14}
$$

Figure 4 shows the squared gain of the (effective) seasonal extraction filter (14) for $\tau = 30.44$, that is of the filter that extracts the day-of-the-month (DOM) pattern from

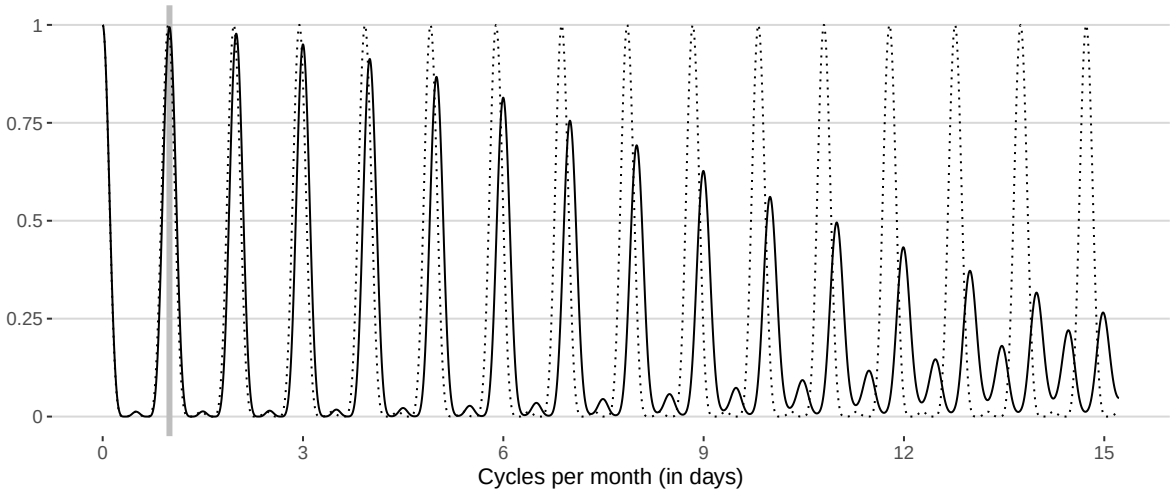


Figure 4: Squared gains of the symmetric 3×3 seasonal extraction filter (14) for $\tau = 30.44$ (*solid*) and $\tau = 31$ (*dotted*). The gray vertical marks the chief day-of-the-month frequency.

detrended daily data (other examples are shown in Figure 14 in the Appendix).[3] Its squared gain is almost one at the chief DOM frequency but from the first harmonic onward the DOM peaks start to decline visibly in almost linear fashion towards 0.25. At the same time, the average level of the squared gain at the infra-DOM frequencies visibly separates from zero between the higher DOM harmonics (with initially small infra-DOM peaks reaching almost 0.25 between the thirteenth and fourteenth harmonic). This divergence effect has already been seen for squared gains of seasonal differencing operators with fractional periodicities (see e.g. Figure 1). Asymmetric seasonal filters are derived accordingly. For example, the concurrent variant of (14) reads

$$\begin{aligned} \hat{s}_t &= \frac{5}{27}\left[\alpha_{2\tau}(\widehat{si})_{t-\lfloor 2\tau\rfloor-1} + (1-\alpha_{2\tau})(\widehat{si})_{t-\lfloor 2\tau\rfloor}\right] \\ &+ \frac{11}{27}\left[\alpha_{\tau}(\widehat{si})_{t-\lfloor \tau\rfloor-1} + (1-\alpha_{\tau})(\widehat{si})_{t-\lfloor \tau\rfloor}\right] \\ &+ \frac{11}{27}(\widehat{si})_t, \end{aligned} \tag{15}$$

the squared gain of which is shown in Figure 5 (a) for $\tau = 30.44$. Compared with the symmetric filter, the latter is similar in shape but has slightly higher and broader peaks at the DOM and "in-between-DOM" frequencies. Panel (b) reveals that the concurrent 3×3 seasonal extraction filter (15) introduces a phase delay of up to 24 days into the low-frequency variation of the filtered series, which may be neglected since the filter is applied to detrended data. Apart from that, no serious phase delays or advances are introduced beyond the chief DOM frequency.

[3] In general, the squared gain of any symmetric $3 \times k$ seasonal extraction filter is calculated as $\left|\mathcal{S}_{3\times k}\left(e^{-i\omega}\right)\right|^2$ with $\mathcal{S}_{3\times k}(B)$ being the filter's polynomial representation given by

$$\mathcal{S}_{3\times k}(B) = \frac{1}{3k}\left(B^{\tau} + 1 + B^{-\tau}\right)\left(B^{(k-1)/2\times\tau} + \cdots + B^{-(k-1)/2\times\tau}\right),$$

where fractional powers of B are defined in (5).

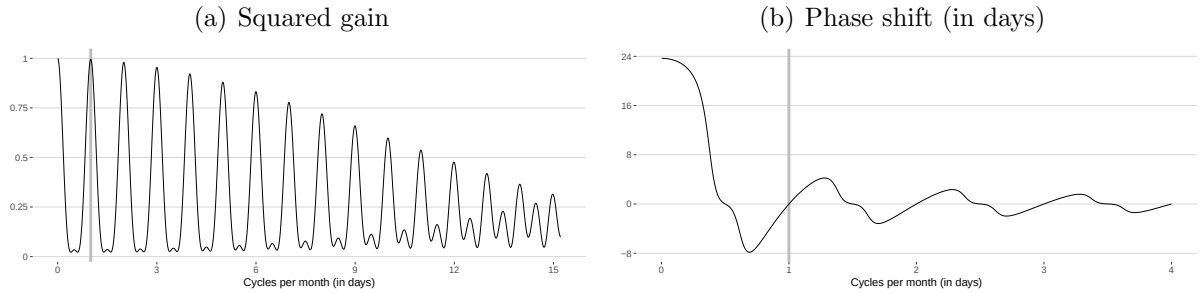


Figure 5: Spectral characterization of the concurrent 3×3 seasonal extraction filter (15) for $\tau = 30.44$. Gray verticals mark the chief day-of-the-month frequency.

Seasonal filters based on principle (5) operate at the exact non-integer seasonal periodicities but inevitably need to compromise some extraction power at the higher seasonal harmonics. Other seasonal adjustment approaches for infra-monthly time series possess a high extraction power at all seasonal harmonics at the expense of operating at slightly inaccurate integer periodicities, which is often achieved through temporary data regularization. For example, the STL-based approach for daily data implemented in the {`dsa`} R package (Ollech, 2021) utilizes spline interpolation in order to temporarily stretch each month to 31 days. Setting $\tau = 31$ instead of $\tau = 30.44$ in (14) mimics the corresponding 5-term seasonal LOESS smoother, and the aforementioned trade-off effect is visualized by the dotted squared gain in Figure 4.

5 STL approach

The classic STL method developed in Cleveland, Cleveland, McRae, and Terpenning (1990) essentially adopts the sequential linear filtering concept of X-11 in its inner loop but incorporates a few modifications. The probably most notable of those is the utilization of trend-cycle and seasonal filters, the weights of which are no longer predefined but derived through locally weighted non-parametric univariate regressions (LOESS). The LOESS objective function is similar to (12) with the kernel weights being replaced by the so-called neighborhood weights obtained through the tricube kernel and usually with $d \in \{1, 2\}$. Another modification is the application of an additional low-pass filter during seasonal extraction. As in X-11, period-wise smoothing of the detrended data produces an estimate of the seasonal component but in STL this estimate is temporary as its low-frequency variation is further removed with a convoluted MA and LOESS filter in order to restrain the final trend-cycle and seasonal smoothers from competing for the same variation in the data.[4] It is also worth noting that asymmetric trend-cycle and seasonal filter variants are not needed in STL as smoothing near the end-points is handled through appropriately modified neighborhood weights in the LOESS regression. The window, or filter, length of the latter remains fixed at $2h + 1$ in the terminology of (12), whereas in

[4]The convolution consists of a $3 \times \tau \times \tau$ MA filter (in X-11 terminology) trailed by a LOESS smoother with a recommended window length given by $\lceil \tau \rceil^{\text{odd}}$. This additional low-pass filtering results in a squared gain of the final seasonal LOESS smoother that remains close to zero in the vicinity of $\omega = 0$, as opposed to X-11 (see Figure 4).

X-11 the filter length shrinks gradually from $2h+1$ in the symmetric case to $h+1$ in the concurrent case. Finally, an optional outer loop provides robustness weights based on the biweight kernel for the irregular component, which is conceptually akin to X-11's extreme-value detection. The neighborhood weights are then multiplied by the robustness weights in the next run through the inner loop, that is the principle of iterated WLS estimation is applied in (12).

The current implementation is mostly in line with the recommendations given in Cleveland et al. (1990) that target "near certainty of convergence" of the trend-cycle and seasonal STL estimates. Therefore, we just mention the key deviations at this point. Two runs through the inner loop are carried out when no robustness weights are needed and 15 runs through the outer loop, each time with one run through the inner loop, are carried out otherwise. The polynomials in the LOESS objective function have orders $d = 1$ for trend-cycle and $d = 0$ for seasonal LOESS smoothers, respectively. If unspecified, the length of the trend-cycle LOESS smoother for extracting the seasonal pattern with periodicity τ is automatically set to $\lceil l_t(\tau) \rceil^{\text{odd}}$ with

$$l_t(\tau) = \left\lfloor \frac{1.5 \lfloor \tau \rfloor}{1 - 1.5 \lfloor l_s(\tau) \rfloor^{-1}} \right\rfloor, \tag{16}$$

where $l_s(\tau)$ is the length of the corresponding seasonal LOESS smoother. The latter also affects the length of the LOESS smoother in the convoluted low-pass filter, which is given by $\lceil l_s(\tau) \rceil^{\text{odd}}$. Also note that (16) already alludes to the way fractional periodicities are dealt with in STL as those are rounded down to the nearest integer, i.e. $\lfloor \tau \rfloor$ is always used, in sharp contrast to the weighted averaging principle (5) utilized in the extended AMB and X-11 approaches.

6 Structural time series models

The pretreatment and filtering methods discussed so far target data linearization and signal extraction in sequential top-down fashion. Structural time series (STS) models enable simultaneous estimation of all UCs in (1) through a bottom-up specification of the relevant dynamics, the aggregate of which then naturally constitutes the model for the observations in (1) (see Harvey, 1989, as a standard reference).

The most general trend-cycle specification within the STS framework is the local linear model given by

$$t_t = t_{t-1} + \nu_{t-1} + \xi_t, \tag{17}$$
$$\nu_t = \nu_{t-1} + \chi_t, \tag{18}$$

where $\{\xi_t\}$ and $\{\chi_t\}$ are mutually uncorrelated Gaussian white-noise processes with zero means and finite variances σ_ξ^2 and σ_χ^2, respectively. Other popular trend-cycle models can be derived directly from this general form. For example, the local level model with drift, which is characterized by a constant slope ν in (17), is obtained from setting $\sigma_\chi^2 = 0$ in (18). Similarly, the so-called smooth trend is obtained from setting $\sigma_\xi^2 = 0$ in (17).

The specification of the seasonal component in (1) is according to the West-Harrison representation (West and Harrison, 1997). Assuming $|\mathbb{S}| = 1$ and $\tau \in \mathbb{N}$ in (3) for the

ease of exposition (subsequent notations naturally extent to the case of multiple seasonal patterns with potentially fractional periodicities, see Proietti and Pedregal (2022) for more details), the general form reads

$$s_t = \mathbf{e}_{1,\tau-1}^\top \tilde{\mathbf{s}}_t,$$

where $\mathbf{e}_{k,n}$ is the k-th unit vector of length n and $\{\tilde{\mathbf{s}}_t\}$ is the $(\tau-1)$-dimensional reduced form of the τ seasonal effects given by

$$\tilde{\mathbf{s}}_t = \mathbf{D}^- \mathbf{P}^{t-1} \mathbf{s}_t,$$

where $\mathbf{s}_t^\top = (s_{1,t}, \ldots, s_{\tau,t})$ is the vector of the τ seasonal effects at time t, $\mathbf{P}$ and $\mathbf{D}$ are permutation and dimension reduction matrices given by

$$\mathbf{P} = \begin{pmatrix} \mathbf{0}_{\tau-1} & \mathbf{I}_{\tau-1} \\ 1 & \mathbf{0}_{\tau-1}^\top \end{pmatrix} \quad \text{and} \quad \mathbf{D} = \begin{pmatrix} \mathbf{I}_{\tau-1} \\ -\mathbf{1}_{\tau-1}^\top \end{pmatrix}$$

with $\mathbf{0}_n$ and $\mathbf{1}_n$ being n-dimensional column vectors of zeros and ones, respectively, and $\mathbf{I}_n$ being the n-dimensional identity matrix, and $\mathbf{D}^- = \left(\mathbf{D}^\top \mathbf{D}\right)^{-1} \mathbf{D}^\top$ is the Moore-Penrose inverse of $\mathbf{D}$. The reduced form is then assumed to follow the first-order vector autoregressive model

$$\tilde{\mathbf{s}}_t = \mathbf{T}\, \tilde{\mathbf{s}}_{t-1} + \boldsymbol{\omega}_t, \quad \mathbf{T} = \mathbf{D}^- \mathbf{P} \mathbf{D} = \begin{pmatrix} \mathbf{0}_{\tau-2} & \mathbf{I}_{\tau-2} \\ & -\mathbf{1}_{\tau-1}^\top \end{pmatrix}, \tag{19}$$

where $\{\boldsymbol{\omega}_t\}$ is a vector Gaussian white-noise process with zero mean and non-singular covariance matrix $\boldsymbol{\Sigma}_\omega$. The latter can be reparametrized as $\boldsymbol{\Sigma}_\omega = \sigma_\omega^2\, \boldsymbol{\Omega}\boldsymbol{\Omega}^\top$ for some seasonal models, where $\boldsymbol{\Omega}$ is a model-specific matrix that explicitly incorporates a stochastic zero-sum restriction in the τ seasonal effects. Its exact form is given in Proietti (2000) for the crude, dummy, Harrison-Stevens and trigonometric seasonal models.

The entire STS model is finally put into a univariate linear Gaussian state space form, which also enables the inclusion of calendar variation from (1) in straightforward fashion by adding the requisite regression variables from (2) to the system matrix of the observation equation and the associated effects contained in $\boldsymbol{\beta}$ to the state vector. Outliers can be dealt with similarly, and an automatic detection procedure (Grassi, Mazzi, and Proietti, 2018) is available, which is quite similar to the routine utilized in pretreatment model (2) and (4), being essentially a "forward-addition-backward-deletion" algorithm based upon the point-wise maximum $\tau_t^{\star 2}$-statistics derived in de Jong and Penzer (1998). The full state space model, including hyper-parameters, is finally estimated with the KFS and from maximizing its derived profile likelihood function, the implementation of which closely follows the algorithms described in Durbin and Koopman (2012).

7 Applications

We consider three real-world examples to highlight key capabilities of the methods discussed in the previous sections: daily births in France, hourly electricity consumption in Germany, and weekly initial claims for unemployment insurance in the United States. The first series is used to exemplify data pretreatment with model (2) and (4) and the extended AMB seasonal adjustment approach (Section 7.1). The second series is used to

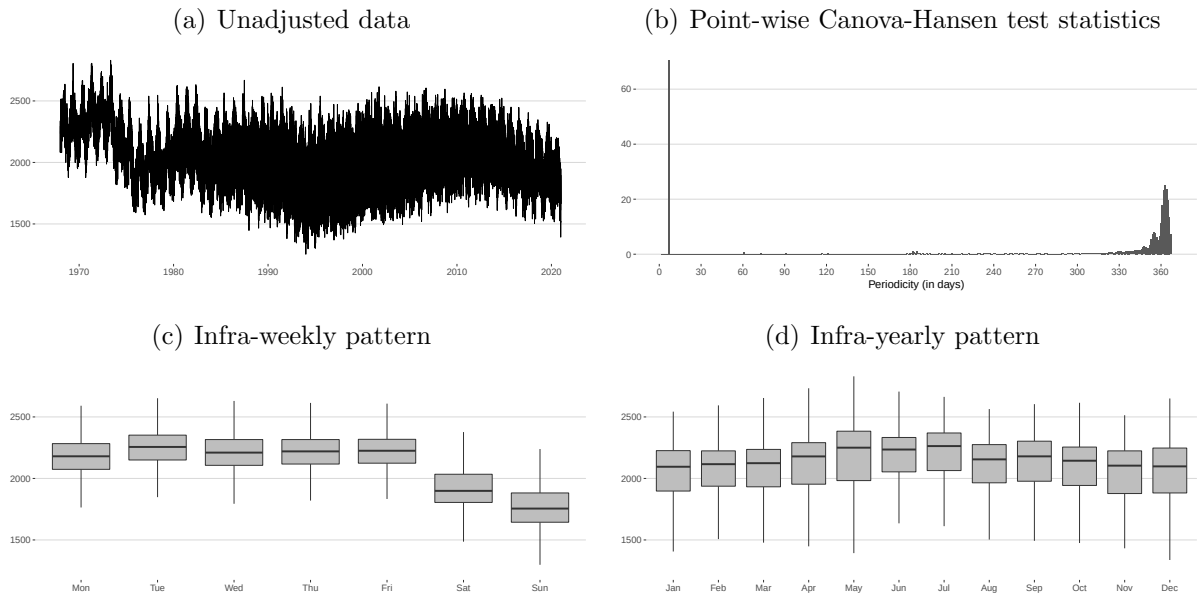


Figure 6: Seasonal profile of daily BIRTHS series.

illustrate the extended X-11 and STL approaches (Section 7.2), whereas the third series is used to demonstrate simultaneous estimation of regression effects and seasonal components with STS models (Section 7.3). Throughout these examples, code snippets are provided consecutively for the sake of brevity.

7.1 Daily births in France

Daily counts of births in France (BIRTHS) are considered from 1 January 1968 to 31 December 2020, resulting in 19,359 observations[5] (Figure 6 (a)).

7.1.1 Identification of seasonal patterns

The seasonal profile of the BIRTHS series is visualized in Figure 6 (b)–(d). For each integer periodicity $\tau \in \{2, \ldots, 367\}$, we calculate the point-wise test statistics of the generalized Canova-Hansen (CH) test (Canova and Hansen, 1995; Busetti and Harvey, 2003) of the null hypothesis of no seasonality against the alternative of either deterministic or non-stationary stochastic (trigonometric) seasonality, using the following command:

```
R> rjd3sa::seasonality.canovahansen(births,
+           p0 = 2, p1 = 367, np = 366, original = TRUE)
```

Under the null hypothesis, the distribution of the generalized CH test statistic belongs to the Cramér-von Mises family, and critical values have been tabulated in Harvey (2001). Panel (b) shows that the CH statistics peak at 7 days and around 365 days, suggesting the presence of day-of-the-week (DOW) and day-of-the-year (DOY) patterns in the BIRTHS series.

[5]The data are freely available from INSEE under URL https://www.insee.fr/fr/statistiques/5414759?sommaire=5414771 (Table T79JNAIS).

Panels (c) and (d) elaborate the appearance of the two patterns by utilizing grouped boxplots. Overall, birth counts are relatively stable from Monday through Friday but visibly lower on the weekends, especially on Sundays. It also seems that births happen more frequently during spring and summer compared to fall and winter. Based on this evidence, we set $\mathbb{S} = \{7, 365.2425\}$ in (4) during subsequent analyses.

7.1.2 Pretreatment

To correct the BIRTHS series for calendar variation, dummy regression variables are created with the companion {`rjd3modelling`} package for the following fixed and moving public holidays in France: New Year's Day (1 Jan), Easter Monday, Labor Day (1 May), Victory Day (8 May, celebrated as of 1982), Ascension Day, Whit Monday, Bastille Day (14 July), Assumption Day (15 Aug), All Saints' Day (1 Nov), Armistice Day (11 Nov) and Christmas Day (25 Dec). Storing those variables in a matrix object named `hol.FR` and assuming a multiplicative UC decomposition in (1), estimation of pretreatment model (2) and (4)—including automatic detection of temporary outliers with data-driven critical values—can be achieved with the `fractionalAirlineEstimation()` function (see Table 4 in the Appendix), using the following command:

```
R> pre.mdl <- rjd3highfreq::fractionalAirlineEstimation(
+             log(births),
+             x = hol.FR,
+             periods = c(7, 365.2425),
+             outliers = c("ao", "wo"),
+             criticalValue = 0)
```

Table 1 reports the estimated calendar effects along with standard errors and t-statistics. Overall, each public holidays has a strong dampening effect on the birth counts, which amounts, for example, to roughly -20% on Christmas Day. The estimated calendar component can be retrieved from the output object via

```
R> births.cal <- exp(pre.mdl$model$X[, 1:n] %*% pre.mdl$model$b[1:n])
```

where `n = ncol(hol.FR)` is the number of user-defined calendar regression variables. In addition, a total of 52 outliers has been detected automatically but the estimated effects are not reported here for the sake of brevity. The estimated extended Airline model is given by

$$\delta(B)\, w_t = \underset{(0.007)}{\left(1 + 0.257\, B\right)} \underset{(0.006)}{\left(1 - 0.865\, B^{7}\right)} \underset{(0.005)}{\left(1 - 0.818\, B^{365.2425}\right)} \varepsilon_t, \tag{20}$$

where $\delta(B) = \delta_1(B)\, \delta_7(B)\, \delta_{365.2425}(B)$, $\{w_t\}$ denotes the linearized logged BIRTHS series, and standard errors are given in parentheses underneath the MA parameter estimates.

7.1.3 Seasonal adjustment

The DOW and DOY patterns are extracted from the linearized logged BIRTHS series with the extended AMB approach in ascending order of their periodicities to minimize risks associated with confounding. This can be achieved through a two-step application of the

Table 1: Estimated calendar effects for BIRTHS series

Event	Estimate	Standard error	t-value
New Year's Day	−0.129	0.0063	−20.3479
Easter Monday	−0.123	0.0056	−22.0609
Labor Day	−0.162	0.0043	−37.3471
Victory Day	−0.163	0.0058	−28.1360
Ascension Day	−0.179	0.0043	−41.8757
Whit Monday	−0.113	0.0056	−20.2993
Bastille Day	−0.175	0.0043	−40.7100
Assumption Day	−0.111	0.0056	−19.8978
All Saints' Day	−0.136	0.0056	−24.2569
Armistice Day	−0.140	0.0056	−24.9068
Christmas Day	−0.201	0.0057	−35.0796

`fractionalAirlineDecomposition()` function (see Table 5 in the Appendix). Starting with the DOW pattern in the first step, the following command initiates a "full" UC decomposition without standard errors and back- and forecasts for the UC estimates:

```
R> amb.dow <- rjd3highfreq::fractionalAirlineDecomposition(
+             pre.mdl$model$linearized, period = 7,
+             sn = FALSE, stde = FALSE,
+             nbcasts = 0, nfcasts = 0)
```

The object `amb.dow` stores the UC estimates and the specified arguments of the AMB run. For example, the estimated DOW pattern can be obtained via

```
R> births.dow <- exp(amb.dow$decomposition$s)
```

In a second step, the DOY pattern can be extracted from the DOW-adjusted linearized logged BIRTHS series using essentially the same options as before except for the seasonal periodicity:

```
R> amb.doy <- rjd3highfreq::fractionalAirlineDecomposition(
+             amb.dow$decomposition$sa,
+             period = 365.2425, ...)
R> births.doy <- exp(amb.doy$decomposition$s)
```

Figure 7 (a) shows the `births.dow` and `births.doy` series. The former has a distinct time-varying amplitude that is largest during the mid-1990's. The latter, on the other hand, has a relatively constant amplitude but apparently undergoes some gradual structural infra-year changes after 1990. Both effects could be a reaction to the changing volatility in the BIRTHS series during the 1980–2010 period. This possible explanation is supported by the fact that the volatility of the seasonally adjusted BIRTHS series calculated as

```
R> births / (births.cal * births.dow * births.doy)
```

remains relatively constant throughout the entire data span (Panel (b)).

(a) Estimated day-of-the-week (*gray*) and day-of-the-year (*black*) patterns

(b) Unadjusted (*gray*) and seasonally adjusted (*black*) data

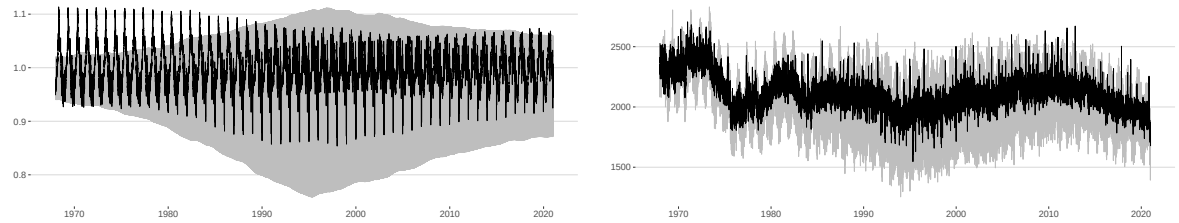


Figure 7: AMB seasonal adjustment of the BIRTHS series.

7.2 Hourly electricity consumption in Germany

Realized electricity consumption (ELCON) in Germany is considered in units of gigawatt hours (GWh) as of 1 January 2015 (00:00 AM) up to 30 June 2022 (11:00 PM), resulting in 65,712 hourly observations. The series covers electricity supplied to the network for the general supply, excluding electricity supplied to the railroad network and to internal industrial and closed distribution networks as well as electricity consumed by the producers.[6]

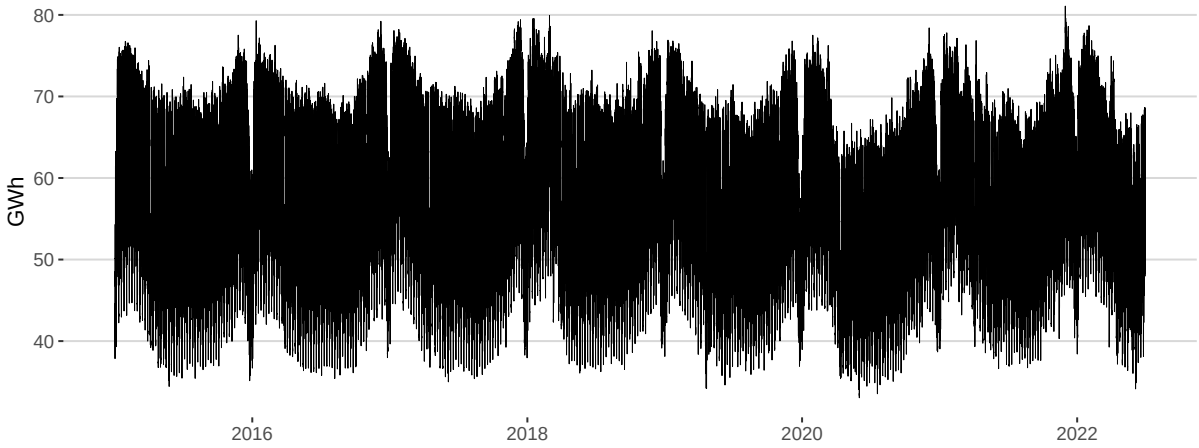
7.2.1 Data regularization and seasonal profile

Daylight saving time (DST) is in effect in Germany and makes the hourly ELCON series irregularly spaced, albeit slightly. More specifically, 24 observations are available for each day except for the last Sunday in March, when the 02:00 AM record is skipped and hence only 23 observations are available, and for the last Sunday in October, when the 02:00 AM value is recorded twice and hence 25 observations are available. Data regularization is thus required and we use the following rather simple technique: an artificial 02:00 AM value is calculated for each DST starting day as the average of the 01:00 and 03:00 AM values from this day and the two 02:00 AM values for each DST ending day are also averaged.

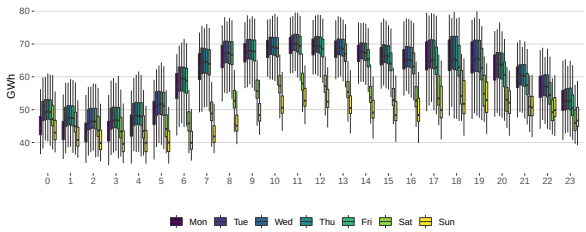
Figure 8 shows the seasonal profile of the regularized hourly ELCON series. As electricity consumption tends to be higher in winter than in summer, the U-shaped infra-yearly pattern seen in Panel (a) may not come as a surprise. In addition, V-shaped dips can be spotted easily around the turn of the years. Panel (b) reveals that, roughly speaking, during a given Monday through Friday the median hourly consumption settles between 45 and 50 GWh before 05:00 AM, then raises up to 70 GWh until lunchtime, and eventually decreases gradually over the afternoon and less gradually after 08:00 PM. The same tendency can be seen on Saturdays and Sundays, although its overall curvature is lower in level and noticeably flatter. Panel (c) confirms the existence of such infra-daily and infra-weekly patterns as the periodogram of the differenced logged ELCON series displays visible peaks at the corresponding fundamental frequencies (highlighted as gray verticals) and their harmonics.

[6]The data are freely available from the Federal Network Agency ("Bundesnetzagentur | SMARD.de") under URL `https://www.smard.de/en` and has been downloaded on 4 July 2022.

(a) Unadjusted data



(b) Infra-daily and infra-weekly patterns: grouped boxplots



(c) Infra-daily and infra-weekly patterns: periodogram (in decibel)

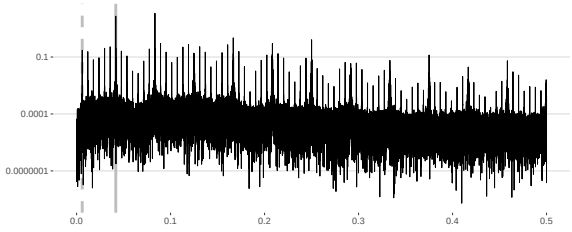


Figure 8: Seasonal profile of hourly ELCON series. Gray verticals in Panel (c) mark the chief hour-of-the-day (*solid*) and hour-of-the-week (*dashed*) frequencies.

7.2.2 Pretreatment

Preliminary empirical studies not discussed here failed to provide sufficient evidence in favor of the assumption of constant hourly calendar effects within each day in the logged regularized hourly ELCON series. To allow for varying hourly calendar effects, we could employ model (2) and (4) with an appropriate set of hourly dummy regression variables and $\mathbb{S} = \{24, 168, 8765.82\}$. However, the former set would be almost prohibitively large since nearly 400 dummy variables would be required to account for calendar variation alone. Consequently, such an approach would be computationally inefficient and most likely lead to rather unstable estimates. For that reason, we adopt a periodic regression approach instead, in line with several other studies of hourly time series (e.g. Cancelo, Espasa, and Grafe, 2008; Ramanathan, Engle, Granger, Vahid-Araghi, and Brace, 1997). To this end, we create 24 daily sub-series $\{y_t^{(h)}\}$, where the h-th sub-series contains logged regularized electricity consumption during hour $h \in \{0, \ldots, 23\}$ at day t, and use for each of those in (2) the same daily calendar regression variables and automatic outlier detection routines that have been applied in an earlier study of the companion daily ELCON series (Webel, 2022), except for the DST variables.[7] The residuals of each daily sub-series are

[7] As in Webel (2022), long-term means have been removed from regression variables for moving holidays to avoid confounding with annual seasonality.

modeled according to (4) with $\mathbb{S} = \{7, 365.2425\}$, so that the entire periodic pretreatment model is given by

$$\delta(B)\, w_t^{(h)} = \left(1 - \theta_1^{(h)} B\right) \left(1 - \theta_7^{(h)} B^7\right) \left(1 - \theta_{365.2425}^{(h)} B^{365.2425}\right) \varepsilon_t^{(h)}, \tag{21}$$

where $\delta(B)$ was defined in (20) and $\{w_t^{(h)}\}$ is the linearized h-th sub-series given by

$$w_t^{(h)} = y_t^{(h)} - \mathbf{x}_t^\top \boldsymbol{\beta}^{(h)}. \tag{22}$$

Note that in (21)–(22) infra-daily seasonality is implicitly accounted for by the varying seasonal parameters across the 24 daily sub-models. The estimated time-varying calendar effects and MA parameters are shown in Figure 9 along with point-wise standard errors (detailed results are available from the authors upon request). The estimated calendar effects display similar infra-daily curvatures that in most cases deviate visibly from the constant benchmark effects estimated for the companion daily ELCON series. The strongest estimates can usually be observed between 06:00 and 09:00 AM and—to a lesser extent—in the afternoon roughly around 03:00 PM, whereas unsurprisingly the weakest estimates tend to be recorded in the evening and during nighttime. The estimated MA parameters remain relatively constant throughout the day. In particular, the estimates for the seasonal parameters are larger than 0.8 for each hour of the day, indicating presence of strong DOW and DOY patterns in the 24 daily sub-series. The linearized versions of the latter are finally transformed back to the original ELCON scale and then combined to form the precleaned hourly ELCON series, which will be denoted by `elcon` in subsequent code snippets.

7.2.3 Seasonal adjustment

Based on visual evidence (Figure 8), we set $\mathbb{S} = \{24, 168, 8765.82\}$ for the precleaned hourly ELCON series and show how those patterns can be extracted sequentially with the extended X-11 and STL approaches, assuming a multiplicative UC decomposition.

X-11 estimates can be obtained with the `x11()` function, an overview of which is given in Table 6 in the Appendix. Starting again with the smallest periodicity, i.e. $\tau_1 = 24$ in this case, the following command runs the extended X-11 approach with a 25-term cubic Henderson trend-cycle filter, 3×9 seasonal filters and default σ-limits on the logged precleaned ELCON series:

```
R> elcon.x11.idp <- rjd3highfreq::x11(
+                   log(elcon), mul = FALSE,
+                   period = 24,
+                   trend.horizon = 12, trend.degree = 3,
+                   trend.kernel = "Henderson",
+                   trend.asymmetric = "CutAndNormalize",
+                   seas.s0 = "S3X9", seas.s1 = "S3X9",
+                   extreme.lsig = 1.5, extreme.usig = 2.5)
```

The infra-weekly and infra-yearly patterns can be estimated in similar fashion. Targeting $\tau_2 = 168$, for example, the following command runs the extended X-11 approach on the

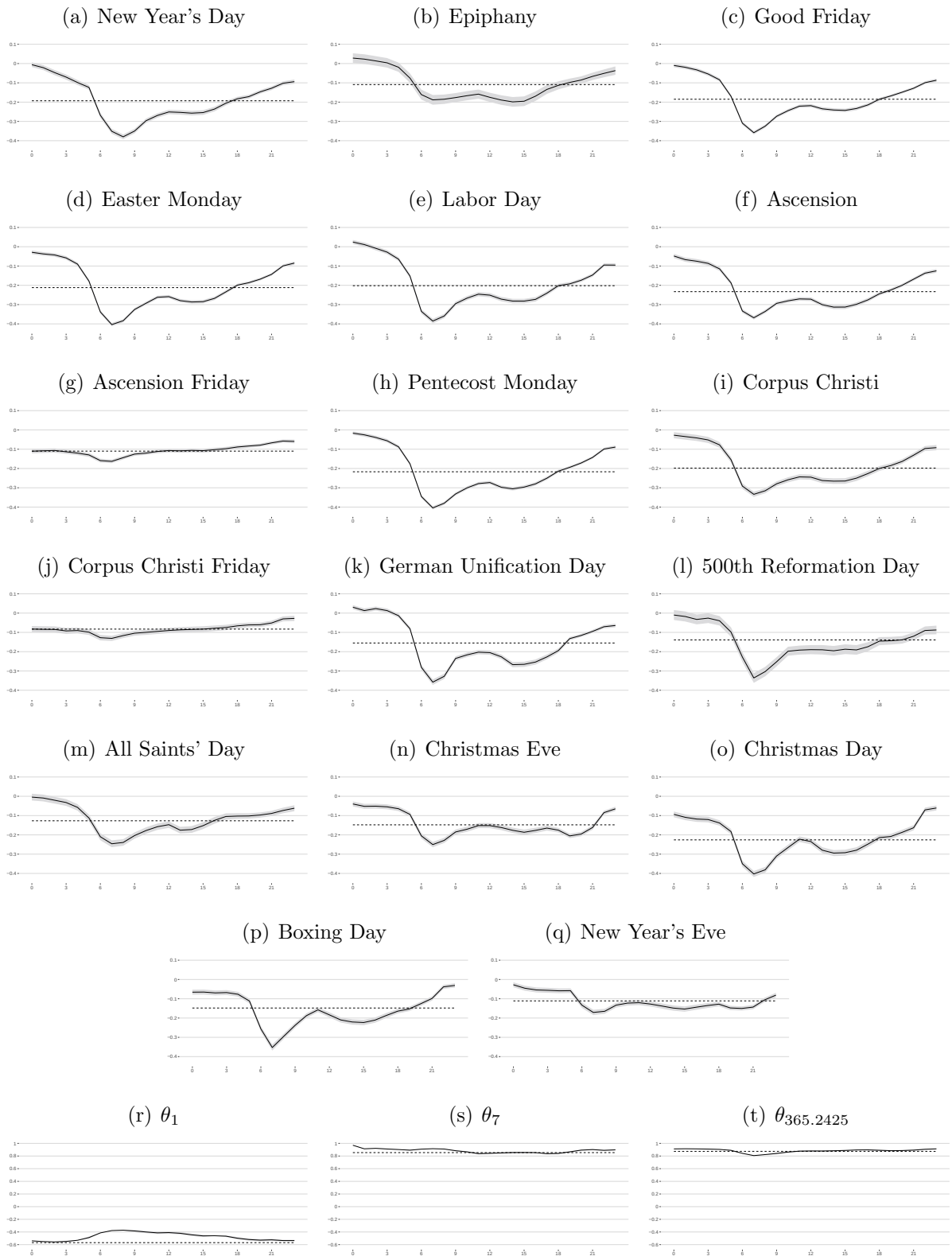


Figure 9: Estimated calendar effects and MA parameters from periodic pretreatment model (21)–(22). Shaded areas mark point-wise ±1 SE intervals. Dashed horizontals mark parameter estimates obtained for the companion daily ELCON series (see Webel, 2022).

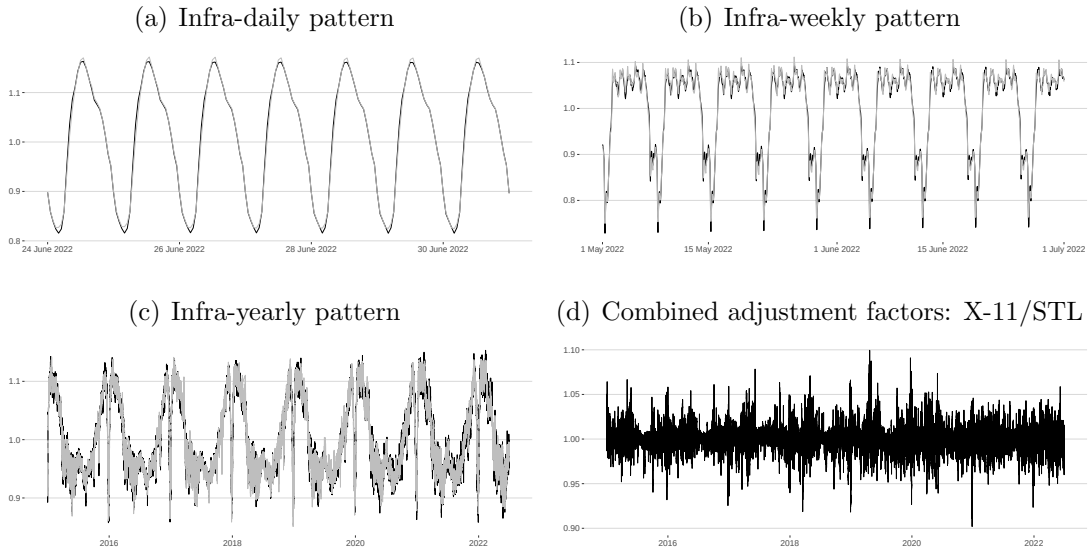


Figure 10: Estimated seasonal patterns for the hourly ELCON series obtained through the extended X-11 (*black*) and STL (*gray*) approaches.

logged precleaned ELCON series after removal of the estimated infra-daily pattern (with unchanged specification of the omitted arguments):

```
R> elcon.x11.iwp <- rjd3highfreq::x11(
+                   elcon.x11.idp$decomposition$sa,
+                   period = 168, trend.horizon = 84, ...)
```

The three estimated seasonal patterns—again rescaled to be centered around one—are shown as black lines in Figure 10 (a)–(c).

STL estimates can be obtained with the `stl()` function, see Table 7 in the Appendix. Targeting again $\tau_1 = 24$ first, the following command runs STL on the logged precleaned ELCON series without robust filter weights, a 25-term LOESS smoother for trend-cycle extraction and an 11-term LOESS smoother for seasonal extraction so that the former two match the lengths of the corresponding X-11 filters:

```
R> elcon.stl.idp <- rjd3highfreq::stl(
+                   log(elcon), multiplicative = FALSE,
+                   period = 24,
+                   swindow = 11, twindow = 25,
+                   robust = FALSE)
```

Note that the `stl()` function has fewer arguments than the `x11()` function, although both methods follow essentially the same iterative steps. The reason is that the weights of both symmetric and asymmetric trend-cycle and seasonal LOESS smoothers as well as the optional robust weights for the irregular are derived directly through prespecified kernels, whereas in X-11 the kernels for trend-cycle extraction, the derivation of asymmetric variants, the seasonal filters and the σ-limits for extreme-value detection can be selected by the user from a multitude of options.

Table 2: Average computation time for the extraction of seasonal patterns (25 replications on a 64-bit Windows OS with an Intel Core i3-8100 CPU @ 3.60 GHz and 8.00 GB RAM)

Seasonal adjustment approach		Infra-daily pattern	Infra-weekly pattern	Infra-yearly pattern
Extended X-11	with default σ-limits	0.19 sec	0.75 sec	34.42 sec
Extended STL	without robustness weights	0.24 sec	1.30 sec	87.37 sec
	with robustness weights	1.72 sec	9.41 sec	681.57 sec

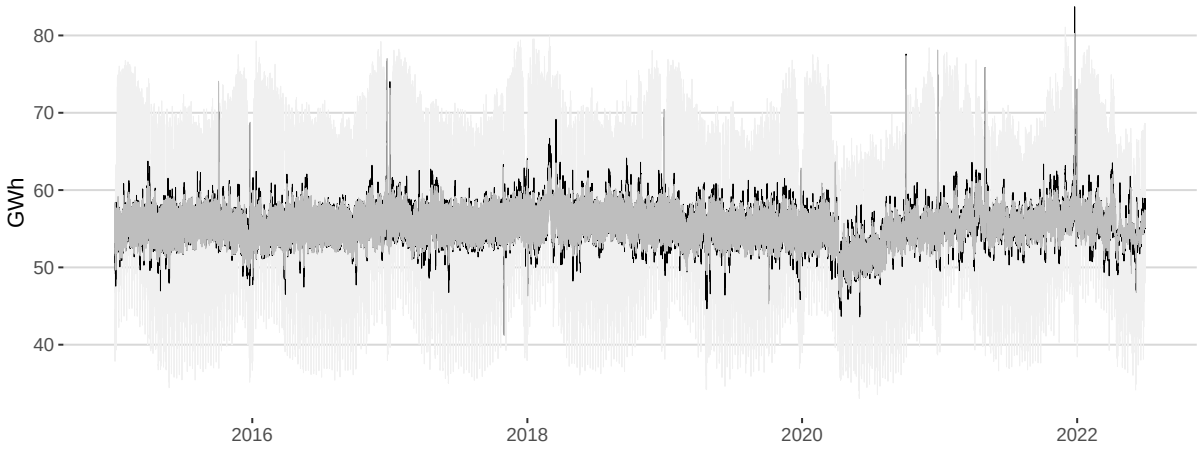


Figure 11: Seasonally adjusted hourly ELCON series obtained through the extended X-11 (*black*) and STL (*gray*) approaches. The light gray line indicates the unadjusted data.

The STL estimates of the infra-weekly and infra-yearly patterns can be obtained in the same way with appropriate changes to the relevant arguments. For example, the `period` argument is again dictated by $\tau \in \mathbb{S}$ and we could set `twindow` to $\lceil \tau \rceil^{\text{odd}}$ and `swindow` to the length of the corresponding $3 \times k$ seasonal filter selected in X-11, which were the 3×9 and 3×3 filters for the infra-weekly and infra-yearly patterns, respectively. The corresponding STL seasonal factors, which are shown as gray lines in Figure 10 (a)–(c), are quite similar to the X-11 counterparts, although the combined adjustment factors occasionally differ by up to 10 % in absolute size. Those large yet rare differences may also be spotted in two versions of the seasonally adjusted hourly ELCON series (Figure 11).

Finally, it should be noted that a repeated application of LOESS smoothers in place of linear filters with predefined weights usually needs to compromise some computation time, especially for longer regression windows. We replicated the X-11 and STL extractions of the three seasonal patterns with the above specifications a total of 25 times—with versus without robust weights in STL—and report the average computation times in Table 2. Being faster in general, X-11's margin over STL is rather slight when extracting the infra-daily and infra-weekly patterns without robustness weights but becomes substantial in any other case, keeping in mind that the total number of runs through the inner and outer loops increases by a factor of 7.5 when robustness weights are utilized.

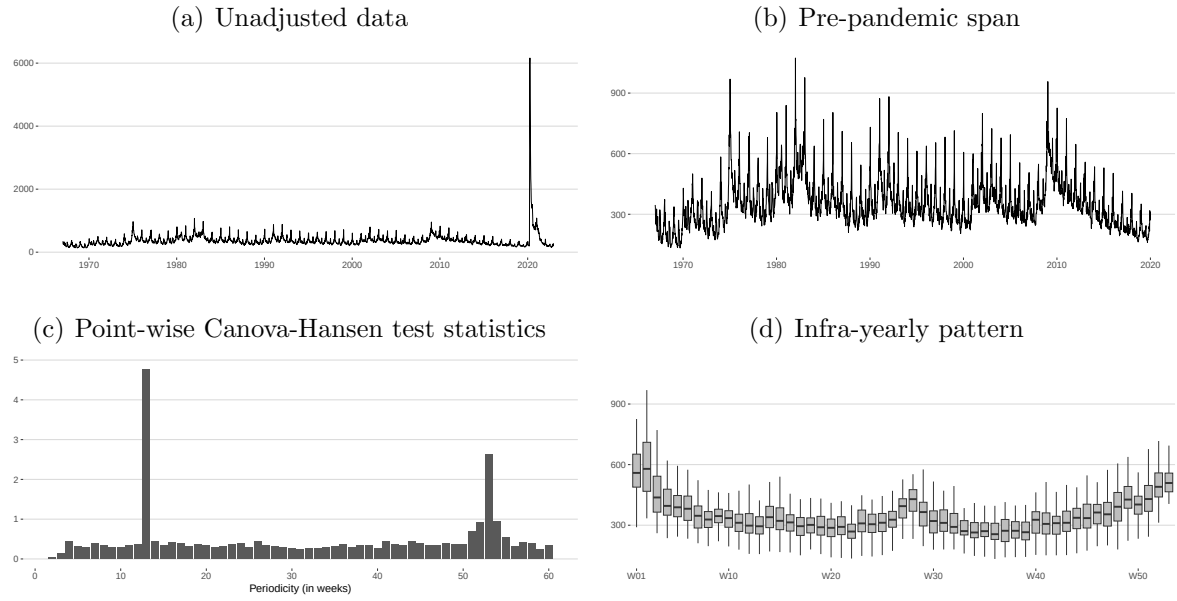


Figure 12: Seasonal profile of weekly CLAIMS series (thousands).

7.3 Weekly initial claims for unemployment insurance in the United States

We now illustrate how regression effects and seasonal dynamics can be estimated jointly within the STS framework. To this end, we consider weekly initial claims for unemployment insurance (CLAIMS) in the United States as of 1967 W01 up to 2022 W48, resulting in 2,918 weekly observations.[8] Shorter versions of this series have also been analyzed in earlier studies on weekly data (see e.g. Cleveland and Scott, 2007; Cleveland, Evans, and Scott, 2018; Proietti and Pedregal, 2022).

7.3.1 Seasonal profile

The CLAIMS series meanders roughly between 200,000 and 900,000 cases, exhibiting periodic phases of higher counts during the colder months (November through February). Is also displays some sharp increases associated with tumultuous economic times, such as the global financial crisis (2008 Q3 through 2009 Q2), and unprecedented high counts immediately after the COVID-19 pandemic outbreak in March 2020. The latter peak slightly above 6.1 million claims in 2020 W14, temporarily plateau around 1 million claims as of 2020 W31, and eventually reach pre-pandemic standards after a gradual decline over the course of the year 2021 (Figure 12 (a)–(b)).

The point-wise generalized CH test statistics clearly peak at 13 and around 53 weeks, signalling presence of strong WOY fluctuations (Figure 12 (c)). Weekly boxplots, which have been calculated from the pre-pandemic data span, confirm this impression (Figure 12 (d)). Their W-shaped nature also reveals mild mid-year increases in the CLAIMS series

[8]The data are freely available from the U.S. Department of Labor's Employment and Training Administration under URL `https://oui.doleta.gov/unemploy/claims.asp` and has been downloaded on 29 December 2022. Note that the weekly figures reflect economic activity from Sunday through Saturday.

that have been partly explained with annual model change-over practices in the automotive industry (Cleveland and Scott, 2007). Overall, the seasonal profile of the CLAIMS series is characterized by a strong WOY pattern with dominant week-of-the-quarter dynamics, so that we could set either $\mathbb{S} = \{13, 52.18\}$ or $\mathbb{S} = \{52.18\}$ in (4) and, for the sake of sparsity, opt for the latter in subsequent analyses.

7.3.2 Basic structural model and seasonal adjustment

We specify an additive basic structural model (BSM) for the untransformed CLAIMS series.[9] The trend-cycle is modeled as a local linear trend according to (17)–(18), the WOY pattern is made up of trigonometric terms, and calendar-related dynamics are captured by a set of 11 weekly dummy regression variables, one for each of the following fixed and moving holidays: New Year's Day (1 Jan)—the effect of which is generally assigned to W02 as in Cleveland et al. (2018)—, Martin Luther King Day (celebrated as of 1986), President's Day, Easter, Memorial Day, Independence Day (4 July), Labor Day, Columbus Day, Veterans Day (11 Nov), Thanksgiving, and Christmas Day (25 Dec).[10] In addition, two level shift sequences are included to capture the atypical pandemic phase in 2020 (W12 through W24 and W28 through W32).

Having stored the weekly regression variables in a matrix object named `claims.reg`, estimation of the specified BSM—including an automatic search for additive outliers and additional level shifts with length-adjusted critical values yet again—is carried out with the following command (see Table 8 in the Appendix and note that fractional periodicities are automatically rounded down to the nearest integer):

```
R> claims.bsm <- rjd3sts::sts.outliers(claims, period = 52.18,
+                 X = claims.reg, X.td = NULL,
+                 level = 1, slope = 1,
+                 seasonal = "Trigonometric", noise = 1,
+                 ao = TRUE, ls = TRUE, so = FALSE, cv = 5, tcv = 5,
+                 estimation.forward = "Full",
+                 estimation.backward = "Full")
```

The object `claims.bsm` stores various details about the estimation process and the final results. For example, the smoothed estimates of the stochastic level, slope, trigonometric WOY and white-noise irregular components can be retrieved jointly through

```
R> claims.bsm$model$components
```

whereas the deterministic calendar and outlier components can be calculated from the `claims.bsm` object using essentially the same commands illustrated for the daily BIRTHS series.

[9] An additive model for the CLAIMS series is in line with current official seasonal adjustment practices at the U.S. Bureau of Labor Statistics. In fact, the decomposition scheme has been changed from multiplicative to additive as a reaction to the COVID-19 pandemic outbreak.

[10] Except for Easter, the holiday dates have been obtained from the {`tis`} package, which lists the dates when the federal holidays are actually celebrated (i.e. in some cases the Friday before or the Monday after the holiday if the holiday happens to fall on a Saturday or Sunday, respectively). This, however, does not affect the subsequent creation of weekly holiday dummies.

Table 3: Estimated calendar effects for CLAIMS series

Event	Estimate	Standard error	t-value
New Year's Day	80.490	4.6750	17.2171
Martin Luther King Day	−47.853	5.4522	−8.7769
President's Day	−24.584	4.0977	−5.9995
Easter	8.610	2.1978	3.9174
Memorial Day	−36.027	4.4594	−8.0789
Independence Day	−9.689	4.4329	−2.1856
Labor Day	−34.003	4.3651	−7.7899
Columbus Day	−23.167	4.3656	−5.3067
Veterans Day	−36.521	4.0480	−9.0221
Thanksgiving	−82.784	4.4758	−18.4958
Christmas Day	−6.221	4.6013	−1.3520

Table 3 reports the estimated calendar effects for the weekly CLAIMS series. Thanksgiving and New Year's Day have the strongest effects, whereas the effect of Independence Day is barely significant at the 5% level of significance, and that of Christmas Day is even insignificant. The estimated calendar component is shown in Figure 13 (a), along with the smoothed trigonometric WOY pattern.

Automatic outlier detection identified 11 additive outliers and 4 additional level shifts. Seven of those outliers lie in the colder months of the pre-1990 era and might be related to exceptionally cold, or warm, winters. One level shift coincides with the maximum CLAIMS counts during the global financial crisis (2009 W02), and four outliers can be associated with the post-pandemic recovery phase. The estimated outlier component, including the two pre-specified level shift sequences, is shown in Figure 13 (b).

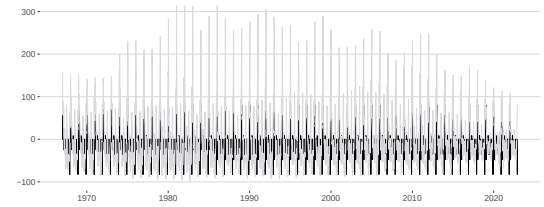
The q-ratios, that is the estimated UC innovation variances expressed as a fraction of their largest value, can be retrieved from the `claims.bsm` object through

```
R> claims.bsm$bsm$final$seasonal
```

for the trigonometric WOY pattern, and similarly for any other stochastic component. The level innovations $\{\xi_t\}$ in (17) turned out to have largest estimated variance, and the other q-ratios are 6.69×10^{-5} for the slope innovations $\{\chi_t\}$ in (18), 1.75×10^{-3} for the WOY innovations $\{\boldsymbol{\omega}_t\}$ in (19) and 0.206 for the white-noise irregular component in (1).

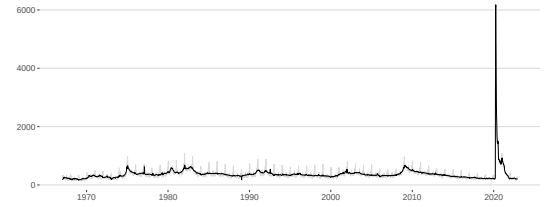
Removal of the calendar and trigonometric WOY components visibly reduces the volatility of the weekly CLAIMS series (Figure 13 (c)–(d)). Thus, the seasonally adjusted data paints a much clearer picture of the underlying ups and downs in the CLAIMS series. Moreover, it reveals some movements that were completely buried in the unadjusted data. A prime example is the short-lived increase to about 370,000 seasonally adjusted claims on average during 2005 W35–W40, i.e. an approximate 10% increase compared to 2005 W34, that results from Hurricane Katrina having struck the U.S. Gulf Coast on 29 August 2005 (see also the discussion in Cleveland et al., 2018).

(a) Calendar component (*black*) and trigonometric WOY pattern (*gray*)



(b) Outlier component

(c) Unadjusted (*gray*) and seasonally adjusted data (*black*)



(d) Unadjusted (*gray*) and seasonally adjusted data (*black*): pre-pandemic span

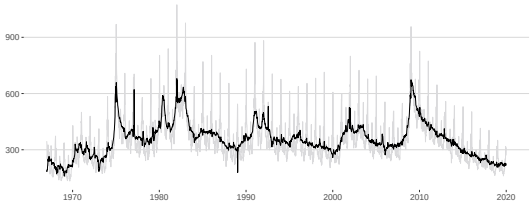


Figure 13: Smoothed estimates for various unobserved components in the weekly CLAIMS series (thousands).

8 Summary

We gave an elaborate description of the R package ecosystem for modeling and seasonally adjusting infra-monthly time series that accompanies JDemetra+ 3.0. In particular, we highlighted key modifications to official statistics' established pretreatment and seasonal adjustment approaches that are needed to take the peculiarities of such data into account. Pretreatment for outliers and calendar variation can be handled through a time series regression in which the disturbances are assumed to follow an extension of the classic Airline model that allows for multiple seasonal patterns as well as fractional powers of the backshift operator. The latter are approximated by a first-order Taylor expansion, which is essentially a weighted average of the two adjacent integer-valued powers. The same model provides the foundation for an extended ARIMA model-based seasonal adjustment approach based on classic Wiener-Kolmogorov theory, and the weighted-averaging principle is also adopted in an extended X-11 seasonal adjustment approach to maintain applicability of classic $3 \times k$ seasonal filters to time series with fractional periodicities. The latter method also contains a rich set of kernel-based trend-cycle extraction filters derived from local polynomial regressions that encompasses X-11's pioneering Henderson filters and Musgrave surrogates. Alternative empirical-based seasonal adjustments that utilize trend-cycle and seasonal LOESS smoothers can also be conducted, the implementation of which essentially emulates the classic STL method except for some minor modifications. Sequential two-step estimation of prior adjustment and seasonal factors is therefore a common strategy for these extended seasonal adjustment methods. However, simultaneous model-based signal extraction, including modeling of covariates and automatic outlier detection, can also be performed within the package ecosystem with the aid of basic structural models. Finally, we illustrated selected capabilities of the aforementioned methods using almost 20,000 observations for daily birth counts in France,

more than 65,000 observations for hourly realized electricity consumption in Germany, and almost 3,000 observations for weekly initial claims for unemployment insurance in the United States.

The ecosystem's methodological toolbox is already quite rich but still leaves room for enhancements. Currently, model-based forecasts for the observed time series can be obtained only indirectly through aggregation of the predicted UC estimates provided by the extended model-based seasonal adjustment approach but not directly from pretreatment model (2) and (4). However, we have no doubt that many users would appreciate the latter solution. Automatic tools for model calibration, such as log/level and seasonality tests as well as tailored quality diagnostics, would be another useful addition. Future research could also pay some attention to adequate data-driven selections of trend-cycle and seasonal extraction filters in extended X-11 and STL seasonal adjustments. However, the period-wise smoothing operations inherent in these two methods implicitly assume different seasonal effects for each point in time along the seasonal periodicity, which could be questioned especially for adjacent, or otherwise close, points in time. Also, saturated models, such as (19), could be over-parametrized and computationally infeasible for large seasonal periodicities. For those reasons, sparse models based on functional base components seem to be interesting alternatives but most likely bring along additional research questions. For example, an automatic state space modeling with seasonality captured through periodic cubic splines (Harvey and Koopman, 1993; Harvey, Koopman, and Riani, 1997) would require a data-driven determination of both the number and positions of the spline's knots for each seasonal pattern (a LOESS-based selection strategy is discussed in Proietti and Pedregal, 2022).

References

Burman, J. P. (1980). Seasonal Adjustment by Signal Extraction. *Journal of the Royal Statistical Society: Series A (General) 143*(3), 321–337.

Busetti, F. and A. Harvey (2003, July). Seasonality Tests. *Journal of Business & Economic Statistics 21*(3), 420–436.

Cancelo, J. R., A. Espasa, and R. Grafe (2008, October–December). Forecasting the electricity load from one day to one week ahead for the Spanish system operator. *International Journal of Forecasting 24*(4), 588–602.

Canova, F. and B. E. Hansen (1995, July). Are Seasonal Patterns Constant Over Time? A Test for Seasonal Stability. *Journal of Business & Economic Statistics 13*(3), 237–252.

Cleveland, R. B., W. S. Cleveland, J. E. McRae, and I. Terpenning (1990, March). STL: A Seasonal-Trend Decomposition Procedures Based on Loess (with comments and rejoinder). *Journal of Official Statistics 6*(1), 3–73.

Cleveland, W. P., T. Evans, and S. Scott (2018). Weekly Seasonal Adjustment: A Locally-Weighted Regression Approach. In G. L. Mazzi, D. Ladiray, and D. A. Riester (Eds.), *Handbook on Seasonal Adjustment*, Chapter 28, pp. 735–755. Luxembourg: Publications Office of the European Union.

Cleveland, W. P. and S. Scott (2007, June). Seasonal Adjustment of Weekly Time Series with Application to Unemployment Insurance Claims and Steel Production. *Journal of Official Statistics 23*(2), 209–221.

de Jong, P. and J. Penzer (1998, June). Diagnosing Shocks in Time Series. *Journal of the American Statistical Association 93*(442), 796–806.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods* (Second ed.). Oxford: Oxford University Press.

Francke, M. K., S. J. Koopman, and A. F. de Vos (2010, November). Likelihood Functions for State Space Models with Diffuse Intial Conditions. *Journal of Time Series Analysis 31*(6), 407–414.

Gasser, T. and H.-G. Müller (1979). Kernel Estimation of Regression Functions. In T. Gasser and M. Rosenblatt (Eds.), *Smoothing Techniques for Curve Estimation*, Volume 757 of *Lecture Notes in Mathematics*, pp. 23–68. Heidelberg: Springer.

Gómez, V. and A. Maravall (1994, June). Estimation, Prediction, and Interpolation for Nonstationary Series With the Kalman Filter. *Journal of the American Statistical Association 89*(426), 611–624.

Gómez, V. and A. Maravall (2001). Seasonal Adjustment and Signal Extraction in Economic Time Series. In D. Peña, G. C. Tiao, and R. S. Tsay (Eds.), *A Course in Time Series Analysis*, pp. 202–247. New York: Wiley.

Grassi, S., G. L. Mazzi, and T. Proietti (2018). Automatic Outlier Detection for the Basic Structural Time Series Model. In G. L. Mazzi, D. Ladiray, and D. A. Riester (Eds.), *Handbook on Seasonal Adjustment*, Chapter 8, pp. 169–194. Luxembourg: Publications Office of the European Union.

Gray, A. G. and P. J. Thomson (2002, March). On a Family of Finite Moving-Average Trend Filters for the Ends of Series. *Journal of Forecasting 21*(2), 125–149.

Grun-Rehomme, M., F. Guggemos, and D. Ladiray (2018). Asymmetric Moving Averages Minimizing Phase Shift. In G. L. Mazzi, D. Ladiray, and D. A. Riester (Eds.), *Handbook on Seasonal Adjustment*, Chapter 15, pp. 391–413. Luxembourg: Publications Office of the European Union.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models, and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey, A. C. (2001, January). Testing in Unobserved Components Models. *Journal of Forecasting 20*(1), 1–19.

Harvey, A. C. and S. J. Koopman (1993, December). Forecasting Hourly Electricity Demand Using Time-Varying Splines. *Journal of the American Statistical Association 88*(424), 1228–1236.

Harvey, A. C., S. J. Koopman, and M. Riani (1997, July). The Modeling and Seasonal Adjustment of Weekly Observations. *Journal of Business & Economic Statistics 15*(3), 354–368.

Jenkins, M. A. and J. F. Traub (1972, February). Algorithm 419: zeros of a complex polynomial. *Communications of the ACM 15*(2), 97–99.

Jones, R. H. (1980, August). Maximum Likelihood Fitting of ARMA Models to Time Series With Missing Observations. *Technometrics 22*(3), 389–395.

Koopman, S. J. (1993, March). Disturbance Smoother for State Space Models. *Biometrika 80*(1), 117–126.

Ladiray, D., J. Palate, G. L. Mazzi, and T. Proietti (2018). Seasonal Adjustment of Daily and Weekly Data. In G. L. Mazzi, D. Ladiray, and D. A. Riester (Eds.), *Handbook on Seasonal Adjustment*, Chapter 29, pp. 757–783. Luxembourg: Publications Office of the European Union.

Ladiray, D. and B. Quenneville (2001). *Seasonal Adjustment with the X-11 Method*, Volume 158 of *Lecture Notes in Statistics*. New York: Springer.

Levenberg, K. (1944). A Method for the Solution of Certain Non-Linear Problems in Least Squares. *Quarterly of Applied Mathematics 2*(2), 164–168.

Ljung, G. M. (1993). On Outlier Detection in Time Series. *Journal of the Royal Statistical Society: Series B (Methodological) 55*(2), 559–567.

Maravall, A. (1995). Unobserved Components in Economic Time Series. In M. H. Pesaran and M. R. Wickens (Eds.), *The Handbook of Applied Econometrics*, Volume 1, pp. 12–72. Oxford: Blackwell Publishing.

Marquardt, D. W. (1963, June). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal on Applied Mathematics 11*(2), 431–441.

McElroy, T. S. and J. A. Livsey (2022, January). Ecce Signum: An R Package for Multivariate Signal Extraction and Time Series Analysis. https://arxiv.org/abs/2201.02148.

Morf, M. and T. Kailath (1975, August). Square-Root Algorithms for Least-Squares Estimation. *IEEE Transactions on Automatic Control 20*(4), 487–497.

Morf, M., G. S. Sidhu, and T. Kailath (1974, August). Some New Algorithms for Recursive Estimation in Constant, Linear, Discrete-Time Systems. *IEEE Transactions on Automatic Control 19*(4), 315–323.

Ollech, D. (2021, July). Seasonal Adjustment of Daily Time Series. *Journal of Time Series Econometrics 13*(2), 235–264.

Ollech, D. (2023, March). Economic analysis using higher-frequency time series: challenges for seasonal adjustment. *Empirical Economics 64*(3), 1375–1398.

Proietti, T. (2000, April). Comparing Seasonal Components for Structural Time Series Models. *International Journal of Forecasting 16*(2), 247–260.

Proietti, T. and A. Luati (2008, December). Real Time Estimation in Local Polynomial Regression, with Application to Trend-Cycle Analysis. *Annals of Applied Statistics 2*(4), 1523–1553.

Proietti, T. and D. J. Pedregal (2022). Seasonality in High Frequency Time Series. *Econometrics & Statistics*. https://doi.org/10.1016/j.ecosta.2022.02.001.

Ramanathan, R., R. Engle, C. W. J. Granger, F. Vahid-Araghi, and C. Brace (1997, June). Short-run forecasts of electricity loads and peaks. *International Journal of Forecasting 13*(2), 161–174.

Shiskin, J., A. H. Young, and J. C. Musgrave (1967). The X-11 Variant of the Census Method II Seasonal Adjustment Program. Technical Paper No 15, U.S. Department of Commerce, Bureau of the Census, Washington, D.C.

Webel, K. (2022). A review of some recent developments in the modelling and seasonal adjustment of infra-monthly time series. Discussion Paper No 31/2022, Deutsche Bundesbank.

West, M. and J. Harrison (1997). *Bayesian Forecasting and Dynamic Models* (Second ed.). New York: Springer.

Wu, L. S.-Y., J. R. M. Hosking, and N. Ravishanker (1993). Reallocation Outliers in Time Series. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 42*(2), 301–313.

Appendix

The methods described and illustrated in this paper are accessible through an ecosystem of R packages available from `https://github.com/palatej`. The version of each package described in this paper is 0.3.1 and Java 17 is required to run any of those packages. Here we document the key functions that have been utilized in the applications (Section 7).

A The {rjd3highfreq} package

This package provides functions for estimation of pretreatment model (2) and (4) and signal extraction with the extended AMB, X-11 and STL approaches. Model estimation is specified through the `fractionalAirlineEstimation()` function (Table 4), and signal extraction is specified through the `fractionalAirlineDecomposition()`, `x11()` and `stl()` functions, respectively (Table 5 through Table 7).

Table 4: The `fractionalAirlineEstimation()` function

Argument	Explanation
`y`	Observed time series $\{y_t\}$ in (1)
`criticalValue`	Critical value for automatic outlier detection in (2)[1)]
`mean` (Boolean)	Add constant mean (after differencing) to $\mathbf{x}_t$ in (2)
`outliers`	Types for automatic outlier detection in $\mathbf{x}_t$ in (2)
`x`	User-defined regression variables in $\mathbf{x}_t$ in (2)
`ar` (Boolean)	Replace $\delta_1(B)$ and $\theta_1(B)$ with a stationary non-seasonal AR(1) operator in (4)[2)]
`ndiff`	Total order of non-seasonal differencing in (4)[3)]
`periods`	Set $\mathbb{S}$ of seasonal periodicities in (4)
`approximateHessian` (Boolean)	Use fast computation of the Hessian in Algorithm 1[4)]
`precision`	Convergence threshold in Algorithm 1 (rescaled by log-likelihood during Levenberg-Marquardt algorithm)

1 Setting `criticalValue = 0` calls the U.S. Census Bureau's modified formula for calculating length-adjusted critical t-values derived in Ljung (1993). **2** This option can be used to remedy numerical problems that may originate from the use of differencing operators. However, note that models with stationary AR polynomials are likely to lack an admissible decomposition, so that variance inflation will be used in such a case (see Section 3). **3** `ndiff` should not be greater than $1 + |\mathbb{S}|$ if `ar = FALSE` and $|\mathbb{S}|$ if `ar = TRUE` to facilitate model identification. Specification of any negative value automatically sets `ndiff` to the respective maximum. **4** Currently, `approximateHessian = TRUE` is always used, irrespective of the user's specification.

Table 5: The `fractionalAirlineDecomposition()` function

Argument	Explanation
`y`	Input time series
`period`	Seasonal periodicity $\tau \in \mathbb{S}$
`sn` (Boolean)	Decompose data into seasonally adjusted data (signal) and seasonal component (noise)
`stde` (Boolean)	Calculate standard deviations of component estimates
`nbcasts`	Backcast horizon
`nfcasts`	Forecast horizon

Table 6: The `x11()` function

Argument	Explanation
`y`	Input time series
`period`	Seasonal periodicity $\tau \in \mathbb{S}$
`mul` (Boolean)	Use multiplicative UC decomposition
`trend.horizon`	Bandwidth h in local trend-cycle model (11)
`trend.degree`	Polynomial order d in local trend-cycle model (11)
`trend.kernel`	Kernel weights $\{\kappa_j\}$ in objective function (12)
`trend.asymmetric`	Calculation of asymmetric trend-cycle filters $\mathbf{w}_q$ $(q < h)$ in (13)
`seas.s0`	Initial $3 \times k$ seasonal filter for Tables B5, C5, D5
`seas.s1`	Final $3 \times k$ seasonal filter for Tables B10, C10, D10
`extreme.lsig`	Lower σ-limit
`extreme.usig`	Upper σ-limit

Table 7: The `stl()` function

Argument	Explanation
`y`	Input time series
`period`	Seasonal periodicity $\tau \in \mathbb{S}$
`multiplicative` (Boolean)	Use multiplicative UC decomposition
`swindow`	Length of LOESS filter for seasonal extraction
`twindow`	Length of LOESS filter for trend-cycle extraction[1)]
`robust` (Boolean)	Use robustness weights for irregular in outer loop

1 Setting `twindow = 0` calls the automatic specification (16) (see Section 5).

B The {rjd3sts} package

This package provides functions for the specification and estimation of structural time series models. Basic structural models with automatic outlier detection are specified through the `sts.outliers()` function (Table 8).

Table 8: The `sts.outliers()` function

Argument	Explanation
`y`	Input time series $\{y_t\}$ in (1)
`period`	Seasonal periodicity $\tau \in \mathbb{S}$
`X`	User-defined regression variables in $\mathbf{x}_t$ in (2)
`X.td`	Groups for pre-defined trading-day contrasts in $\mathbf{x}_t$ in (2)[1)]
`level`	Type of level in (17)[2)]
`slope`	Type of slope in (18)[2)]
`noise`	Type of white-noise irregular in (1)[2)]
`seasonal`	Type of seasonal pattern in (19)
`ao` (Boolean)	Search for additive outliers
`ls` (Boolean)	Search for level shifts
`so` (Boolean)	Search for seasonal outliers
`cv`	Critical value for point-wise maximum $\tau_t^{\star^2}$-statistics in forward-addition step
`tcv`	Critical value for point-wise minimum $\tau_t^{\star^2}$-statistics in backward-deletion step
`estimation.forward`	Search direction for initial numerical likelihood maximization in forward-addition step[3)]
`estimation.backward`	Search direction for initial numerical likelihood maximization in backward-deletion step[3)]

1 The trading-day contrasts are internally calculated with the `rjd3modelling::td.forTs()` function. If `X.td` is used, then `y` must be provided as a time series object. **2** The component can be specified as being stochastic (`1`), fixed (`0`) or absent (`-1`). **3** Valid choices are `"Full"`, `"Point"` and `"Score"`.

C Squared gains of X-11 seasonal filters for selected fractional periodicities

Figure 14 shows the squared gains of all (effective) symmetric $3 \times k$ seasonal filters for extracting the most common periodic patterns with a fractional periodicity from infra-monthly time series. Those patterns are the day-of-the-month and day-of-the-year patterns typically seen in daily time series and the week-of-the-year pattern often observable in weekly time series. The underlying calculations are explained in Footnote 3, and the weights of the symmetric 3×3 seasonal extraction filter are stated explicitly in (14).

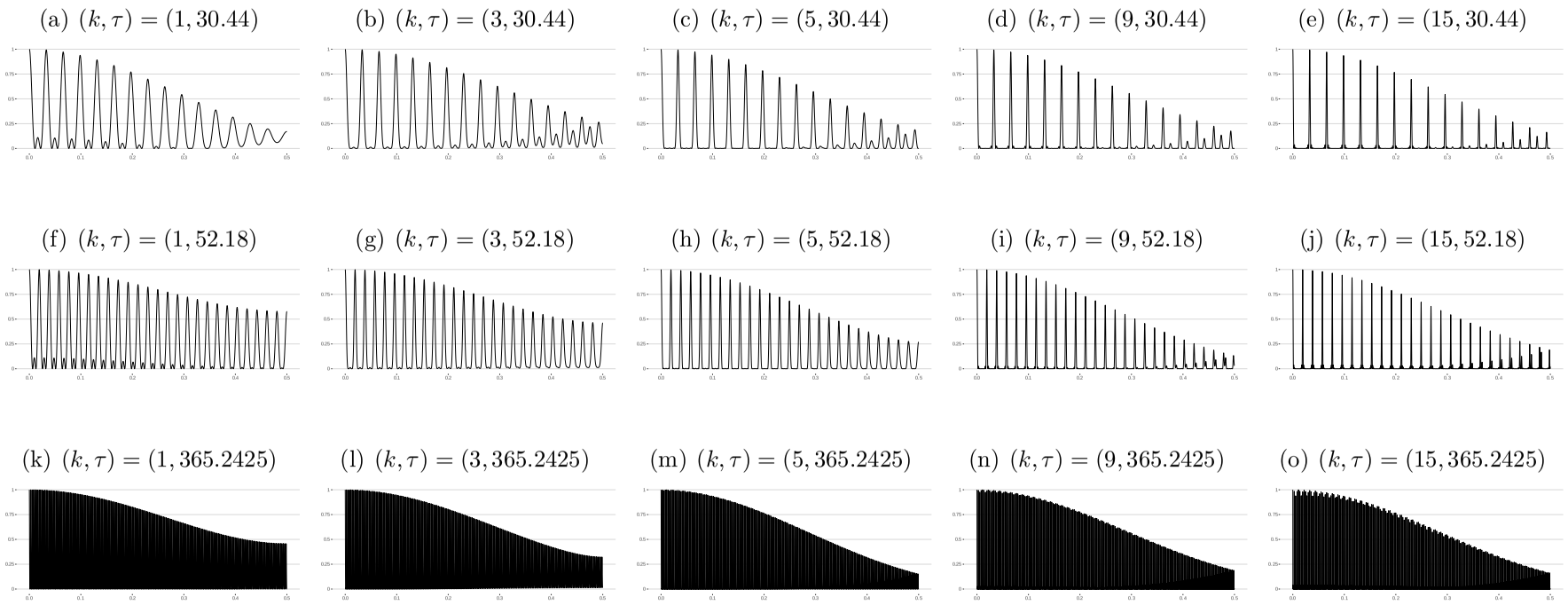


Figure 14: Squared gains of the symmetric $3 \times k$ seasonal extraction filter for $k \in \{1, 3, 5, 9, 15\}$ and $\tau \in \{30.44, 52.18, 365.2425\}$.