Econometrics and Statistics 27 (2023) 62–82

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

Seasonality in High Frequency Time Series

Tommaso Proietti[a,*], Diego J. Pedregal[b]

[a] Universitá di Roma "Tor Vergata", Dipartimento di Economia e Finanza, Via Columbia 2, Roma 00133, Italy
[b] ETSI Industrial, Universidad de Castilla–La Mancha, Ciudad Real 13071, Spain

ARTICLE INFO

Article history:
Received 19 January 2021
Revised 3 February 2022
Accepted 3 February 2022
Available online 11 February 2022

JEL classification:
C22
C52
C58

Keywords:
State Space Models
Robust filtering
Seasonal Adjustment
Variable selection

ABSTRACT

Time series observed at higher frequencies than monthly frequency display complex seasonal patterns that result from the combination of multiple seasonal patterns (with annual, monthly, weekly and daily periodicities) and varying periods, due to the irregularity of the calendar. Seasonality in high frequency data is modelled from two main perspectives: the stochastic harmonic approach, based on the Fourier representation of a periodic function, and the time-domain random effects approach. An encompassing representation illustrates the conditions under which they are equivalent. Three major challenges are considered: the first deals with modelling the effect of moving festivals, holidays and other breaks due to the calendar. Secondly, robust estimation and filtering methods are needed to tackle the level of outlier contamination, which is typically high, due to the lower level of temporal aggregation and the raw nature of the data. Finally, model selection strategies play an important role, as the number of harmonic or random components that are needed to account for the complexity of seasonality can be very large.

1. Introduction

Time series obtained from the analysis of social networks (indicators of sentiment), the system of payment (flow of electronic payments) and the internet of things (commercial traffic data) complement traditional aggregate economic data and are typically available at a high-observation frequency. The econometric literature has started investigating their contribution to macroeconomic assessment and forecasting and their use to distil timely signals that can surrogate more accurate measurements, that are nevertheless available with significant delays.

Seasonality is one of the most prominent sources of variability of time series that are observed at the sub-annual frequency. The focus of this paper is time series whose observation frequency is higher than monthly, i.e., are weekly, daily, intra-daily, and so forth. Though we will refer to regular discrete time, the methods can be extended to irregularly sampled time series.

The statistical analysis of high frequency time series must address the complexity of the seasonal pattern, which results from the superposition, and possibly interaction, of multiple periodic pattern (annual, monthly, weekly and daily cycles) and calendar effects (holidays, moving festivals). Furthermore, institutional and climatic factors, and the behaviour of economic agents (tax collection, payment of salaries, consumer preferences), determine a concentration of activity in particular days or weeks, manifesting itself with periodic spikes.

* Corresponding author.
E-mail addresses: tommaso.proietti@uniroma2.it (T. Proietti), diego.pedregal@uclm.es (D.J. Pedregal).

https://doi.org/10.1016/j.ecosta.2022.02.001
2452-3062/

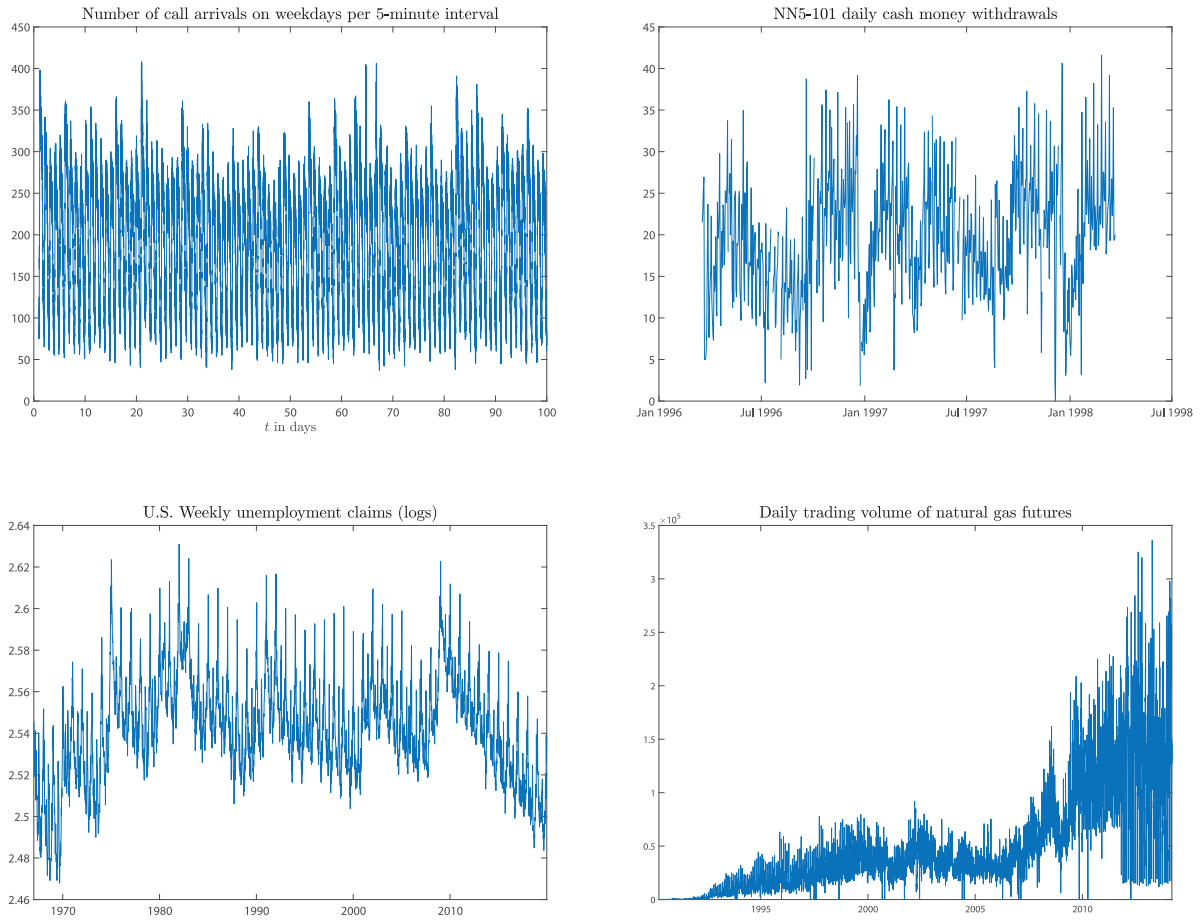

Fig. 1. Four high frequency time series. The top left panel displays the number of retail banking call arrivals per 5-minute interval between 7:00 a.m. and 9:05 p.m. during 100 days. The top right plot displays the series NN5-101, a daily time series of cash money withdrawals, for the sample period from March 18, 1996 to March 22, 1998. The bottom left plot is the weekly number of initial unemployment claims in the US (logarithms). The bottom right panel is the daily volume of natural gas futures traded on the New York Mercantile Exchange during the period between April 3, 1990 and March 6, 2014.

The number of fixed and random effects that are needed to accommodate such complexity can be very large, which poses variable selection or regularization problems of the kind that are typical of a high-dimensional inferential setting. Moreover, increasing the frequency of observations induces an incidental problem, leading to the introduction of additional higher frequency periodic components (e.g., a day of the week effect, when we move from weekly to daily data), which are aliased by temporal aggregation.

The analysis is further complicated by the fact that the calendar year does not contain an integral or constant number of weeks or days; the same holds for the months. For instance, the location or time stamp of the week within a year varies, either when the data are a systematic sample of a stock variable taken at a particular weekday, or represent the weekly total of a flow; also, holidays occurring on fixed days can fall in different weeks in different years. Hence, the seasonal period of the annual and the monthly cycles are neither constant nor integral.

Finally, high frequency data are prone to outliers, and other irregularities, such as missing data, due to their raw and unadjusted nature. Hence, the need for inferential methods that are robust to outliers, which is reinforced by the fact that the effects of outlying observations are not smoothed by temporal aggregation and that they are relatively more frequent.

These problems can be illustrated with reference to the time series displayed in Fig. 1. The first time series plotted in the top left panel is the number of retail banking call arrivals per 5-minute interval between 7:00 a.m. and 9:05 p.m. The most prominent features are an intra-daily pattern with period 169 observations and a weekly cycle with period 845 (169×5). The source is De Livera et al. (2011). It provides an interesting case study for the issue of handling multiple nested seasonal cycles.

The top right plot concerns the series NN5-101, a daily time series of cash money withdrawals, available from March 18, 1996 to March 22, 1998, for a total of 735 observations. The series is part of the NN5 forecasting competition dataset, available at http://www.neural-forecasting-competition.com. The main challenge is modelling the annual cycle, and in particular the effects of Christmas and Easter, with only two years of data, along with that of modelling the day of the week effect.

The third time series plotted in bottom left panel of Fig. 1 is the logarithm of the number of initial unemployment claims in the US. This is a weekly time series posing a number of challenges, due to the presence of annual and monthly seasonality with non-integer period. As we will see, the seasonal pattern is very complicated to interpret and a rather dense model is required.

Finally, the time series plotted in the last panel refers to the daily volume of natural gas futures traded on the New York Mercantile Exchange during the period between April 3, 1990 and March 6, 2014. The market is open only during weekdays (Monday to Friday) and a total of 5,993 observations are available. The series represents a case study in robust filtering and has been investigated by Calvet et al. (2015). It features a monthly seasonal component with an irregular period around 21 days, whose amplitude and phase have been evolving rather dramatically over time. Furthermore, the series is heavily contaminated by outlying observations.

The paper aims at addressing the above challenges. Our approach is fully parametric, being grounded in the class of unobserved components models proposed by Harvey (1989) and West and Harrison (1997), according to which seasonality is a latent component that can be modelled by suitable stochastic processes, so that we can learn about it from the available time series by computationally efficient algorithms that perform the projection of the components onto the available data. The methods have the flexibility to handle data irregularities, the possibility of encompassing both fixed and random effects. Furthermore, the signal extraction filters are easily robustified against outliers.

Our treatment draws from three essential references. A key reference is the article by Pierce et al. (1984), which introduced the idea of modelling weekly data starting from a model formulated at the daily level; moreover, it offered solutions for handling components with time-varying and non integral periods. Finally, it achieved parsimony in modelling the annual and monthly cycles by careful modelling of the holiday component. Secondly, Harvey et al. (1997) considers modelling complex and multiple seasonal patterns in daily data using time-varying periodic splines, within an unobserved components framework with multiple sources of error. The contribution by De Livera et al. (2011) deals with forecasting time series with complex seasonal patterns by a class of models featuring multiple stochastic components driven by a single disturbance. It introduces also a variable selection method for the identification of the relevant harmonic cycles. Obviously, other approaches are available in the literature; for the nonparametric approach see Ladiray et al. (2018), and the references therein.

The approach taken in this paper is based on a linear mixed model, with random and fixed effects representing trends, periodic and calendar components. The specifications adopted are encompassed by the following representation:

$$y_t = \mathbf{z}_t'\boldsymbol{\alpha}_t + \mathbf{x}_t'\boldsymbol{\beta} + \epsilon_t, \quad \mathbf{z}_t'\boldsymbol{\alpha}_t = \mu_t + \gamma_t. \tag{1}$$

The linear combination of the elements of the vector random process $\boldsymbol{\alpha}_t$ with possibly time-varying coefficients $\mathbf{z}_t$ yields the sum of the trend component, μ_t and the seasonal component γ_t. For the former we will consider the stochastic linear trend representation

$$\begin{aligned} \mu_{t+1} &= \mu_t + \varsigma_t + \eta_t, \qquad \eta_t \sim \text{i.i.d. } \mathrm{N}(0, \sigma_\eta^2), \\ \varsigma_{t+1} &= \varsigma_t + \zeta_t, \qquad \zeta_t \sim \text{i.i.d. } \mathrm{N}(0, \sigma_\zeta^2). \end{aligned} \tag{2}$$

The following sections will concentrate on the specifications of γ_t, primarily, that may be just a single stochastically evolving seasonal component as is the case of monthly data or a mixture of multiple seasonal components, as yearly, monthly, and daily periodic components. The regressors in $\mathbf{x}_t$ capture calendar effects, such as time varying holidays, interventions and exogenous variables. The calendar effects, expressed in deviation from the long run component, will contribute to the total seasonality. Finally, it will be assumed that the error term ϵ_t is normally distributed with mean zero and variance σ_ϵ^2, denoted $\epsilon_t \sim \text{i.i.d. } \mathrm{N}(0, \sigma_\epsilon^2)$, independently of any other random effect in the model, and of $\mathbf{x}_t$.

The plan of the paper is the following. We set off by considering the case in which γ_t is a deterministic component with regular period (section 2). The Fourier series representation and the fixed effects dummy seasonal model offer two approaches to modelling a deterministic periodic function of time. Section 3 exposes the stochastic harmonic approach, such that the seasonal component results from the sum of trigonometric cycles whose coefficients have random walk type evolution. In section 4 we consider the time-domain random effects model: we start with the case when a random effect is associated with each season, derive an encompassing specification which nests the stochastic harmonic model, and we finally express the seasonal cycle in terms of a smaller set of random effects.

Section 5 deals with the specification of the regression component in (1). In particular, it deals with the construction of intervention variables that are suitable to capture the effect of holidays and moving festivals. The estimation of the model parameters and of the unobserved components is dealt with in section 6, which also deals with robust estimation and filtering. Section 7 proposes model selection strategies for the stochastic harmonic model and the time-domain random effects model. The four time series displayed in Fig. 1 are analyzed in section 8. Section 9 concludes the paper.

2. Deterministic seasonality

The function $\gamma : t \in \mathbb{R} \to \gamma(t) \in \mathbb{R}$ is said to be periodic with period s if it repeats itself over intervals of fixed length, called the period. Denoting the latter by s and assuming that it is fixed, then s is the smallest number such that $\gamma(t) = \gamma(t + ks), k \in \mathbb{Z}$.

Table 1
Trigonometric cyclical components in the weekly cycle for hourly data.

Component	1	2	3	4	5	6	7	⋯	14	⋯	21	⋯	28	⋯	84
Frequency	$\frac{\pi}{84}$	$\frac{\pi}{42}$	$\frac{\pi}{28}$	$\frac{\pi}{21}$	$\frac{\pi}{16.8}$	$\frac{\pi}{14}$	$\frac{\pi}{12}$	⋯	$\frac{\pi}{6}$	⋯	$\frac{\pi}{4}$	⋯	$\frac{\pi}{3}$	⋯	π
Period	168	84	56	42	33.6	28	24	⋯	12	⋯	8	⋯	6	⋯	2
Cycles per week	1	2	3	4	5	6	7	⋯	14	⋯	21	⋯	28	⋯	84

According to the Fourier series representation of a periodic function, any piecewise continuous function $\gamma(t)$ with period s, satisfying $\int_0^s \gamma(t)dt = 0$ and $\int_0^s |\gamma(t)|^2 dt < \infty$, can be expressed as an infinite linear combination of sinusoids with frequencies $\lambda_j = \frac{2\pi j}{s}$, $j = 1, 2, \ldots$, i.e.,

$$\gamma(t) = \sum_{j=1}^{\infty} \left[b_j \cos(\lambda_j t) + b_j^* \sin(\lambda_j t) \right].$$

The sine and cosine functions evaluated at the seasonal frequencies constitute an orthogonal basis, and thus $b_j = s^{-1}2\int_0^s \gamma(t)\cos(\lambda_j t)dt$, and $b_j^* = s^{-1}2\int_0^s \gamma(t)\sin(\lambda_j t)dt$.

Assume that the periodic function is observed at equally spaced times labelled $t = 1, 2, \ldots, n$. Then, the smallest period that is observed is 2, corresponding to the frequency π. All the frequencies $\lambda_j, j > \lfloor s/2 \rfloor$ are aliased (i.e., they are indistinguishable from $\lambda_k, k = 1, \ldots, \lfloor s/2 \rfloor$). Here $\lfloor s/2 \rfloor$ equals $s/2$ for s even, and $(s-1)/2$, for s odd.

Then, a deterministic periodic series can be represented as the sum of $\lfloor s/2 \rfloor$ cycles defined at the seasonal frequencies λ_j:

$$\gamma_t = \sum_{j=1}^{\lfloor s/2 \rfloor} \gamma_{jt}, \qquad \gamma_{jt} = \alpha_j \cos(\lambda_j t) + \alpha_j^* \sin(\lambda_j t).$$

λ_1 is the fundamental frequency, and it always resides between 0 and $\pi/\lfloor s/2 \rfloor$. The remaining ones are the harmonic frequencies and are obtained as integral multiples of λ_1. If $\gamma(t)$ is band-limited, meaning that $b_j = b_j^* = 0$ for $j > \lfloor s/2 \rfloor$, then $\alpha_j = b_j$ and $\alpha_j^* = b_j^*$ for $j \leq \lfloor s/2 \rfloor$. When s is even the sine term disappears for $j = s/2$, so the number of trigonometric terms is always $s - 1$.

For instance, a weekly cycle in daily data has $s = 7$, so that $\gamma_t = \gamma_{1t} + \gamma_{2t} + \gamma_{3t}$. The three components are defined at the frequencies $\lambda_1 = 2\pi/7$, $\lambda_2 = 4\pi/7$, and $\lambda_3 = 6\pi/7$, corresponding respectively to 1, 2, and 3 cycles per week, each with period 7, 3.5 and $2.\bar{3}$ days. With hourly observations, the weekly cycle is defined at the fundamental frequency $\pi/84$ and has 83 harmonic components, as illustrated by Table 1.

The daily cycle is nested in the weekly one. It should be noticed that the 7th harmonic coincides with the fundamental frequency of the daily cycle and the harmonics $7 \times j$, $j = 1, \ldots, 12$, contribute to the daily cycle.

A deterministic seasonal component with period s can also be modelled in the time domain by defining a complete set of dummy variables, one for each season, $D_{jt}, j = 1, \ldots, s$, taking value 1 in season j and 0 otherwise. Denoting by δ_j the fixed effect of season j, we can write:

$$\gamma_t = \sum_{j=1}^{s} \delta_j D_{jt},$$

or, equivalently, $\gamma_t = \boldsymbol{w}_t'\boldsymbol{\delta}$, with $\boldsymbol{w}_t = [D_{1t}, D_{2t}, \ldots, D_{st}]'$ and $\boldsymbol{\delta} = [\delta_1, \delta_2, \ldots, \delta_s]'$.

If the model includes a trend component or the global mean, then we need to constrain the coefficients δ_j to sum to zero in order to identify seasonality (as a component in deviation from trend).

Models of stochastic seasonality can be specified in the time domain, by attributing a specific effect to a particular season, e.g., to a particular day of the week, or day of the month, or day of the year, and letting it vary over time according to a Markovian random process.

3. Stochastic harmonic approach

Following Hannan et al. (1970), a stochastic periodic component can be obtained by letting the coefficients α_j and α_j^* vary over time according to two independent random walks:

$$\begin{aligned} \gamma_t &= \sum_{j=1}^{\lfloor s/2 \rfloor} \gamma_{jt}, \\ \gamma_{jt} &= \alpha_{jt}\cos(\lambda_j t) + \alpha_{jt}^* \sin(\lambda_j t), \\ \alpha_{j,t+1} &= \alpha_{jt} + \kappa_{jt}, \\ \alpha_{j,t+1}^* &= \alpha_{jt}^* + \kappa_{jt}^*, \end{aligned} \tag{3}$$

where κ_{jt} and κ^*_{jt} are mutually independent i.i.d. $N(0, \sigma^2_{\kappa_j})$ disturbances.

The seasonal period can vary over time due to the characteristics of the Gregorian calendar: for instance, in daily data the yearly cycle has period $s_t = 365$ in regular years and 366 in leap years; the monthly cycle has period 30 days in April, June, September and November, 31 days in the remaining months, except February, where it can last 28 or 29 days. Also, with weekly time series, the annual cycle has non integer period, as the number of weeks in a year can vary from 52 to 53, with an average of 52.1775 weeks per year.

In such cases, the seasonal effect at time t arises from the combination of a set of stochastic cycles defined at the seasonal frequencies $\lambda_{jt} = 2\pi j/s_t$, $j = 1, \ldots, k$, with $k \leq \lfloor s_t/2 \rfloor$, where s_t denotes the seasonal period in days.

For a seasonal component with period s_t, the trigonometric seasonal model (3) is modified as follows:

$$\gamma_t = \sum_{j=1}^{k} \gamma_{jt}, \qquad \gamma_{jt} = \alpha_{jt} \cos(\lambda_{jt} d_t) + \alpha^*_{jt} \sin(\lambda_{jt} d_t), \tag{4}$$

$$\begin{bmatrix} @\alpha_{j,t+1} \\ \alpha^*_{j,t+1} \end{bmatrix} = \begin{bmatrix} @\alpha_{j,t} \\ \alpha^*_{j,t} \end{bmatrix} + \begin{bmatrix} @\kappa_{j,t} \\ \kappa^*_{j,t} \end{bmatrix}, \quad j = 1, \ldots, k, \tag{5}$$

where $d_t = 1, 2, \ldots, s_t$, is the time of the season at which the observation is taken. This approach is taken in Pierce et al. (1984), who however focus on the deterministic case.

The seasonal component can result from the combination of periodic components with different periodicity. For instance, with daily observations, the seasonal component may result from the sum of the weekly, monthly and annual seasonal components, so that we can write $\gamma_t = \gamma_t^{(A)} + \gamma_t^{(M)} + \gamma_t^{(W)}$, where the annual (A), monthly (M) and weekly (W) cycle are represented as (5). In particular, the weekly seasonal component can be modelled according to (5), with s_t representing the length of the month; intra-weekly seasonality can be modelled as the sum of three trigonometric cycles defined at $\lambda_1 = 2\pi/7$ (one cycle per week), $\lambda_2 = 4\pi/7$ (2 cycles per week) and $\lambda_3 = 6\pi/7$ (three cycles per week). For weekly data, $\gamma_t = \gamma_t^{(A)} + \gamma_t^{(M)}$.

If the year consisted exactly of 52 weeks and 13 months (each with 4 weeks), then the intra-monthly and intra-weekly seasonal components would result from the combination of harmonic cycles of the annual main cycle. The difficulty in disentangling the three components lies in the fact that the harmonics of the yearly and monthly seasonality occur at frequency that are very close. The same is true of the monthly and weekly cycle. Fig. 2 illustrates the distribution of the seasonal frequency of $\gamma_t^{(A)}$, $\gamma_t^{(M)}$ and $\gamma_t^{(W)}$ in a regular year. Leap years contribute to the denseness of the seasonal frequencies even further. The issue concerning the selection of the relevant harmonic cycles will be taken up in section 7.

An extension of the stochastic trigonometric model has been provided by Pedregal and Young (2006). A modulated cycle is defined as follows:

$$\gamma_{jt} = \alpha_{jt} \cos(\lambda_{jt} s_t) + \alpha^*_{jt} \sin(\lambda_{jt} s_t),$$

$$\begin{bmatrix} @\alpha_{jt} \\ \alpha^*_{jt} \end{bmatrix} = \begin{bmatrix} \cos\lambda^*_j & \sin\lambda^*_j \\ -\sin\lambda^*_j & \cos\lambda^*_j \end{bmatrix} \begin{bmatrix} @\alpha_{j,t-1} \\ \alpha^*_{j,t-1} \end{bmatrix} + \begin{bmatrix} @\kappa_{j,t} \\ \kappa^*_{j,t} \end{bmatrix}. \tag{6}$$

Here, λ^*_j is the modulating frequency (not necessarily a seasonal frequency) and $\kappa_{j,t}$ and $\kappa^*_{j,t}$ are i.i.d. Gaussian disturbances with zero mean and variance σ^2_κ. The objective of combining modulated cycle is that of fitting multiplicative periodic components keeping the number of parameters in the model much lower than in a standard unobserved components model without modulation. Pedregal and Young (2006) present an application to hourly electricity demand. Notice that if $\lambda_{jt} = 0$ we obtain the trigonometric seasonal model in Harvey (1989, p.39).

4. Time-domain random effects models

Consider modelling a regular seasonal pattern with integer period s time units in discrete time. Let $\boldsymbol{w}_t$ denote a $k \times 1$ non random vector, containing the values of $k \leq s$ deterministic periodic explanatory variables, so that $\boldsymbol{w}_{t+rs} = \boldsymbol{w}_t$, $r = 1, 2 \ldots \lfloor n/s \rfloor$.

A fixed seasonal pattern can be modelled by combining the elements of $\boldsymbol{w}_t$ with coefficients $\boldsymbol{\delta}$, such that the sum of the seasonal effects across s consecutive seasons is zero. In order to guarantee that $\sum_{j=0}^{s-1} \gamma_{t-j} = 0$, the orthogonality constraint $\bar{\boldsymbol{w}}'\boldsymbol{\delta} = 0$, where $\bar{\boldsymbol{w}} = \sum_{j=0}^{s-1} \boldsymbol{w}_{t-j}$, needs to be enforced.

The time-domain random effects model allows the coefficients to vary over time according to a (singular) random walk:

$$\begin{aligned} \gamma_t &= \boldsymbol{w}_t'\boldsymbol{\delta}_t, t = 1, \ldots, n, \\ \boldsymbol{\delta}_{t+1} &= \boldsymbol{\delta}_t + \boldsymbol{\omega}_t, \boldsymbol{\omega}_t \sim N(\boldsymbol{0}, \boldsymbol{\Omega}), \end{aligned} \tag{7}$$

where $\boldsymbol{\Omega}$ is matrix with rank $k-1$ that lies in the null space of $\bar{\boldsymbol{w}}$, and the initial vector $\boldsymbol{\delta}_1$ satisfies $\bar{\boldsymbol{w}}\boldsymbol{\delta}_1 = 0$. The singularity guarantees that $\bar{\boldsymbol{w}}'\boldsymbol{\delta}_t = 0$, for all t, and ultimately that the expected value of $\sum_{j=0}^{s-1} \gamma_{t-j} = 0$ is zero.

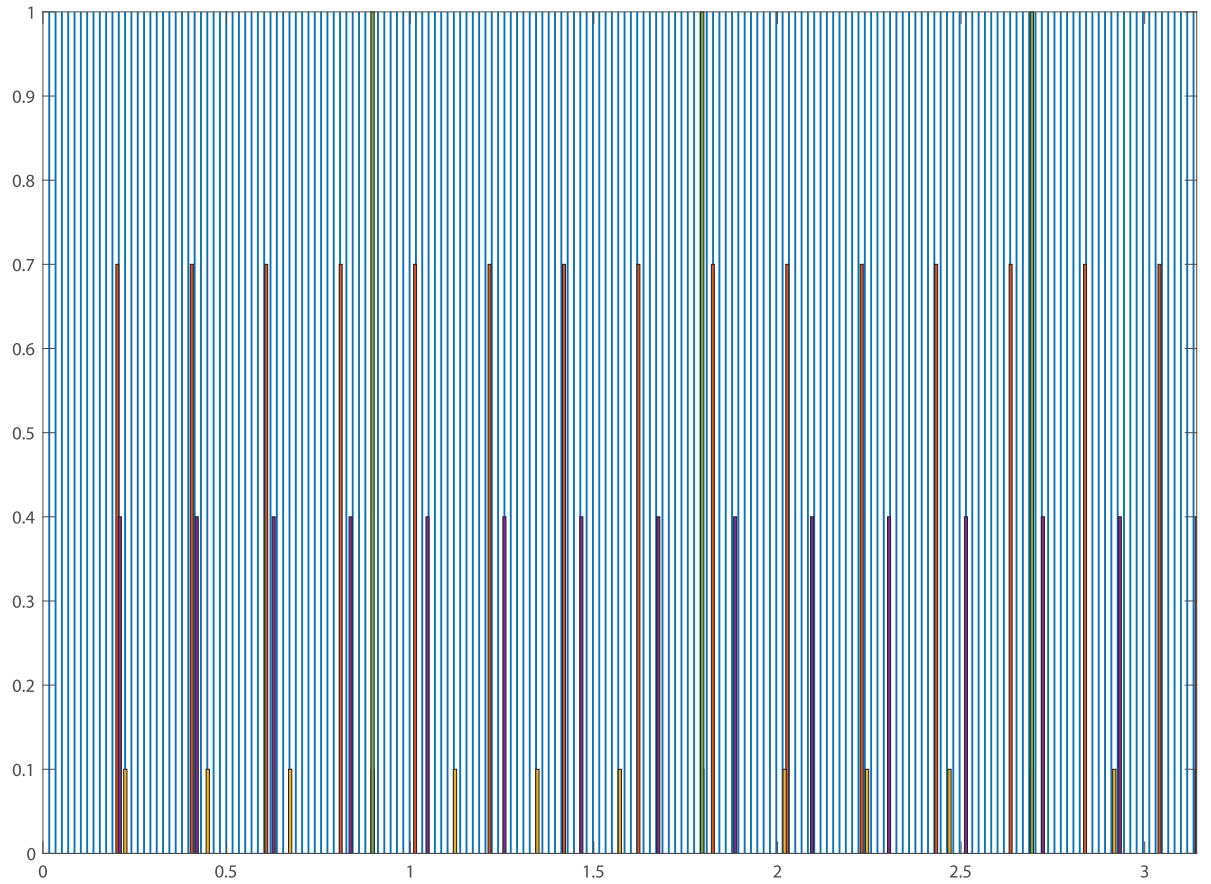

Fig. 2. Distribution in the range $[0, \pi]$ of the seasonal frequencies of the annual cycle (blue bars), of the monthly cycle with period 31 days (red bars with height 0.7) 28 days (yellow bars with height 0.1) and 30 days (bars with height 0.4), and of the weekly cycle (green bars).

There are two alternative specifications that enforce the condition $\boldsymbol{\Omega}\bar{\boldsymbol{w}} = \mathbf{0}$. For a positive definite and symmetric matrix $\boldsymbol{\nabla}$, we can set

$$\boldsymbol{\Omega} = \boldsymbol{\nabla} - \boldsymbol{\nabla}\bar{\boldsymbol{w}}(\bar{\boldsymbol{w}}'\boldsymbol{\nabla}\bar{\boldsymbol{w}})^{-1}\bar{\boldsymbol{w}}'\boldsymbol{\nabla}, \tag{8}$$

and $\bar{\boldsymbol{w}}'\boldsymbol{\delta}_1 = 0$. The vector $\boldsymbol{\delta}_t$ is a multivariate random walk, allowing the seasonal pattern to evolve over time in a persistent way, with singular distribution, as $\bar{\boldsymbol{w}}'\boldsymbol{\delta}_t = 0$.

The second alternative specification is

$$\boldsymbol{\Omega} = \boldsymbol{M}_w\boldsymbol{\nabla}\boldsymbol{M}_w, \quad \boldsymbol{M}_w = \mathbf{I}_k - \frac{\bar{\boldsymbol{w}}\bar{\boldsymbol{w}}'}{\bar{\boldsymbol{w}}'\bar{\boldsymbol{w}}}.$$

The most popular specifications for the matrix $\boldsymbol{\nabla}$ are the scalar matrix $\boldsymbol{\nabla} = \sigma^2_{\omega}\mathbf{I}_k$ and the diagonal specification $\boldsymbol{\nabla} = \mathrm{diag}(\sigma^2_{\omega,1}, \ldots, \sigma^2_{\omega,2})$, which allows for seasonal heteroscedasticity.

Model (7)-(8) implies that the sum of s consecutive seasonal values is a zero mean moving average process of order not larger than $s-2$:

$$\sum_{j=0}^{s-1}\gamma_{t-j} = \boldsymbol{w}_t'\boldsymbol{\omega}_{t-1} + (\boldsymbol{w}_t + \boldsymbol{w}_{t-1})'\boldsymbol{\omega}_{t-2} + \cdots (\boldsymbol{w}_t + \boldsymbol{w}_{t-1} + \cdots + \boldsymbol{w}_{t-s+2})'\boldsymbol{\omega}_{t-s+1}.$$

The result follows from substituting $\gamma_{t-j} = \boldsymbol{w}'_{t-j}(\boldsymbol{\omega}_{t-j-1} + \cdots + \boldsymbol{\omega}_{t-s+1} + \boldsymbol{\delta}_{t-s+1})$ and noticing $\bar{\boldsymbol{w}}'\boldsymbol{\delta}_{t-s+1} = 0$. If $\boldsymbol{\delta}_t = \boldsymbol{\delta}_1 \neq \mathbf{0}$ for all $t > 1$, then γ_t is a deterministic periodic function of time, centred at zero, $\sum_{j=0}^{s-1}\gamma_{t-j} = 0$.

We now focus on the choice of the periodic basis function, determining the loading vector $\boldsymbol{w}_t$. We start with the case when $k = s$ and $\boldsymbol{w}_t$ is an $s \times 1$ vector selecting the relevant season, which occurs when $\boldsymbol{w}_t = [D_{1t}, D_{2t}, \ldots, D_{st}]'$. It is convenient to write $\boldsymbol{w}'_t = \mathbf{e}'_1\mathbf{P}^{t-1}$, where

$$\mathbf{P} = \begin{bmatrix} @ & \mathbf{0}_{s-1}, & \mathbf{I}_{s-1} \\ & \mathbf{e}'_1 & \end{bmatrix}, \quad \mathbf{e}_1 = \begin{bmatrix} @1 \\ \mathbf{0}_{s-1} \end{bmatrix},$$

$\mathbf{I}_m$ denotes the identity matrix of order m and $\mathbf{0}_m$ a vector of m zeroes. The permutation matrix $\mathbf{P}$ is s-cyclic, so that, for any integer l, $\mathbf{P}^{k+ls}=\mathbf{P}^k$, and orthogonal, $\mathbf{P}'\mathbf{P}=\mathbf{P}\mathbf{P}'=\mathbf{I}_s$. Hence, $\boldsymbol{w}_t$ corresponds to the first column of the $(t-1)$-th power permutation, or circular shift (we assume without loss of generality that the first observation corresponds to the first season), and $\bar{\boldsymbol{w}}=\mathbf{i}_s$. If we further assume $\boldsymbol{\nabla}=\mathbf{I}_s$, we find the Harrison and Stevens (1976) seasonal model as a special case.

Appendix A establishes the equivalence between the time-domain random effects seasonal model and the stochastic harmonic model, for a suitable specification of $\boldsymbol{\Omega}$, generalizing results in Proietti (2000).

4.1. Subsets of seasonal effects

We consider the case when a subset of size $k<s$ of seasonal effects is modelled. This arises from imposing $s-k$ restrictions $\boldsymbol{C}'\boldsymbol{\delta}_t=\mathbf{0}$, where $\boldsymbol{C}$ is an $s\times(s-k-1)$ matrix whose range space does not include $\mathbf{i}_s$. This implies that $s-k-1$ effects arise as linear combinations of k principal effects. For instance, Harvey (1989, Section 2.3.5) considers the model for weekly seasonality ($s=7$) such that all weekdays (Monday to Friday) are alike, but Saturday and Sunday are different, in which case $k=3$ and

$$\boldsymbol{C}'=\begin{bmatrix}-1&1&0&0&0&0&0\\-1&0&1&0&0&0&0\\-1&0&0&1&0&0&0\\-1&0&0&0&1&0&0\end{bmatrix}.$$

Let $\boldsymbol{\Theta}$ be an $s\times k$ matrix spanning the kernel of $\boldsymbol{C}$, $\boldsymbol{\Theta}'\boldsymbol{C}=\mathbf{0}$; in the above example, $\boldsymbol{\Theta}$ is block-diagonal with diagonal blocks $\mathbf{i}_5/5$ and $\mathbf{I}_2$. The corresponding seasonal model is (7)-(8) with

$$\boldsymbol{w}_t'=\mathbf{e}_1'\mathbf{P}^{t-1}\boldsymbol{\Theta}(\boldsymbol{\Theta}'\boldsymbol{\Theta})^{-1},\qquad \bar{\boldsymbol{w}}=(\boldsymbol{\Theta}'\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}'\mathbf{i}_s.$$

If the range space of $\boldsymbol{C}$ also spans the vector $\mathbf{i}$, then $\bar{\boldsymbol{w}}=\mathbf{0}$ and the model (7)-(8) is modified by leaving $\boldsymbol{\Omega}$ unrestricted.

4.2. Time–varying Periodic Splines

The saturated seasonal model, postulating a specific effect for each of the s seasons is overparameterized and infeasible for large s. Also it needs to be adapted to the case when the seasonal period varies with time, as it is the case with the annual and monthly cycle in daily data.

Let s_t denote the seasonal period (e.g., 28 days in February, 31 days in March, etc.), t_0 denote the beginning of the season and let $u(t)=(t-t_0)/s_t$ be the rescaled time, taking values in [0,1]. A periodic spline $\gamma(u)$, is a continuous and periodic function with continuous and periodic derivatives, $\gamma^{(m)}(u)=d^m\gamma(u)/(du)^m$, $m=0,1,2$, i.e., up to the second order. We assume that $k-1$ internal knots τ_j, $j=1,\dots,k-1$, have been identified in the interval [0,1], in addition to $\tau_0=0$ and $\tau_k=1$, and we let $\delta_j=\gamma(\tau_j)$, $j=1,\dots k$, with $\gamma(\tau_0)=\gamma(\tau_k)$, due to the periodic nature of the function, and $c_j=\gamma^{(2)}(\tau_j)$.

Here we follow the derivation by Poirier (1976), Harvey and Koopman (1993) and Harvey et al. (1997). For a comprehensive treatment of splines we refer to De Boor (1978) and Hastie et al. (2001). The starting point is the periodic linear piecewise representation for the second derivative at $\tau_{j-1}\le u\le\tau_j$, $j=1,\dots,k$:

$$\gamma^{(2)}(u)=\frac{\tau_j-u}{h_j}c_{j-1}+\frac{u-\tau_{j-1}}{h_j}c_j,$$

where we have defined $h_j=\tau_j-\tau_{j-1}$, and $c_0=c_k$ by periodicity.

Integrating twice, evaluating the constant of integration using $\delta_j=\gamma(\tau_j)$, and imposing the continuity and periodicity conditions on $\gamma^{(1)}(u)$ at the knots, yields a system of k equations in the unknown elements of $\mathbf{c}=[c_1,\dots,c_k]'$, $\mathbf{Q}\mathbf{c}=\mathbf{R}\boldsymbol{\delta}$, with $\boldsymbol{\delta}=[\delta_1,\dots,\delta_k]'$,

$$\mathbf{Q}=\begin{bmatrix}2(h_1+h_2)&h_2&0&\cdots&0&h_1\\h_2&2(h_2+h_3)&h_3&\ddots&0&0\\0&h_3&2(h_3+h_4)&h_4&\ddots&0\\\vdots&\ddots&\ddots&\ddots&\ddots&0\\0&0&\cdots&h_{k-1}&2(h_{k-1}+h_k)&h_k\\h_1&0&\cdots&0&h_k&2(h_k+h_1)\end{bmatrix},$$

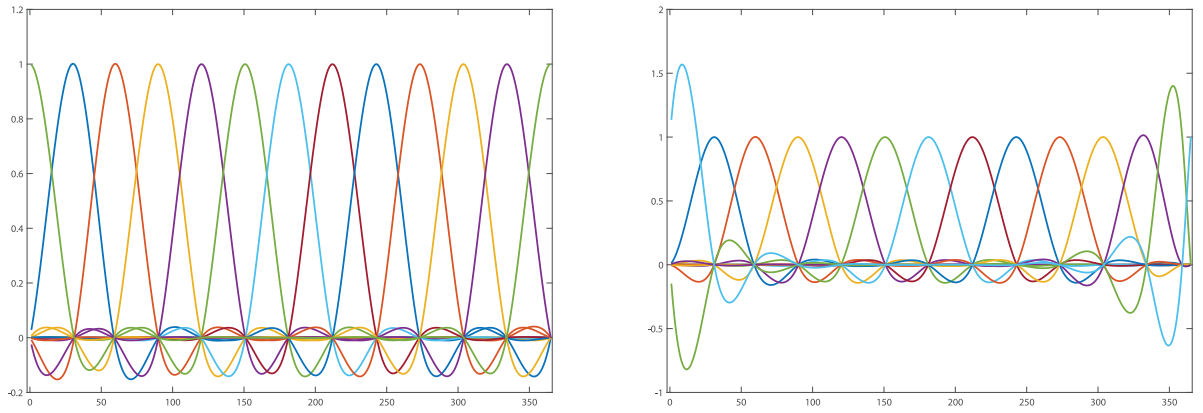

Fig. 3. Periodic spline basis. In the left panel the knots are located at the end of each calendar month [0, 31, 59, 90,..., 365]. In the right panel an additional knot has been included at t=359 for the Christmas effect.

and

$$\mathbf{R} = 6\begin{bmatrix} -(h_1^{-1}+h_2^{-1}) & h_2^{-1} & 0 & \cdots & 0 & h_1^{-1} \\ h_2^{-1} & -(h_2^{-1}+h_3^{-1}) & h_3^{-1} & \ddots & 0 & 0 \\ 0 & h_3^{-1} & -(h_3^{-1}+h_4^{-1}) & h_4^{-1} & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & h_{k-1}^{-1} & -(h_{k-1}^{-1}+h_k^{-1}) & h_k^{-1} \\ h_1^{-1} & 0 & \cdots & 0 & h_k^{-1} & -(h_k^{-1}+h_1^{-1}) \end{bmatrix}.$$

Moreover, for $\tau_{j-1} \le u \le \tau_j$, $j = 1, \ldots, k$, $\gamma(u)$ is a linear combination of the elements of $\mathbf{c}$ and $\boldsymbol{\delta}$, namely,

$$\gamma(u) = \frac{\tau_j - u}{6h_j}[(\tau_j - u)^2 - h_j^2]c_{j-1} + \frac{u - \tau_{j-1}}{6h_j}[(u - \tau_{j-1})^2 - h_j^2]c_j + \frac{\tau_j - u}{h_j}\delta_{j-1} + \frac{u - \tau_{j-1}}{h_j}\delta_j,$$

so that, after replacing $\mathbf{c} = \mathbf{Q}^{-1}\mathbf{R}\boldsymbol{\delta}$, we can write $\gamma(u) = \boldsymbol{w}(u)'\boldsymbol{\delta}$.

Hence, returning to calendar time, we can write $\gamma_t = \boldsymbol{w}_t'\boldsymbol{\delta}$, where $w_{jt} = w_j(u)$, and u is given by the ratio of the time lapsed from the start of the season to the period s_t. The spline can be made time-varying by allowing the coefficients $\boldsymbol{\delta}$ to evolve over time according to a singular random walk, as in (7). In the definition of $\boldsymbol{\Omega}$, given in (8), the vector $\tilde{\boldsymbol{w}}$ needs to be replaced by $\tilde{\boldsymbol{w}}_t = \sum_{j=0}^{s_t-1} \boldsymbol{w}_{t-j}$.

Figure 3 displays the spline basis, i.e., the elements of the vector $\boldsymbol{w}_t$, as a function of time, for $t = 1, \ldots, 365$, when the knots are located at the end of each calendar month [0, 31, 59, 90,..., 365] (left panel) and when an additional knot at 359 is inserted for capturing the effect of Christmas (right panel). One important issue is the selection of k and the location of the knots. Harvey et al. (1997) modelled γ_t for daily data by a periodic spline; in their specification $\boldsymbol{\Omega} = \boldsymbol{M}_w$.

5. Modelling holiday effects and the calendar component

The effect of fixed holidays and moving festivals can be modelled via the regression component $\mathbf{x}_t'\boldsymbol{\beta}$, where the j-th variable in the vector $\mathbf{x}_t$ contains the intervention signature of a particular holiday and β_j measures its effect. The calendar component requires careful modelling. Failure to capture it appropriately will result in the need to add high frequency harmonic cycles to the seasonal component. This is so since the variation of the series around holidays contributes strongly to the high frequency components of the spectrum. For hourly data, the switch from standard time to daylight saving time and back can be dealt with by spline interpolation.

A crucial element is the a priori specification of the intervention signature $\{x_{jt}, t = 1, \ldots, n\}$. Hereby we mention a few possibilities. In the sequel, we denote by τ the time of the intervention and by $\underline{\tau} < \tau < \overline{\tau}$ the initial and the final day between which the holiday effect occurred. Notice that all the interventions repeat periodically in different years, possibly with time varying period. Also, it is customary to consider the calendar component in deviation from the trend. This is achieved if the calendar regressors in $\mathbf{x}_t$ have zero average, which can be guaranteed by subtracting the long run average of the intervention variable.

Dummy interventions can be used to model the once for all effect on y_t, by specifying $x_{jt} = I(t = \tau)$ modulo s_t, where $I(\cdot)$ is the indicator function and s_t is the periodicity of the holiday. If the effect of a holiday spreads uniformly around its time, we can use a temporary level shift: $x_{jt} = I(\underline{\tau} < t < \overline{\tau})$ modulo s_t, or the contrast of two smooth transition interventions:

$x_{jt} = F(\underline{\tau}, \gamma_1, t) - F(\overline{\tau}, \gamma_2, t)$, where $F(\cdot)$ is the logistic function

$$F(\tau, \gamma, t) = \frac{1}{1 + \exp(-\gamma(t - \tau))}.$$

The parameter γ regulates the speed of the transition of the function from 0 to 1.

Periodic interventions may also result from combining periodic spline basis functions, as in the previous section. The problem is the localization of the spline basis function. A periodic B-spline basis is more adequate. For modelling the effect of Easter and Christmas we will consider wavelet interventions formed from the normalized first and second derivative of the Gaussian density (1-wavelet and Mexican hat) with bandwidth h:

$$\psi_1(u) = \frac{\sqrt{2}ue^{-\frac{u^2}{2h^2}}}{h^{3/2}\pi^{1/4}}, \quad \psi_2(u) = \frac{2}{\sqrt{3}h\pi^{1/4}}\left(1 - u^2/h^2\right)e^{-\frac{u^2}{2h^2}},$$

where, e.g., $u = t - t_{Easter}$. The bandwidth controls how localized is the effect of a particular festival or interventions. The 1-wavelet is a contrast between before and after the intervention; when considered in conjunction with $\psi_2(u)$ enables to capture potential asymmetries.

Unrestricted modelling can take place by the put-k-dummy-in approach by Penzer (2007), which is related to the dummy saturation approach by Santos et al. (2008), Hendry and Doornik (2014), applied to seasonal time series by Marczak and Proietti (2016). The dummy saturation approach provides also a methodology for automatic outlier detection in seasonal time series.

6. Statistical inference and robustness

The time series models for y_t considered in this paper are encompassed by the following state space specification:

$$\begin{aligned} y_t &= \mathbf{z}_t'\boldsymbol{\alpha}_t + \mathbf{x}_t'\boldsymbol{\beta} + \mathbf{g}_t'\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \text{i.i.d. N}(\mathbf{0}, \sigma^2\mathbf{I}_q), \quad t = 1, ..., n, \\ \boldsymbol{\alpha}_{t+1} &= \mathbf{T}_t\boldsymbol{\alpha}_t + \mathbf{W}_t\boldsymbol{\beta} + \mathbf{H}_t\boldsymbol{\varepsilon}_t, \end{aligned} \tag{9}$$

where in the *measurement equation* $\boldsymbol{\alpha}_t$ is an $(m \times 1)$ vector of states, $\mathbf{z}_t$ is an $m \times 1$ nonstochastic vector, e.g. $\mathbf{z}_t = [1,\ 0,\ \mathbf{w}_t']$ in the trend plus seasonal model, such that $\mathbf{z}_t'\boldsymbol{\alpha}_t = \mu_t + \gamma_t$, $\mathbf{x}_t$ is a $(K \times 1)$ vector of exogenous regressors, $\boldsymbol{\beta}$ is a $(K \times 1)$ vector. In the *transition equation* $\mathbf{T}_t$ is a $(m \times m)$ matrix, $\mathbf{H}_t$ is a $(m \times q)$ matrix, and $\mathbf{W}_t$ is $(m \times K)$. The system matrices $\mathbf{z}_t, \mathbf{g}_t, \mathbf{T}_t, \mathbf{H}_t$ are non-stochastic, can be time varying, and in general contain unknown parameters denoted by $\boldsymbol{\theta}$, referred to as hyperparameters, that have to be estimated along with the fixed effects in $\boldsymbol{\beta}$; $\mathbf{x}_t$ and $\mathbf{W}_t$ contain the exogenous measurements.

The initial state vector is specified as follows:

$$\boldsymbol{\alpha}_1 = \tilde{\boldsymbol{\alpha}}^*_{1|0} + \mathbf{W}_0\boldsymbol{\beta} + \mathbf{H}_0\boldsymbol{\xi}_0, \quad \boldsymbol{\xi}_0 \sim \text{N}(0, \mathbf{I}_q), \tag{10}$$

where $\tilde{\boldsymbol{\alpha}}^*_{1|0}$, $\mathbf{W}_0$, and $\mathbf{H}_0$ are known quantities and $\boldsymbol{\xi}_0$ is assumed to be orthogonal to $\boldsymbol{\varepsilon}_t$ at all leads and lags.

The vector $\boldsymbol{\beta}$ and the matrices $\mathbf{x}_t, \mathbf{W}_t, \mathbf{W}_0$ are partitioned as follows:

$$\boldsymbol{\beta} = \begin{bmatrix} @\boldsymbol{\alpha}_0^\dagger \\ \boldsymbol{\beta}_x \\ \boldsymbol{\beta}_w \end{bmatrix}, \quad \begin{aligned} \mathbf{x}_t' &= \left[\mathbf{0}', \mathbf{x}_t^{\dagger\prime}, \mathbf{0}'\right], \\ \mathbf{W}_t &= \left[\mathbf{0}, \mathbf{0}, \mathbf{W}_t^\dagger\right], \\ \mathbf{W}_0 &= \left[\mathbf{T}^\dagger, \mathbf{0}, \mathbf{W}_0^\dagger\right], \end{aligned}$$

where $\boldsymbol{\alpha}_0^\dagger$ are a subset of initial states corresponding to nonstationary elements of $\boldsymbol{\alpha}_t$, $\mathbf{x}_t^\dagger$ is an $(N \times k_x)$ matrix of explanatory variables affecting the response variable, $\mathbf{W}_t^\dagger$ is an $(m \times k_w)$ matrix of explanatory variables affecting $\boldsymbol{\alpha}_{t+1}$, and $\mathbf{T}^\dagger$ is a matrix relating $\boldsymbol{\alpha}_1$ to $\boldsymbol{\alpha}_0^\dagger$.

The statistical treatment of the model refers to the estimation of the unknown hyperparameters, and to the extraction of the unobserved components, feature and outlier detection. The estimation of the hyperparameters is carried out by maximum likelihood with the support of the Kalman filter. The components are extracted by a smoothing algorithm. See Harvey (1989) and Durbin and Koopman (2012).

The augmented Kalman filter (AKF), see Rosenberg (1973) and de Jong (1991), is an essential algorithm for likelihood inferences on the parameters of a state space model and for linear prediction. Given the parameter values, it evaluates the likelihood via the prediction error decomposition, and once the parameters are estimated by maximizing the likelihood, it enables the out-of-sample prediction of the series and the estimation of the states in real time.

Our treatment is based on de Jong (1989), de Jong (1991), and Durbin and Koopman (2012). In the next subsection we start from the specification of a general state space model encompassing the daily and weekly seasonal models discussed in the paper. Algorithmic details on filtering, maximum likelihood estimation and smoothing are provided in Appendix B.

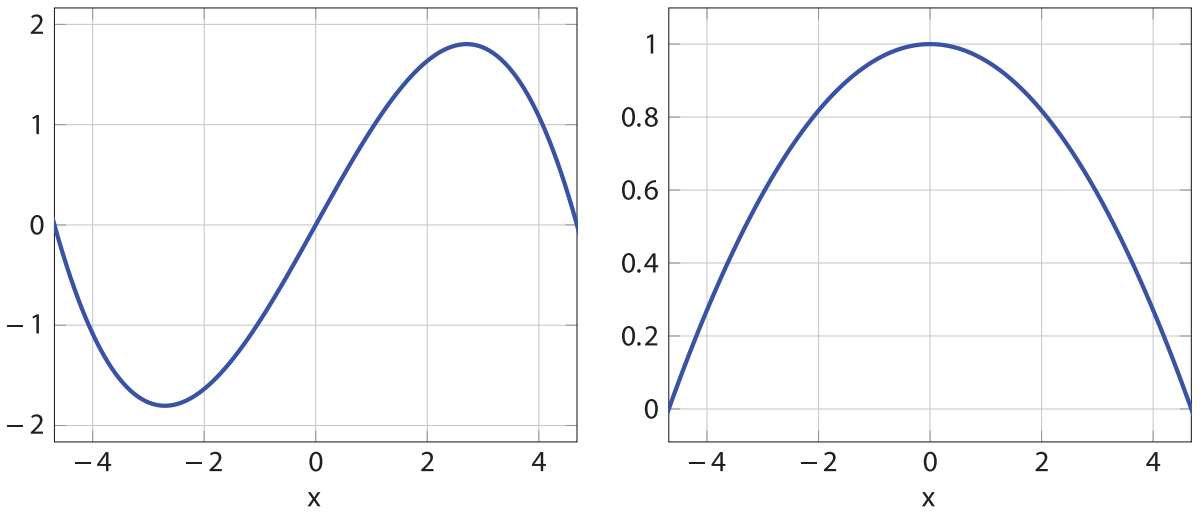

Fig. 4. Tukey's biweight function, $\psi(x)$ and its associated weight function $w(x) = \psi(x)/x$.

6.1. Robust Filtering and Signal Extraction

High frequency time series are affected by outliers. Recently, there has been renewed interest on the robustification of the inferences. This approach differs from outlier correction, which also leads to robustified inferences. Yet another alternative approach is based on endogenizing the outlier generation process by assuming that the disturbances driving the components are characterized by a heavy tailed distribution, or are distributed as a scale mixture of known distributions. See Bruce and Jurke (1996) and Aston et al. (2006) for this approach, which is perhaps more consistent theoretically, but requires computational inference based on importance sampling or Markov Chain Monte Carlo methods.

A theoretically consistent and empirically viable approach to robustness in time series analysis has been recently proposed by Harvey (2013). The approach, however, deals with unobserved components whose dynamics is driven by the conditional score of the observation density and, unlike our proposed method, cannot handle models with multiple source of errors.

In this section we present a robust filter, which draws essentially on the approach proposed by Masreliez and Martin (1977), Martin (1979) and Martin and Thomson (1982), see also Maronna et al. (2018). The idea is that in the presence of an outlying observation the robust filter shrinks the real time estimate of the state components towards the one-step-ahead prediction, not using the current observation, and replaces the observation after distilling out the contaminated part.

The influence function ψ is an essential component of the robust filter. In our applications, we shall use Tukey's biweight function which, for a variable x, is given by

$$\psi(x) = \begin{cases} x\left(1 - \frac{x^2}{c^2}\right)^2, & |x| \le c, \\ 0, & |x| > c. \end{cases}$$

The tuning constant c regulates the trade-off between the so-called breakdown point and the efficiency of the estimator. The breakdown point is a measure of robustness of an estimator, as it gives the fraction of bad data the estimator can tolerate before giving results towards the boundary of the parameter space. Lower values of c increase the breakdown point but reduce efficiency. We consider for c the values $c = 4.685$, which guarantees 95% efficiency when sampling from the normal distribution and a breakdown point about 12%, and $c = 7.0414$, achieving 99% efficienty and a 6% breakdown point; see also Maronna et al. (2018). The plot of the biweight function, $\psi(x)$, and the corresponding weight function, $w(x) = \psi(x)/x$, using $c = 4.685$, is available in Fig. 4.

The robust augmented Kalman filter extends the data– cleaning filter by Masreliez and Martin (1977), to the general state space model featuring nonstationary and regression effects. The robust filter shrinks the observations towards their one–step–ahead prediction based on the past, by bounding the effect of the information carried by a new observation according to an influence function. When maximum likelihood estimation is carried out on the replacement data, an M–type estimator is obtained.

Assume that the observation at time t is contaminated by an additive outlier, $o_t \sim \mathrm{N}(0, \lambda\sigma^2), \lambda > 0$, independently of $\boldsymbol{\alpha}_t$ and $\boldsymbol{\varepsilon}_t$, so that $y_t^+ = y_t + o_t$ is observed instead of y_t. Also, for the uncontaminated series, let $\mathrm{E}(y_t|\mathcal{F}_{t-1}) = v_t$ and $\mathrm{Var}(y_t|\mathcal{F}_{t-1}) = \sigma^2 f_t$, where $\mathcal{F}_{t-1}$ denotes the information up to and including time $t-1$. As a result, the Kalman filter innovation will be $v_t^+ = v_t + o_t$, with $v_t^+ \sim \mathrm{N}(0, \sigma^2 f_t^+)$, with $f_t^+ = f_t + \lambda$.

Consider the standardized innovation $v_t^+ = v_t^+/(\sigma\sqrt{f_t^+})$, and let $\mathrm{w}_t = \psi(v_t^+)/v_t^+$. Then, we can think of distilling the uncontaminated innovation from v_t^+ by the fraction $\mathrm{w}_t v_t^+$. Interpreting w_t as the ratio

$$\frac{\mathrm{Cov}(v_t, v_t^+)}{\mathrm{Var}(v_t^+)} \equiv \frac{f_t}{f_t^+},$$

we can retrieve

$$f_t^+ = \mathrm{w}_t^{-1} f_t.$$

The robust AKF is thus obtained by replacing f_t^{-1} by $\mathrm{w}_t f_t^{-1}$ in the updating equations for the state and the regression effects, thereby shrinking the contribution of an outlying observation. See Appendix B for details.

A cleaned time series is obtained by replacing the contaminated observations y_t^+ by

$$\tilde{y}_t = \mathbf{z}_t'\tilde{\boldsymbol{\alpha}}_{t|t} + \mathbf{x}_t'\tilde{\boldsymbol{\beta}}_t + \mathbf{G}_t\tilde{\boldsymbol{\varepsilon}}_{t|t},$$

where $\tilde{\boldsymbol{\alpha}}_{t|t}$ and $\tilde{\boldsymbol{\beta}}_t$ are the robustified real time state estimate of the states and the regression coefficients, and $\tilde{\boldsymbol{\varepsilon}}_{t|t}$ is that of the error term. The cleaned value is coincident with the observed one if $\mathrm{w}_t = 1$; if, however, w_t tends towards zero, the robust AKF shrinks the outlying observation towards the one-step-ahead prediction $\tilde{y}_{t|t-1} = \mathbf{z}_t'\tilde{\boldsymbol{\alpha}}_{t|t-1} + \mathbf{x}_t'\tilde{\boldsymbol{\beta}}_{t-1}$, where $\tilde{\boldsymbol{\alpha}}_{t|t-1}$ is the one-step-ahead prediction of the state and $\tilde{\boldsymbol{\beta}}_{t-1}$ is the estimate of the regression coefficients based only on $\mathcal{F}_{t-1}$. The sequence $\{\tilde{y}_t\}$, $t = 1, \ldots, n$, represents a cleaned data set which can be used for robust parameter estimation.

A robust M–type estimate of the parameters of the model, denoted by $\boldsymbol{\theta}$, can be obtained by the following procedure:

1. Compute the maximum likelihood estimates of $\boldsymbol{\theta}$ and obtain a robust scale estimate, e.g., replacing (B.4) by the median absolute deviation of the scaled AKF innovations:
$$\tilde{\sigma}^2_{MAD} = \left[\mathrm{med}\Big(|v_t/\sqrt{f_t} - \mathrm{med}(v_t/\sqrt{f_t})|\Big)/0.6745\right]^2,$$
where $\mathrm{med}(\cdot)$ is the median of the distribution. See Maronna et al. (2018, section 2.6).
2. Run the robust AKF to obtain a clean series $\{\tilde{y}_t, t = 1, \ldots, n\}$. The argument of the influence function and weights is $v_t/(\tilde{\sigma}_{MAD}\sqrt{f_t})$.
3. Estimate the parameters $\boldsymbol{\theta}$ by maximising the likelihood of $\{\tilde{y}_t, t = 1, \ldots, n\}$.

Steps 2–3 may be iterated until the robust AKF coincides with the AKF and no further corrections to the series are made.

7. Model selection

7.1. Model selection for the stochastic harmonic approach

The complexity of the seasonal pattern manifests itself with the necessity of bringing in a large number of harmonics or knots, leading to a large number of states and/or explanatory variables. For instance, to model the weekly cycle with minute by minute time series, 720 trigonometric components are involved ($s = 1440$). Obviously, the complete trigonometric model is infeasible, as the number of harmonics is too large to be handled, and the general trigonometric representation poses a fundamental variable selection problem, which has often been addressed by truncating the representation, i.e., selecting only the fundamental and a few first harmonics for each cycle.

The trigonometric approach is preferable when there is a single periodic component or if multiple seasonality are nested, i.e., the period of the main cycle is a multiple of that of the other seasonal components. In this case, the sines and cosines defined at the harmonic frequencies are orthogonal and the selection of the relevant trigonometric components is simplified. Moreover, for regular cycles, the harmonics are hierarchically ranked, the fundamental and the first harmonics being dominant.

The decision as to whether to include a particular harmonic lag must be accompanied by that as to whether the component is stochastically or deterministically varying. This may be thought inessential, as after all the variance of the component can be estimated equal to zero. However, if the component is deterministic there is no need to include it in the state vector and its coefficients can be estimated in closed form by generalized least squares.

Canova and Hansen (1995) and Busetti and Harvey (2003) (CHBH, henceforth) have derived the locally best invariant test of the null that seasonality is stable versus the alternative that it is stochastically evolving. The null hypothesis that the trigonometric cycle at frequency λ_j is deterministic is then formulated as $H_0 : \sigma^2_{\kappa_j} = 0$, versus $H_1 : \gamma_0 = 0, \sigma^2_{\kappa_j} > 0$. The test statistic is

$$\varpi_j = \frac{2}{n^2\hat{\sigma}_L^2}\sum_{t=1}^{n}\left[\left(\sum_{i=1}^{t} e_i \cos\lambda_j i\right)^2 + \left(\sum_{i=1}^{t} e_i \sin\lambda_j i\right)^2\right], \tag{11}$$

where e_t are the ordinary least squares residuals of the regression of the original series on a constant, $\cos(\lambda_j t)$ and $\sin(\lambda_j t)$, and $\hat{\sigma}_L^2$ is an estimator of the long run variance of the residuals. Under the null, ϖ_j is asymptotically distributed according

to a Cramér-von Mises distribution with 2 degrees of freedom; the critical values have been tabulated by Harvey (2001) in his Table I(b).

The selection method that we propose has the following steps:

1. Estimate the null model (1) with trigonometric seasonality by including in γ_t the component defined at the fundamental frequency and the first two harmonic components for each seasonal cycle (annual, monthly, etc.). Include among the explanatory variables the calendar regressors and any known interventions.
2. Compute the standardized Kalman filter innovations (see section 6), denoted v_t^+. Conduct the CHBH test, and if the largest value of the test statistic is significant at the prescribed level, after a Bonferroni correction, include the corresponding harmonic in γ_t.
3. Reestimate the model and iterate until no further stochastic cycles are included.
4. Conduct the the Wald test of the restriction $\alpha_j = \alpha_j^* = 0$ for the remaining harmonics. Include in $\mathbf{x}_t'$ the harmonics that are significant at the prescribed significance level with Bonferroni correction. This can be done by forward addition of the maximally significant harmonic, one at a time with reestimation, or by including all those which are significant. Forward addition (with possible alternation of backward deletion) is preferable in the presence of multiple non nested seasonal components.

A similar selection strategy has been adopted by De Livera et al. (2011), who proposed a forward stepwise procedure, which gradually adds deterministic cycles $\alpha_j \cos(\lambda_j t) + \alpha_j^* \sin(\lambda_j t)$ one at a time testing the significance of each one by an F-test for the null $H_0 : \alpha_j = \alpha_j^* = 0$, with nominal size 0.001. The forward addition stops when the Akaike Information Criterion has no further reduction. A selection approach based on the periodogram for seasonal long memory models is taken by Leschinski and Sibbertsen (2019).

7.2. Model selection for the time domain random effects model

When the seasonal component is modelled according to this approach we propose the following model selection strategy:

1. Filter the series with a low-pass filter to have a preliminary estimate of the trend. A model based filter based on an integrated random walk with cut-off frequency close to the zero frequency, such that its gain is close to zero at the seasonal frequencies, can be appropriate. Otherwise, use a preliminary specification of model 1.
2. Plot the detrended series against the day of the year pertaining to the observation. Estimate the conditional mean function by a nonparametric local polynomial smoother, such as LOESS; see Cleveland et al. (1990).
3. Identify the knots as the change-points of the fitted conditional mean function, i.e. the points at which level, slope and curvature change most.
4. If there are multiple seasonal components, plot the detrended series against the the day of the month, week, etc., pertaining to the observation and identify the knots as the change-points of the regression function estimated nonparametrically.

Forward addition of knots in a periodic spline setting can also be performed, though the insertion of a knot affects all other elements of the basis, rendering the selection more complicated. A periodic cubic B-spline basis can be more amenable for the purpose of model selection, being more localized.

8. Illustrations

8.1. Call Center Data

The series deals with the number of retail banking call arrivals per 5-minute interval between 7:00 a.m. and 9:05 p.m. It has been analyzed by De Livera et al. (2011). The most prominent features are the presence of daily seasonality with period $s_1 = 169$ and a weekly cycle with period $s_2 = 845$ (169×5). We also explore the possibility of a hourly cycle with period equal to 12 observations. We consider $n = 16,900$ observations over 100 days (20 weeks, 5 days a week).

The initial model features the fundamental and two harmonic seasonal cycles for the weekly (W) and daily (D) components, and only the fundamental cycle for the hourly component:

$$y_t = \mu_t + \sum_{j=1}^{3} \gamma_{jt}^{(W)} + \sum_{j=1}^{3} \gamma_{jt}^{(D)} + \gamma_{1t}^{(H)} + \epsilon_t.$$

The remaining harmonics are selected based on the Wald test of residual seasonality at the seasonal frequencies and the CHBH test. This has led to collecting in $\mathbf{x}_t$ cosine and sine terms defined at the harmonics $\{6, 9\}$ of the weekly cycles and at the harmonics $\{4, 5, 6, 8, 11, 12, 13, 26, 27\}$ of the daily cycles. It should be remarked that the higher frequencies of the daily cycles are very close to the fundamental and the harmonics of the hourly cycle.

Figure 5 displays the series with the underlying level, which is very stable (top left). The estimated weekly seasonal component captures most of the variability of the series (top right panel). The sample available, consisting only of 100 days does not allow us to verify whether the fluctuations of the amplitude are due to the presence of a monthly cycle. On the

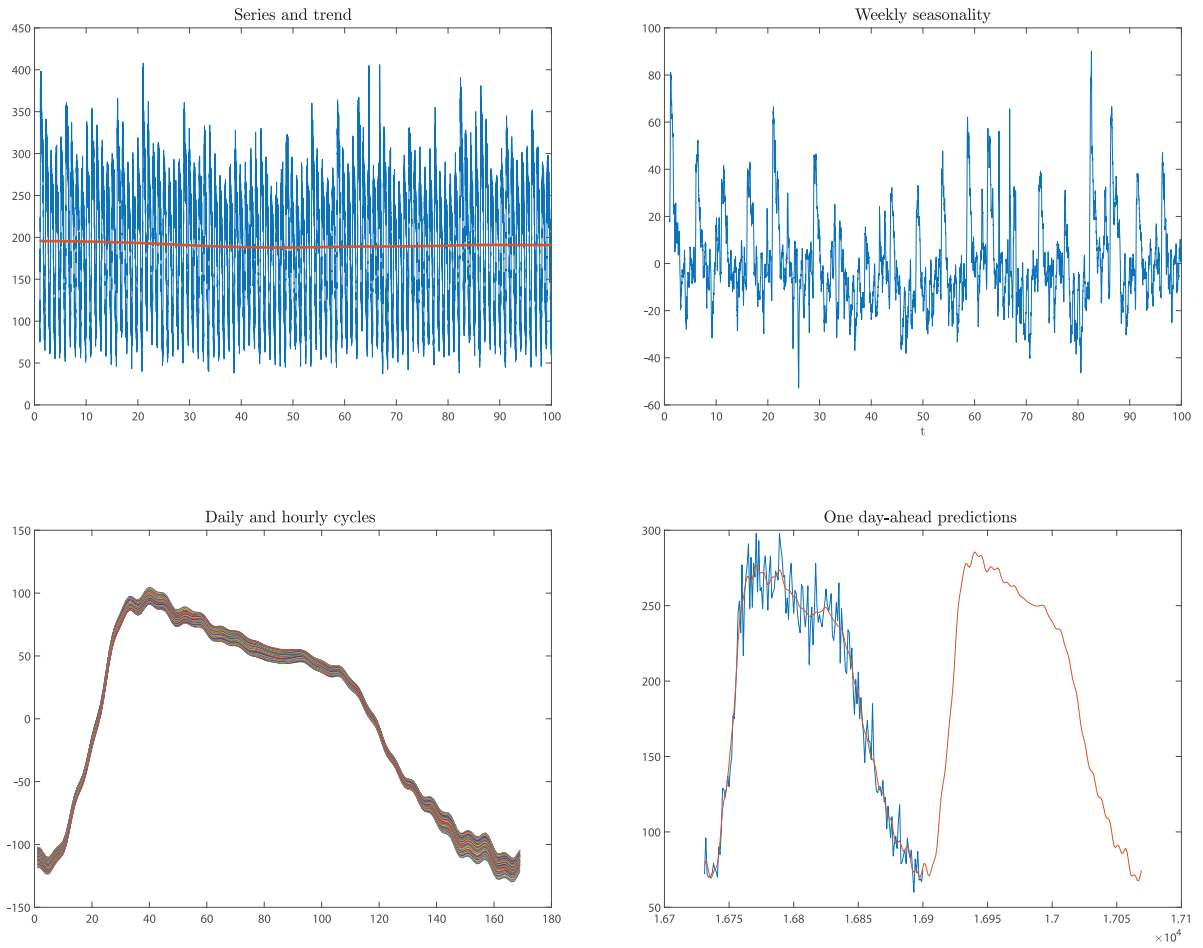

Fig. 5. Call center data. Series and smoothed estimates of level component (top left). Smoothed estimates of the weekly seasonal component (top right). Smoothed monthly and hourly seasonal components versus the 169 5-minutes intervals making up each day (bottom left). Multistep predictions for day 101 (bottom right).

contrary, the daily pattern is very regular: the bottom right plot displays the 100 estimated daily and hourly cycles versus the index of the five minutes intervals, $t = 1, \ldots, 169$. The daily and hourly components are considered together due to the inherent difficulty in separating the two components. The plot confirms that there is a detectable hourly cycle, manifesting itself with a clustering of the calls at the beginning and towards the end of each hour. The effect is more sizable at the beginning of the day. The last plot displays the smoothed estimates of the signal for the last day, with the observed number of calls superimposed, and the multistep predictions for the next day (169 predictions).

8.2. Daily cash money withdrawal

The NN5 daily cash withdrawal time series features prominent Christmas and end of year effects. Secondly, the Easter holidays need to be taken into consideration. For these effects we created two sets of wavelet interventions, as in section 5 using a bandwidth $h = 7$ (seven days).

Our initial model contains a level component, μ_t that absorbs all the low frequency components, essentially capturing the annual cycle, which could not be autonomously estimated with only two years of data, a monthly seasonal component, modelled using the fundamental trigonometric cycle and two harmonic cycles, a weekly seasonal component, also modelled considering only the first three trigonometric components, and the four intervention variables, 1-wavelet and Mexican hat for Christmas and Easter: $y_t = \mu_t + \sum_{j=1}^{3} \gamma_{jt}^{(M)} + \sum_{j=1}^{3} \gamma_{jt}^{(W)} + \mathbf{x}_t'\beta + \epsilon_t$.

Then, the variable selection method of section 7 leads to the addition of deterministic components at the frequencies corresponding to the annual, semiannual and one-third of a year cycles.

The top right panel of Fig. 6 displays the smoothed estimates of the combined level and regression components. The Christmas and Easter effects result from the combination of the wavelet interventions plotted in the right panel: the 1-wavelet is a smooth contrast between what occurred after and before the holiday; the Mexican hat wavelet is a contrast

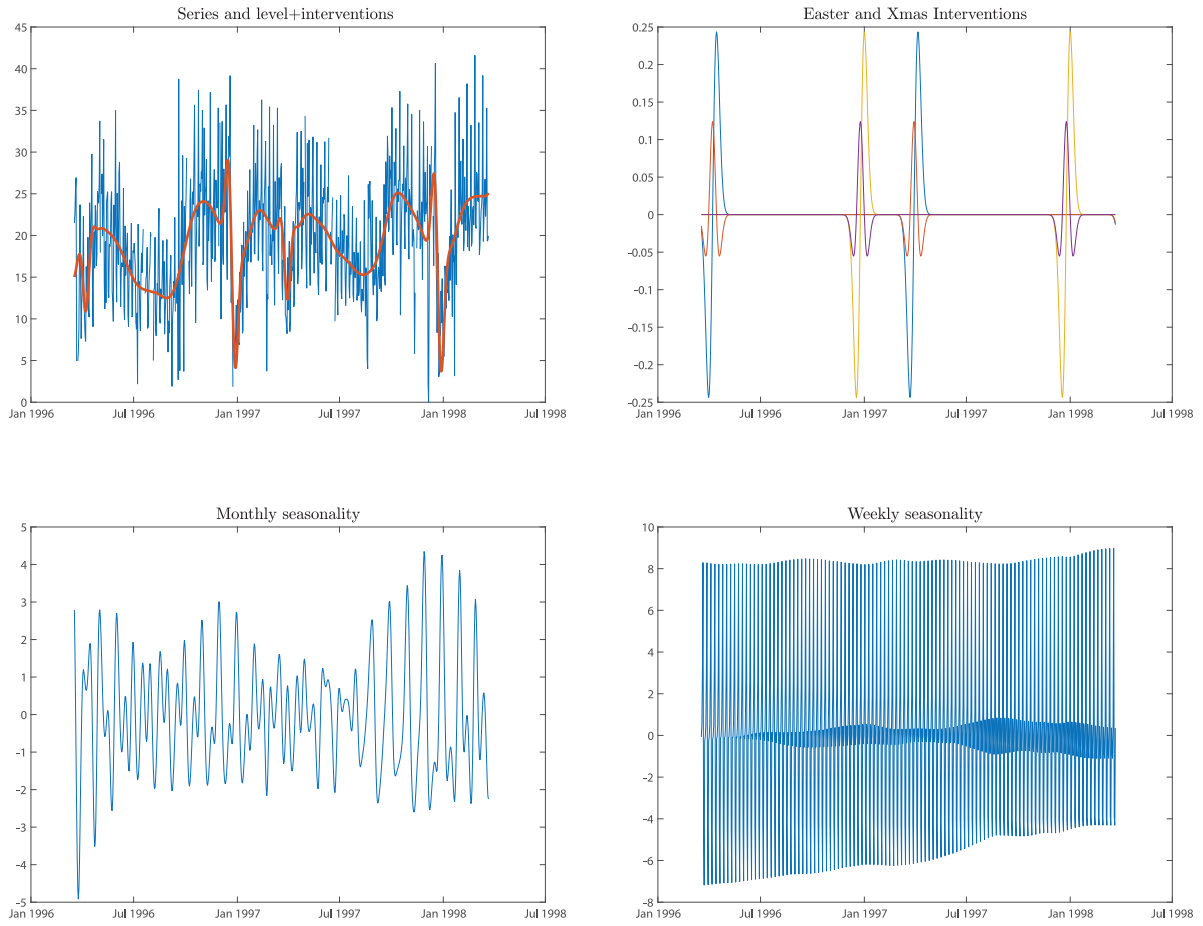

Fig. 6. NN5-101 Daily cash money withdrawal. Series and smoothed estimates of level and annual seasonal component (top left). Easter and Christmas wavelet intervention variables (top right). Smoothed monthly seasonal component (bottom left). Smoothed weekly seasonal component (bottom right).

between the value at the time of the intervention and before and after. The amplitude of the monthly component, shown in the bottom left panel, may vary with the annual cycle, being smaller in the summer months. This interaction among the multiple seasonal components is captured by the time-varying pattern of the monthly seasonal component. The weekly seasonal component, displayed in the bottom right panel of Fig. 6 is more regular, although it is responsible for a large share of the daily withdrawals variation.

8.3. US initial unemployment claims

The weekly US initial unemployment claims series is an important timely indicator of the US business cycle, as a flexible labour market is very sensitive to the state of the economy. It is characterized by a strong annual seasonal pattern, and unemployment claims are also varying within the month. For analyzing this time series we applied the strategy proposed in section 7.2. The reason why it is preferable to use a periodic spline is related to the very fast movement of the series at the beginning of the year and in July. This is better accommodated by an appropriate selection of knots. For the yearly seasonal component we selected 13 knots, the first one located at 0.0274 in rescaled annual time, and the remaining ones located at $j/12$, $j = 1, \ldots, 12$. As for the monthly cycle, the following knots were identified in rescaled monthly time: $\{0, 0.22, 0.37, 0.44, 0.58, 0.73, 0.84, 1\}$; these are the points at which the minima and maxima of the regression function of the detrended observations on rescaled time, estimated by LOESS, were located. This may be the reflection of the fact that the monthly cycles have idiosyncratic features, so that unemployment claims peak at different points of rescaled time.

Easter and Christmas wavelet interventions were also included and the null model so identified was estimated by maximum likelihood. While the Easter effect was not significant, the analysis of the standardized innovations and their sample spectrum highlighted the presence of significant residual autocorrelation and unaccounted effects at the seasonal frequencies. Rather than inserting more knots we took a hybrid strategy and we found more suitable to sequential identification of trigonometric components according to the strategy outlined in section 7.1. This led to the identification of the following harmonics of the yearly cycle: $\{8, 18, 16, 19, 22, 20\}$. Also, since we suspected a substantial interaction of the monthly

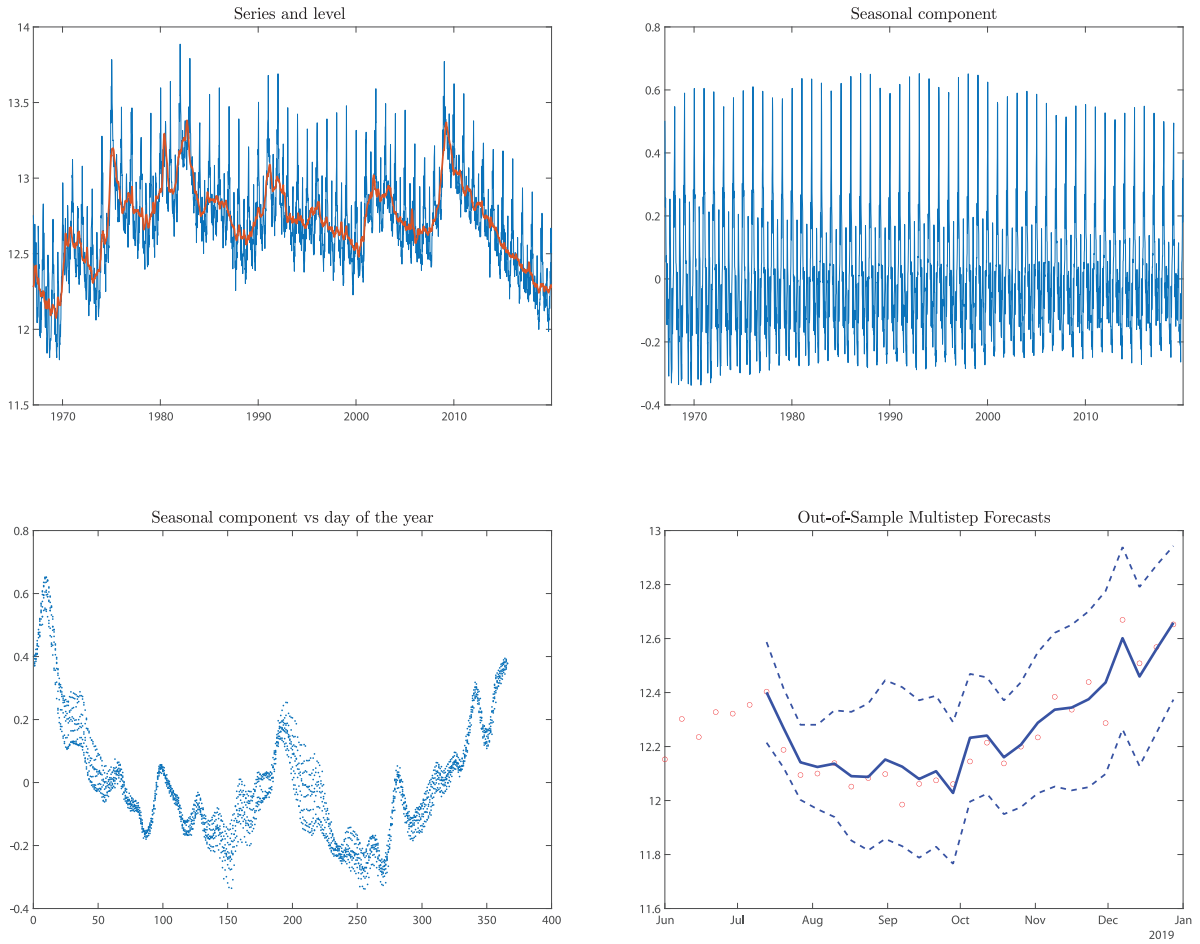

Fig. 7. US initial weekly unemployment claims. Series and smoothed estimates of the trend component (top left). Smoothed estimates of the seasonal component $\gamma_t^{(A)} + \gamma_t^{(M)}$ (top right). Estimated seasonal component $\gamma_t^{(A)} + \gamma_t^{(M)}$ versus the day of the year (bottom left). Series and out-of-sample interval multistep predictions for the last 25 weeks of 2019 (bottom right).

and the annual cycle, we also included among the explanatory variables the product between cosines and sines evaluated at the fundamental annual frequency with those evaluated at the monthly fundamental frequency. This eventually led to a satisfactory model.

Figure 7 displays the original time series (logarithms) and the smoothed estimates of the level in (B.9). The periodic component, plotted in the top right panel, is obtained by summing the annual and monthly periodic spline estimates and the regression component, which accounts for the deterministic additional cycles and the interaction effects. The bottom left panel displays the periodic component in annual time. The plot illustrates the variability of the monthly cycles within the year. Hence, we are led to conclude that it is not suitable to consider the seasonal component as the addition of the monthly and annual components. There are substantial dependencies between the two cycles, such that the monthly pattern of unemployment claims varies with the position in the annual cycle. Finally, the bottom right panel displays the multistep ahead point and the 95% interval forecast obtained for the last 25 weeks, when the selected model is estimated leaving out the corresponding observations.

8.4. Robust inference for NG daily volume

The daily volume of natural gas futures traded on the New York Mercantile Exchange series, NG series hereafter, was modelled after a variance stabilizing Box-Cox transformation with parameter 0.125. Similar results are obtained if the logarithmic transformation is used. The transformed series is plotted in the first panel of Fig. 8 along with the final robust estimated trend.

The series is characterised by a sizable and time varying monthly cycle with irregular period around 21 working days (during weekends NG is not traded), for modelling which we considered the trigonometric model (4)-(5), where $j = 1, 2, 3, 4,$ and in defining the time-varying frequencies s_t is the length of the month in which the observation falls

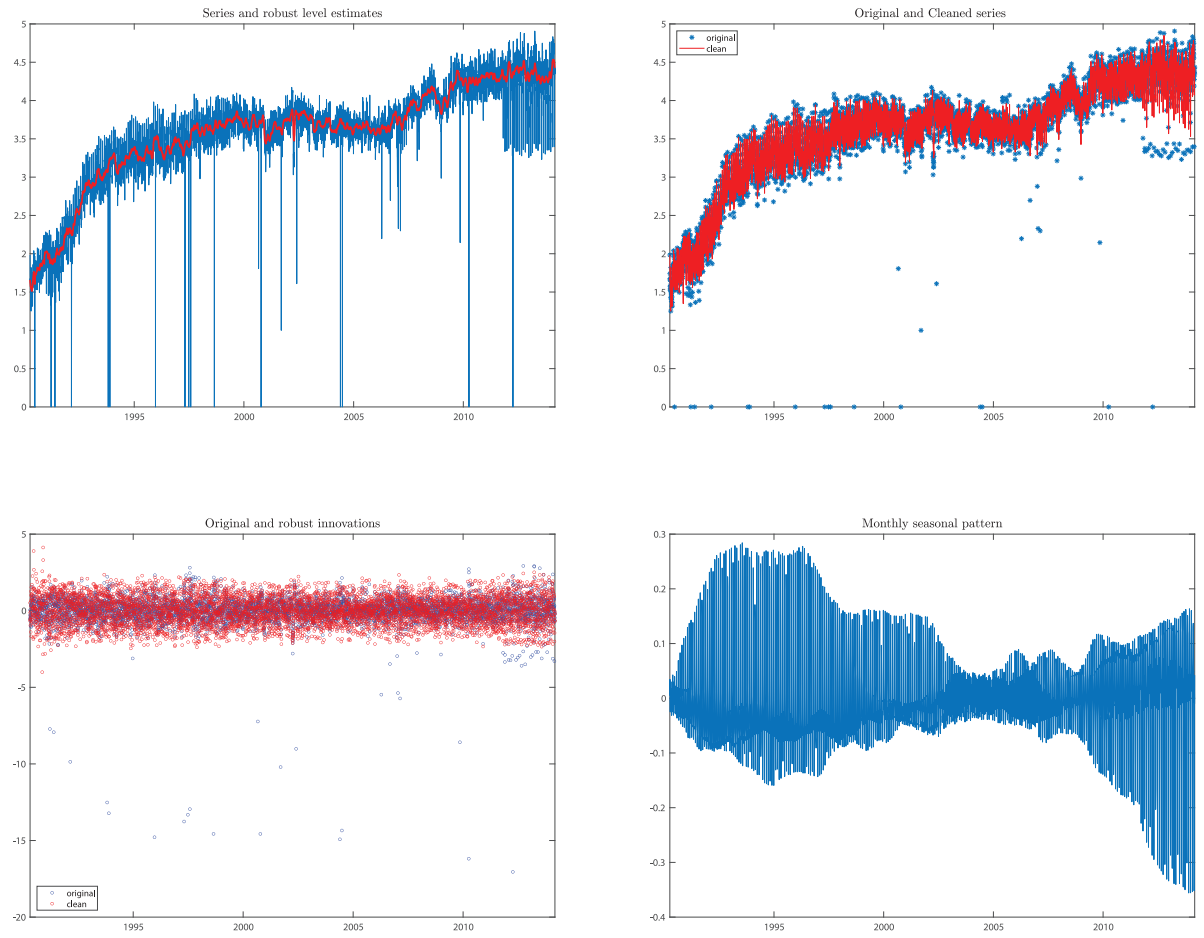

Fig. 8. Daily volume (Box-Cox transformation with parameter 1/8) of natural gas futures traded on the New York Mercantile Exchange during the period between April 3, 1990 and March 6, 2014. The top left panel shows the original series and the robust estimates of the underlying trend. The top right panel shows the cleaned observations resulting from the data cleaning filter of section 6.1. The bottom left panel displays the original and robust standardized innovations. The bottom right panel shows the robust smoothed estimates of the monthly seasonal cycle.

and d_t is its position within the month. The Easter and Christmas effects were captured by including wavelet interventions, as detailed in section 5.

The high level of outlier contamination affecting the series is clearly visible from its plot. Indeed the series provided a case study in robust filtering in Calvet et al. (2015). The robust estimation and filtering method of section 6.1 was applied by considering the value $c = 7.0414$ for the tuning parameter of Tukey's biweight influence function. Robust estimates of the parameters were obtained by two iterations of the procedure consisting of obtaining a robust scale estimate, run the robust augmented Kalman filter to obtain a clean series, and re-estimate the parameters.

9. Conclusions

We have considered three main challenges that need to be taken when modelling seasonality in high frequency time series, dealing respectively with modelling calendar events, such as holidays and moving festivals, achieving robustness in estimation and signal extraction, and the selection of the seasonal model. We think that there are very interesting research directions to be taken, relating to what modelling strategies should be adopted to move beyond the linear additive framework in the presence of substantial interactions between the annual cycle and its subannual components. Secondly, model selection with periodic spline models should consider choosing a periodic B-spline basis. We also plan to extend our methods to seasonal long memory models, of the type investigated by Asai et al. (2020) and Voges and Sibbertsen (2021). Finally, the role of nonlinearities and seasonal shifts along the lines of He et al. (2019) should be investigated.

Acknowledgements

Diego J. Pedregal acknowledges support from the European Regional Development Fund and Junta de Comunidades de Castilla-La Mancha (JCCM/FEDER, UE) under the project with reference SBPLY/19/180501/000151 and by the Vicerrectorado de Investigación y Política Científica from UCLM through the research group fund program (PREDILAB; [2021-GRIN-31210]).

Appendix A. The equivalence between time-domain and trigonometric seasonal models

The time-domain random effects model (7)-(8) encompasses the stochastic harmonic model of section 3. To show this, we shall focus on the canonical basis $\boldsymbol{w}_t = \mathbf{P}^{t-1'}\mathbf{e}_1$, for which $\bar{\boldsymbol{w}} = \mathbf{i}_s$ and consider the specification of $\boldsymbol{\Omega}$ as a circular covariance matrix.

Let $\boldsymbol{\omega}_t^*$ denote the s-dimensional circular white noise process, $\boldsymbol{\omega}_t^* \sim$ i.i.d. $\mathrm{N}(\mathbf{0}, \sigma_\omega^2 \mathbf{N}_s)$, where $\mathbf{N}_s$ is the circular correlation matrix:

$$\mathbf{N}_s = \begin{cases} \mathbf{I}_s + \sum_{k=1}^{s/2-1} \rho_k(\mathbf{P}^k + \mathbf{P}^{-k}) + \frac{1}{2}\rho_{s/2}(\mathbf{P}^{s/2} + \mathbf{P}^{-s/2}), & s \text{ even}, \\ \mathbf{I}_s + \sum_{k=1}^{\lfloor s/2 \rfloor} \rho_k(\mathbf{P}^k + \mathbf{P}^{-k}), & s \text{ odd}, \end{cases} \tag{A.1}$$

and ρ_k is the correlation coefficient at lag k. The spectral decomposition of $\mathbf{N}_s$, see theorem 6.5.3 in Anderson (2011), is $\mathbf{N}_s = \mathbf{V}_s \boldsymbol{\Lambda}_s \mathbf{V}_s'$, $\mathbf{V}_s'\mathbf{V}_s = \mathbf{V}_s\mathbf{V}_s' = \mathbf{I}_s$. When s is even,

$$\boldsymbol{\Lambda}_s = \mathrm{diag}\big(\varsigma_0, \varsigma_1, \varsigma_1, \ldots, \varsigma_{s/2-1}, \varsigma_{s/2-1}, \varsigma_{s/2}\big),$$

$$\mathbf{V}_s = \frac{1}{\sqrt{s}}[\mathbf{i}_s, \mathbf{c}_1, \mathbf{s}_1, \mathbf{c}_2, \mathbf{s}_2, \ldots, \mathbf{c}_{s/2-1}, \mathbf{s}_{s/2-1}, \mathbf{c}_{s/2}],$$

$$\mathbf{c}_k = \sqrt{2}\begin{bmatrix} \cos(\lambda_k) \\ \cos(2\lambda_k) \\ \vdots \\ \cos\left((s-1)\lambda_k\right) \\ \cos(s\lambda_k) \end{bmatrix}, \ \mathbf{s}_k = \sqrt{2}\begin{bmatrix} \sin(\lambda_k) \\ \sin(2\lambda_k) \\ \vdots \\ \sin\left((s-1)\lambda_k\right) \\ \sin(s\lambda_k) \end{bmatrix}, \ k = 1, \ldots, s/2-1,$$

and $\mathbf{c}_{s/2} = [-1, 1, \ldots, -1, 1]$, where $\lambda_k = \frac{2\pi k}{s}$, $k = 1, 2, \ldots, s/2$, are the seasonal frequencies, λ_1 being the fundamental frequency, corresponding to a period of s observations, and $\lambda_2, \ldots, \lambda_{s/2}$ are the harmonic frequencies.

The eigenvalues of $\mathbf{N}_s$ are

$$\varsigma_k = 1 + 2\sum_{j=1}^{s/2-1} \rho_j \cos(\lambda_k j) + \rho_{s/2}\cos(\lambda_k s/2),$$

$k = 0, 1, \ldots, s/2$, where we have set $\lambda_0 = 0$. Hence, we can write

$$\mathbf{N}_s = \frac{\varsigma_0}{s}\mathbf{i}_s\mathbf{i}_s' + \sum_{k=1}^{s/2-1} \frac{\varsigma_k}{s}(\mathbf{c}_k\mathbf{c}_k' + \mathbf{s}_k\mathbf{s}_k') + \frac{\varsigma_{s/2}}{s}\mathbf{c}_{s/2}\mathbf{c}_{s/2}'. \tag{A.2}$$

When s is odd,

$$\mathbf{N}_s = \frac{\varsigma_0}{s}\mathbf{i}_s\mathbf{i}_s' + \sum_{k=1}^{(s-1)/2} \frac{\varsigma_k}{s}(\mathbf{c}_k\mathbf{c}_k' + \mathbf{s}_k\mathbf{s}_k').$$

The detrended matrix $\boldsymbol{\Omega} = \sigma_\omega^2(\mathbf{N}_s - \frac{\varsigma_0}{s}\mathbf{i}_s\mathbf{i}_s')$ provides a model for the disturbances of the model (7)-(8). In particular, if all the circular correlations ρ_j are zero, $\mathbf{N}_s = \mathbf{I}_s$, and $\varsigma_0 = 1$, which is the Harrison and Stevens (1976) seasonal model.

When the circular correlations are nonzero we have the equivalent trigonometric seasonal representation

$$\gamma_t = \begin{cases} \sum_{j=1}^{s/2-1}\left(\alpha_{kt}\cos(\lambda_k t) + \alpha_{kt}^*\sin(\lambda_k t)\right) + \alpha_{\frac{s}{2},t}(-1)^t, & s \text{ even}, \\ \sum_{k=1}^{(s-1)/2}\left(\alpha_{kt}\cos(\lambda_k t) + \alpha_{kt}^*\sin(\lambda_k t)\right), & s \text{ odd}, \end{cases} \tag{A.3}$$

where α_{kt} and α_{kt}^* evolve as independent random walks with disturbance variance proportional to ς_k.

Considering for simplicity the case when s is even and writing

$$\mathbf{V}_s^* = \frac{1}{\sqrt{s}}\big[\mathbf{c}_1, \mathbf{s}_1, \mathbf{c}_2, \mathbf{s}_2, \ldots, \mathbf{c}_{s/2-1}, \mathbf{s}_{s/2-1}, \mathbf{c}_{s/2}\big],$$

gives the trigonometric seasonal model

$$\gamma_t = \boldsymbol{w}_t'\mathbf{V}_s\mathbf{V}_s'\boldsymbol{\delta}_t,$$

$$
\begin{aligned}
&= \boldsymbol{w}_t' \mathbf{V}_s^* \mathbf{V}_s^{*'} \boldsymbol{\delta}_t, \\
&= \left[\boldsymbol{w}_t'\mathbf{c}_1, \boldsymbol{w}_t'\mathbf{s}_1, \boldsymbol{w}_t'\mathbf{c}_2, \boldsymbol{w}_t'\mathbf{s}_2, \ldots, \boldsymbol{w}_t'\mathbf{c}_{s/2}\right]\boldsymbol{\delta}_t^\dagger, \\
&= [\cos(\lambda_1 t), \sin(\lambda_1 t), \cos(\lambda_2 t), \sin(\lambda_2 t), \ldots, \cos(\pi t)]\boldsymbol{\delta}_t^\dagger,
\end{aligned}
$$

with

$$
\boldsymbol{\delta}_{t+1}^\dagger \quad = \quad \boldsymbol{\delta}_t^\dagger + \boldsymbol{\omega}_t^\dagger, \quad \boldsymbol{\omega}_t^\dagger \sim \text{i.i.d. N}\big(\mathbf{0}, \text{diag}(\sigma_1^2, \sigma_1^2, \sigma_2^2, \sigma_2^2, \ldots, \sigma_{s/2}^2)\big),
$$

where $\sigma_k^2 = \sigma_\omega^2 \varsigma_k / s$.

The first line follows from the orthogonality of the matrix $\mathbf{V}_s$, $\mathbf{V}_s\mathbf{V}_s' = \mathbf{I}_s$, the second from the decomposition $\mathbf{V}_s\mathbf{V}_s' = \mathbf{V}_s^*\mathbf{V}_s^{*'} + \frac{\mathbf{i}_s\mathbf{i}_s'}{s}$ and from $\mathbf{i}_s'\boldsymbol{\delta}_t = 0$, for all t. In the third line $\boldsymbol{\delta}_t^* = \mathbf{V}_s^{*'}\boldsymbol{\delta}$ has alternating elements $\alpha_{kt} = \mathbf{c}_k'\boldsymbol{\delta}_t$ and $\alpha_{kt}^* = \mathbf{s}_k'\boldsymbol{\delta}_t$.

Appendix B. State Space Methods

B1. The Augmented Kalman Filter

Consider the state space model (9), with initial conditions stated in (10). Setting $\mathbf{A}_{1|0} = -\mathbf{W}_0$, $\mathbf{P}_{1|0}^* = \mathbf{H}_0\mathbf{H}_0'$, the augmented Kalman filter (de Jong, 1991, AKF henceforth) consists of the following set of recursions and definitions. For $t = 1, \ldots, n$:

$$
\begin{aligned}
v_t^* &= y_t - \mathbf{z}_t'\tilde{\boldsymbol{\alpha}}_{t|t-1}^*, \quad \mathbf{v}_t' = \mathbf{x}_t' - \mathbf{z}_t'\mathbf{A}_{t|t-1}, \\
f_t^* &= \mathbf{z}_t'\mathbf{P}_{t|t-1}^*\mathbf{z}_t' + \mathbf{g}_t'\mathbf{g}_t, \quad \mathbf{k}_t^* = \mathbf{T}_t\mathbf{P}_{t|t-1}^*\mathbf{z}_t f_t^{*-1}, \\
\tilde{\boldsymbol{\alpha}}_{t+1|t}^* &= \mathbf{T}_t\tilde{\boldsymbol{\alpha}}_{t|t-1}^* + \mathbf{k}_t^* v_t^*, \quad \mathbf{A}_{t+1|t} = \mathbf{T}_t\mathbf{A}_{t|t-1} - \mathbf{W}_t + \mathbf{k}_t^*\mathbf{v}_t', \\
\mathbf{P}_{t+1|t}^* &= \mathbf{T}_t\mathbf{P}_{t|t-1}^*\mathbf{T}_t' + \mathbf{H}_t\mathbf{H}_t' - f_t^*\mathbf{k}_t^*\mathbf{k}_t^{*'}.
\end{aligned} \tag{B.1}
$$

The starred quantities correspond to the usual Kalman filter applied to y_t with $\boldsymbol{\beta} = \mathbf{0}$, when the state vector is initialized by $\tilde{\boldsymbol{\alpha}}_{1|0}^*$ and no explanatory variables are considered. The scalar v_t^* represents the conditional innovations, $v_t^* = y_t - \text{E}(y_t|\mathcal{Y}_{t-1}, \boldsymbol{\beta} = \mathbf{0})$, whereas f_t^* denotes its variance, $f_t^* = \sigma^{-2}\text{Var}(y_t|\mathcal{Y}_{t-1}, \boldsymbol{\beta} = \mathbf{0})$. Here, $\mathcal{Y}_{t-1} = \{y_1, y_2, \ldots, y_{t-1}\}$ is the past history of y_t. The Kalman gain matrix has the following interpretation: $\mathbf{k}_t^* = \text{Cov}(\boldsymbol{\alpha}_t, y_t|\mathcal{Y}_{t-1}, \boldsymbol{\beta} = \mathbf{0})[\text{Var}(y_t|\mathcal{Y}_{t-1}, \boldsymbol{\beta} = \mathbf{0})]^{-1}$. The vector $\tilde{\boldsymbol{\alpha}}_{t+1|t}^* = \text{E}(\boldsymbol{\alpha}_{t+1}|\mathcal{Y}_t, \boldsymbol{\beta} = \mathbf{0})$ is the one–step–ahead prediction of the state vector given the information in period t and conditional on $\boldsymbol{\beta} = \mathbf{0}$. The corresponding conditional covariance matrix is $\mathbf{P}_{t+1|t}^*$.

The time series innovations are $y_t - \text{E}(y_t|\mathcal{Y}_{t-1}, \mathcal{X}_t, \mathcal{W}_{t-1}) = v_t^* - \mathbf{v}_t'\boldsymbol{\beta}$ and the one step ahead prediction of the state vector is $\text{E}(\boldsymbol{\alpha}_{t+1}|\mathcal{Y}_t, \mathcal{X}_t, \mathcal{W}_t) = \tilde{\boldsymbol{\alpha}}_{t+1|t}^* - \mathbf{A}_{t+1|t}\boldsymbol{\beta}$. Here we denoted $\mathcal{X}_t = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t\}$ and $\mathcal{W}_t = \{\mathbf{W}_0, \mathbf{W}_1, \ldots, \mathbf{W}_t\}$.

Given the availability of a sample of $t > K$ observations, $\{\mathcal{Y}_t, \mathcal{X}_t, \mathcal{W}_t\}$, the vector $\boldsymbol{\beta}$ can be estimated by a weighted least squares regression of v_i^* on $\mathbf{v}_i$, by minimizing the criterion function

$$
S(\boldsymbol{\beta}) = \sum_{i=1}^{t} \frac{(v_i^* - \mathbf{v}_i'\boldsymbol{\beta})^2}{f_i^*}.
$$

Defining

$$
\mathbf{s}_t = \sum_{i=1}^{t} \frac{\mathbf{v}_i v_i^*}{f_i^*}, \qquad \mathbf{S}_t = \sum_{i=1}^{t} \frac{\mathbf{v}_i\mathbf{v}_i'}{f_i^*},
$$

for $t \geq K$ we obtain $\tilde{\boldsymbol{\beta}}_t = \mathbf{S}_t^{-1}\mathbf{s}_t$. The variance-covariance matrix of the regression coefficients (scaled by σ^{-2}) is $\mathbf{B}_t = \mathbf{S}_t^{-1}$. Also, the estimator of σ^2 based on t observations is

$$
\hat{\sigma}_t^2 = \frac{1}{t}S(\tilde{\boldsymbol{\beta}}_t) = \frac{1}{t}\left[\sum_{i=1}^{t} \frac{v_i^{*2}}{f_i^*} - \mathbf{s}_t'\mathbf{S}_t^{-1}\mathbf{s}_t\right].
$$

Replacing $\boldsymbol{\beta}$ by its estimator, we obtain the innovations $v_t = y_t - \text{E}(y_t|\mathcal{Y}_{t-1}, \mathcal{X}_t, \mathcal{W}_{t-1})$, the one–step–ahead prediction of the state vector, $\tilde{\boldsymbol{\alpha}}_{t|t-1} = \text{E}(\boldsymbol{\alpha}_t|\mathcal{Y}_{t-1}, \mathcal{X}_{t-1}, \mathcal{W}_{t-1})$, and the corresponding estimation error covariance matrices, as follows:

$$
\begin{aligned}
v_t &= v_t^* - \mathbf{v}_t'\tilde{\boldsymbol{\beta}}_{t-1}, \qquad f_t = f_t^* + \mathbf{v}_t'\mathbf{S}_{t-1}^{-1}\mathbf{v}_t, \\
\tilde{\boldsymbol{\alpha}}_{t|t-1} &= \tilde{\boldsymbol{\alpha}}_{t|t-1}^* - \mathbf{A}_{t|t-1}\tilde{\boldsymbol{\beta}}_{t-1}, \qquad \mathbf{P}_{t|t-1} = \mathbf{P}_{t|t-1}^* + \mathbf{A}_{t|t-1}\mathbf{S}_{t-1}^{-1}\mathbf{A}_{t|t-1}'.
\end{aligned} \tag{B.2}
$$

B2. Estimation of model hyperparameters

The maximum likelihood estimators of $\boldsymbol{\theta}$ are the maximizers of the profile marginal likelihood function

$$L(\boldsymbol{\theta}) = -\frac{1}{2}\left[(n-K)(\ln\hat{\sigma}^2+1) + \sum_{t=1}^{n}\ln|f_t^*| + \ln|\mathbf{S}_n|\right], \tag{B.3}$$

where $\hat{\sigma}^2$ is the estimator of the scale parameter σ^2:

$$\hat{\sigma}^2 = \frac{1}{n-K}\left[\sum_{t=1}^{n}\frac{v_t^{*2}}{f_t^*} - \mathbf{s}_n'\mathbf{S}_n^{-1}\mathbf{s}_n\right]. \tag{B.4}$$

The likelihood (B.3) is based on reduced rank linear transformation of the series that eliminates dependence on $\boldsymbol{\beta}$; see de Jong (1991) and Francke et al. (2010).

B3. Real time filtering, prediction and smoothing

At time t, the updated estimate of the vector $\boldsymbol{\beta}$ is $\tilde{\boldsymbol{\beta}}_t = \mathbf{S}_t^{-1}\mathbf{s}_t$. Recalling that $\mathbf{S}_t = \mathbf{S}_{t-1} + f_t^{*-1}\mathbf{v}_t\mathbf{v}_t'$, and applying the Sherman–Woodbury–Morrison matrix inversion lemma Henderson and Searle (1981),

$$\mathbf{S}_t^{-1} = \mathbf{S}_{t-1}^{-1} - \frac{\mathbf{S}_{t-1}^{-1}\mathbf{v}_t\mathbf{v}_t'\mathbf{S}_{t-1}^{-1}}{f_t}.$$

Also, $\mathbf{s}_t = \mathbf{s}_{t-1} + f_t^{*-1}\mathbf{v}_t v_t^*$ and (B.2) yield the recursive formula:

$$\tilde{\boldsymbol{\beta}}_t = \tilde{\boldsymbol{\beta}}_{t-1} + \mathbf{S}_{t-1}^{-1}\mathbf{v}_t\frac{v_t}{f_t}. \tag{B.5}$$

The real-time estimates of the state vector, $\tilde{\boldsymbol{\alpha}}_{t|t} = \mathrm{E}(\boldsymbol{\alpha}_t|\mathcal{Y}_t,\mathcal{X}_t,\mathcal{W}_t)$, and its covariance matrix $\mathrm{Var}(\boldsymbol{\alpha}_t|\mathcal{Y}_t,\mathcal{X}_t,\mathcal{W}_t) = \sigma^2\mathbf{P}_{t|t}$, are obtained by running the filter

$$\begin{aligned}
&\tilde{\boldsymbol{\alpha}}_{t|t}^* = \tilde{\boldsymbol{\alpha}}_{t|t-1}^* + \mathbf{P}_{t|t-1}^*\mathbf{z}_t\tfrac{v_t^*}{f_t^*}, && \mathbf{A}_{t|t} = \mathbf{A}_{t|t-1} + \mathbf{P}_{t|t-1}^*\mathbf{z}_t\tfrac{\mathbf{v}_t'}{f_t^*},\\
&\mathbf{P}_{t|t}^* = \mathbf{P}_{t|t-1}^* - \mathbf{P}_{t|t-1}^*\tfrac{\mathbf{z}_t\mathbf{z}_t'}{f_t^*}\mathbf{P}_{t|t-1}^*,
\end{aligned} \tag{B.6}$$

and setting

$$\tilde{\boldsymbol{\alpha}}_{t|t} = \tilde{\boldsymbol{\alpha}}_{t|t}^* - \mathbf{A}_{t|t}\tilde{\boldsymbol{\beta}}_t, \quad \mathbf{P}_{t|t} = \mathbf{P}_{t|t}^* + \mathbf{A}_{t|t}\mathbf{S}_t^{-1}\mathbf{A}_{t|t}'. \tag{B.7}$$

Then, the prediction step for the state vector gives:

$$\begin{aligned}
&\tilde{\boldsymbol{\alpha}}_{t+1|t}^* = \mathbf{T}_t\tilde{\boldsymbol{\alpha}}_{t|t}^*, && \mathbf{A}_{t+1|t} = \mathbf{T}_t\mathbf{A}_{t|t} - \mathbf{W}_t,\\
&\mathbf{P}_{t+1|t}^* = \mathbf{T}_t\mathbf{P}_{t|t}^*\mathbf{T}_t' + \mathbf{H}_t\mathbf{H}_t'.
\end{aligned} \tag{B.8}$$

Missing values are handled by skipping the KF updating operations: if y_t is missing at time t, the recursions (B.1) are replaced by

$$\begin{aligned}
&\tilde{\boldsymbol{\alpha}}_{t+1|t-1}^* = \mathbf{T}_t\tilde{\boldsymbol{\alpha}}_{t|t-1}^*, && \mathbf{A}_{t+1|t-1} = \mathbf{T}_t\mathbf{A}_{t|t-1} - \mathbf{W}_t,\\
&\mathbf{P}_{t+1|t-1}^* = \mathbf{T}_t\mathbf{P}_{t|t-1}^*\mathbf{T}_t' + \mathbf{H}_t\mathbf{H}_t'.
\end{aligned}$$

The smoothed estimates $\tilde{\boldsymbol{\alpha}}_{t|n} = \mathrm{E}(\boldsymbol{\alpha}_t|\mathcal{Y}_n,\mathcal{X}_n,\mathcal{W}_n)$, and their covariance matrix $\mathbf{P}_{t|n} = \mathrm{E}[(\boldsymbol{\alpha}_t - \tilde{\boldsymbol{\alpha}}_{t|n})(\boldsymbol{\alpha}_t - \tilde{\boldsymbol{\alpha}}_{t|n})'|\mathcal{Y}_n,\mathcal{X}_n,\mathcal{W}_n]$, are computed by the following backwards recursive formulae, given by de Jong (1989), starting at $t = n$, with initial values $\mathbf{r}_n^* = 0, \mathbf{R}_n = \mathbf{0}$ and $\mathbf{N}_n^* = 0$: for $t = n-1,\ldots,1$,

$$\begin{aligned}
\mathbf{r}_{t-1}^* &= \mathbf{L}_t'\mathbf{r}_t^* + \mathbf{z}_t f_t^{-1}v_t^*, \qquad \mathbf{R}_{t-1} = \mathbf{L}_t'\mathbf{R}_t + \mathbf{z}_t f_t^{-1*}\mathbf{v}_t', \quad \mathbf{L}_t = \mathbf{T}_t - \mathbf{k}_t\mathbf{z}',\\
\mathbf{N}_{t-1}^* &= \mathbf{L}_t'\mathbf{N}_t^*\mathbf{L}_t + \mathbf{z}_t f_t^{-1*}\mathbf{z}_t',\\
\tilde{\boldsymbol{\alpha}}_{t|n} &= \tilde{\boldsymbol{\alpha}}_{t|t-1}^* - \mathbf{A}_{t|t-1}\tilde{\boldsymbol{\beta}}_n + \mathbf{P}_{t|t-1}^*\left(\mathbf{r}_{t-1}^* - \mathbf{R}_{t-1}\tilde{\boldsymbol{\beta}}_n\right),\\
\mathbf{P}_{t|n} &= \mathbf{P}_{t|t-1}^* - \mathbf{P}_{t|t-1}^*\mathbf{N}_{t-1}^*\mathbf{P}_{t|t-1}^* + (\mathbf{A}_{t|t-1} + \mathbf{P}_{t|t-1}^*\mathbf{R}_{t-1})\mathbf{S}_n^{-1}(\mathbf{A}_{t|t-1} + \mathbf{P}_{t|t-1}^*\mathbf{R}_{t-1})'.
\end{aligned} \tag{B.9}$$

B4. Robust Augmented Kalman Filter

The robust AKF is thus obtained by replacing f_t^{-1} by $\mathrm{w}_t f_t^{-1}$ in the updating equations for the state and the regression effects. In particular, after running (B.1) and (B.2), $t = k+1, \ldots, n$, we compute

$$\begin{aligned} \tilde{\boldsymbol{\beta}}_t &= \tilde{\boldsymbol{\beta}}_{t-1} + \mathrm{w}_t \mathbf{S}_{t-1}^{-1} \mathbf{v}_t' \tfrac{v_t}{f_t}, \\ \mathbf{S}_t^{-1} &= \mathbf{S}_{t-1}^{-1} - \mathrm{w}_t \mathbf{S}_{t-1}^{-1} \tfrac{\mathbf{v}_t \mathbf{v}_t'}{f_t} \mathbf{S}_{t-1}^{-1}. \end{aligned} \tag{B.10}$$

Note that $\mathrm{w}_t = 1$ yields the usual updated estimate, whereas if $\mathrm{w}_t = 0$, $\tilde{\boldsymbol{\beta}}_t = \tilde{\boldsymbol{\beta}}_{t-1}$ and $\mathbf{S}_t^{-1} = \mathbf{S}_{t-1}^{-1}$, so that the updating of the inferences on $\boldsymbol{\beta}$ does not occur.

The robustified real-time estimates of the state vector are:

$$\begin{aligned} \tilde{\boldsymbol{\alpha}}_{t|t} &= \tilde{\boldsymbol{\alpha}}_{t|t-1}^* - \mathbf{A}_{t|t-1} \tilde{\boldsymbol{\beta}}_t + \mathrm{w}_t \mathbf{P}_{t|t-1}^* \mathbf{z}_t \tfrac{v_t}{f_t}, \\ \mathbf{P}_{t|t} &= \mathbf{P}_{t|t-1}^* - \mathrm{w}_t \mathbf{P}_{t|t-1}^* \tfrac{\mathbf{z}_t \mathbf{z}_t'}{f_t^*} \mathbf{P}_{t|t-1}^* + \mathbf{A}_{t|t} \mathbf{S}_t^{-1} \mathbf{A}_{t|t}'. \end{aligned} \tag{B.11}$$

The robustified real-time estimate of the disturbance vector $\boldsymbol{\varepsilon}_t$ is:

$$\tilde{\boldsymbol{\varepsilon}}_{t|t} = \mathbf{g}_t \mathrm{w}_t \frac{v_t}{f_t}. \tag{B.12}$$

If $\mathrm{w}_t = 0$, it can be easily seen from (B.11) that $\tilde{\boldsymbol{\alpha}}_{t|t} = \tilde{\boldsymbol{\alpha}}_{t|t-1}$. Moreover, $\tilde{\boldsymbol{\alpha}}_{t+1|t}$ is a two-step-prediction in this case, which becomes evident from writing $\tilde{\boldsymbol{\alpha}}_{t+1|t} = \mathbf{T}_t \tilde{\boldsymbol{\alpha}}_{t|t-1} + \mathbf{W}_t \tilde{\boldsymbol{\beta}}_t + \mathbf{H}_t \tilde{\boldsymbol{\varepsilon}}_{t|t}$, and replacing $\tilde{\boldsymbol{\beta}}_t$ by (B.10), and $\tilde{\boldsymbol{\varepsilon}}_{t|t}$ by (B.12).

The one-step-ahead prediction equations are corrected as follows:

$$\begin{aligned} \tilde{\boldsymbol{\alpha}}_{t+1|t}^* &= \mathbf{T} \tilde{\boldsymbol{\alpha}}_{t|t-1}^* + \mathrm{w}_t \mathbf{k}_t v_t^*, & \mathbf{A}_{t+1|t} &= \mathbf{T} \mathbf{A}_{t|t-1} + \mathrm{w}_t \mathbf{k}_t \mathbf{v}_t', \\ \mathbf{P}_{t+1|t}^* &= \mathbf{T} \mathbf{P}_{t|t-1}^* \mathbf{T}' + \mathbf{H}\mathbf{H}' - \mathrm{w}_t f_t^* \mathbf{k}_t \mathbf{k}_t'. \end{aligned} \tag{B.13}$$

References

Anderson, T.W., 2011. The statistical analysis of time series, 19. John Wiley & Sons.

Asai, M., McAleer, M., Peiris, S., 2020. Realized stochastic volatility models with generalized gegenbauer long memory. Econometrics and Statistics 16, 42–54.

Aston, J., Koopman, S.J., et al., 2006. A non-gaussian generalization of the airline model for robust seasonal adjustment. Journal of Forecasting 25 (5), 325.

Bruce, A.G., Jurke, S.R., 1996. Non-gaussian seasonal adjustment: X-12-arima versus robust structural models. Journal of Forecasting 15 (4), 305–328.

Busetti, F., Harvey, A.C., 2003. Seasonality tests. Journal of Business & Economic Statistics 21 (3), 420–436.

Calvet, L.E., Czellar, V., Ronchetti, E., 2015. Robust filtering. Journal of the American Statistical Association 110 (512), 1591–1606.

Canova, F., Hansen, B.E., 1995. Are seasonal patterns constant over time? a test for seasonal stability. Journal of Business & Economic Statistics 13 (3), 237–252.

Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I., 1990. STL: A seasonal-trend decomposition procedure based on LOESS. Journal of Official Statistics 6 (1), 3–73.

De Boor, C., 1978. A practical guide to splines, 27. springer-verlag New York.

de Jong, P., 1989. Smoothing and Interpolation with the State-Space Model. Journal of the American Statistical Association 84 (408), 1085–1088.

de Jong, P., 1991. The Diffuse Kalman Filter. The Annals of Statistics 19, 1073–1083.

De Livera, A.M., Hyndman, R.J., Snyder, R.D., 2011. Forecasting time series with complex seasonal patterns using exponential smoothing. Journal of the American Statistical Association 106 (496), 1513–1527.

Durbin, J., Koopman, S.J., 2012. Time Series Analysis by State Space Methods, second Oxford University Press, Oxford.

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

Hannan, E.J., Terrell, R., Tuckwell, N., 1970. The seasonal adjustment of economic time series. International Economic Review 11 (1), 24–52.

Harrison, P.J., Stevens, C.F., 1976. Bayesian forecasting. Journal of the Royal Statistical Society. Series B (Methodological) 205–247.

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

Harvey, A.C., 2001. Testing in unobserved components models. Journal of Forecasting 20 (1), 1–19.

Harvey, A.C., 2013. Dynamic Models for Volatility and Heavy Tails: with Applications to Financial and Economic Time Series. Cambridge University Press.

Harvey, A.C., Koopman, S.J., 1993. Forecasting hourly electricity demand using time-varying splines. Journal of the American Statistical Association 88 (424), 1228–1236.

Harvey, A.C., Koopman, S.J., Riani, M., 1997. The modeling and seasonal adjustment of weekly observations. Journal of Business & Economic Statistics 15 (3), 354–368.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The elements of statistical learning, 1. Springer series in statistics New York.

He, C., Kang, J., Teräsvirta, T., Zhang, S., 2019. The shifting seasonal mean autoregressive model and seasonality in the central england monthly temperature series, 1772–2016. Econometrics and Statistics 12, 1–24.

Henderson, H.V., Searle, S.R., 1981. On Deriving the Inverse of a Sum of Matrices. Siam Review 23 (1), 53–60.

Hendry, D.F., Doornik, J.A., 2014. Empirical model discovery and theory evaluation: automatic selection methods in econometrics. MIT Press.

Ladiray, D., Palate, J., Mazzi, G., Proietti, T., 2018. Seasonal Adjustment of Daily and Weekly Data. In: Mazzi, G., Ladiray, D., Rieser, D. (Eds.), Handbook on Seasonal Adjustment. Eurostat: Publications Office of the European Union, Luxembourg., pp. 759–782.

Leschinski, C., Sibbertsen, P., 2019. Model order selection in periodic long memory models. Econometrics and statistics 9, 78–94.

Marczak, M., Proietti, T., 2016. Outlier detection in structural time series models: The indicator saturation approach. International Journal of Forecasting 32 (1), 180–202.

Maronna, R.A., Martin, R.D., Yohai, V.J., Salibián-Barrera, M., 2018. Robust statistics: theory and methods (with R). John Wiley & Sons.

Martin, R.D., 1979. Robust Estimation for Time Series Autoregressions. In: Launer, R.L., Wilkinson, G.N. (Eds.), Robustness in Statistics. Academic Press, New York, pp. 147–176.

Martin, R.D., Thomson, D.J., 1982. Robust–resistant Spectrum Estimation. Proceedings of the IEEE 70, 1097–1115.

Masreliez, C.J., Martin, R.D., 1977. Robust Bayesian Estimation for the Linear Model and Robustifying the Kalman Filter. IEEE Transactions on Automatic Control AC–22, 361–371.

Pedregal, D.J., Young, P.C., 2006. Modulated cycles, an approach to modelling periodic components from rapidly sampled data. International Journal of Forecasting 22 (1), 181–194.

Penzer, J., 2007. State space models for time series with patches of unusual observations. Journal of Time Series Analysis 28 (5), 629–645.
Pierce, D.A., Grupe, M.R., Cleveland, W.P., 1984. Seasonal adjustment of the weekly monetary aggregates: A model-based approach. Journal of Business & Economic Statistics 2 (3), 260–270.
Poirier, D.J., 1976. econometrics of structural change, with special emphasis on spline functions. North-Holland Pub. Co..
Proietti, T., 2000. Comparing seasonal components for structural time series models. International Journal of Forecasting 16 (2), 247–260.
Rosenberg, B., 1973. Random Coefficients Models: the Analysis of a Cross Section of Time Series by Stochastically Convergent Parameter Regression. In: Annals of Economic and Social Measurement, 2. NBER, pp. 399–428.
Santos, C., Hendry, D.F., Johansen, S., 2008. Automatic selection of indicators in a fully saturated regression. Computational Statistics 23 (2), 317–335.
Voges, M., Sibbertsen, P., 2021. Cyclical fractional cointegration. Econometrics and Statistics 19, 114–129.
West, M., Harrison, J., 1997. Bayesian Forecasting and Dynamic Models, 2nd Springer–Verlag, New York.