



**HAL**  
open science

## Vers une désaisonnalisation des séries temporelles infra-mensuelles avec JDemetra+.

Smyk Anna, Webel Karsten

### ► To cite this version:

Smyk Anna, Webel Karsten. Vers une désaisonnalisation des séries temporelles infra-mensuelles avec JDemetra+.. 2024. <hal-05069646>

**HAL Id: hal-05069646**

**<https://hal.science/hal-05069646v1>**

Preprint submitted on 15 May 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Vers une désaisonnalisation des séries temporelles infra-mensuelles avec JDemetra+

---

Document de travail

N° M2024-04 – Octobre 2024





**M 2024/04**

**Vers une désaisonnalisation des séries temporelles  
infra-mensuelles avec JDemetra+**

**Anna SMYK**  
Insee

**Karsten WEBEL**  
Deutsche Bundesbank

**OCTOBRE 2024**

*Note : Ce document de travail est la traduction en français d'un travail originellement paru en septembre 2023 sous la forme d'un Discussion Paper de la Deutsche Bundesbank (No 24/2023). Seules les références aux packages R et les extraits de code ont été actualisés.*

*Remerciements :*

*Les auteurs remercient Christiane Hofer, James Livsey et Jean Palate pour leurs précieux commentaires*



# Résumé non technique

## Problématique

Ces dernières années, les statistiques officielles ont accordé de plus en plus d'attention aux séries temporelles infra-mensuelles, principalement en raison de la transformation numérique de la collecte de données et surtout en réaction à la demande croissante d'indicateurs économiques précoces à la suite de l'épidémie de COVID-19 en 2020. Cependant, les séries temporelles infra-mensuelles présentent souvent des dynamiques saisonnières complexes et d'autres particularités que l'on ne retrouve généralement pas dans les données compilées à un rythme mensuel ou trimestriel. Par conséquent, les méthodes traditionnelles de modélisation et de désaisonnalisation des séries chronologiques ne peuvent être appliquées sans modifications appropriées.

## Contribution

Nous donnons une description méthodologique approfondie des modifications apportées aux trois méthodes traditionnelles mises en œuvre dans JDemetra+, le logiciel officiel de désaisonnalisation utilisé pour produire des statistiques harmonisées dans le cadre du Système statistique européen et du Système européen de banques centrales. Nous discutons également des principales propriétés statistiques de ces modifications d'un point de vue théorique et examinons trois exemples afin de mettre en évidence leurs principales caractéristiques.

## Résultats

Nos illustrations suggèrent que les modifications apportées aux approches traditionnelles de modélisation et de désaisonnalisation fournissent des résultats convaincants pour divers types de séries temporelles infra-mensuelles et, par conséquent, des alternatives crédibles à des solutions *ad-hoc* qui ne sont actuellement pas intégrées dans les logiciels utilisés par la statistique publique. Toutefois, des recherches supplémentaires sont nécessaires pour parvenir à une acceptation définitive de ces modifications parmi les praticiens. Nous partageons enfin quelques réflexions sur leurs voies de développement.

# Vers une désaisonnalisation des séries temporelles infra-mensuelles avec JDemetra+. \*

Karsten Webel  
Deutsche Bundesbank

Anna Smyk  
Insee

## Résumé

Les séries économiques infra-mensuelles sont devenues de plus en plus populaires dans les statistiques officielles ces dernières années. Cette évolution a été largement favorisée par la transformation numérique de la dernière décennie. La pandémie de COVID-19 en 2020 a renforcé ce phénomène, car de nombreux utilisateurs de données ont immédiatement demandé des données hebdomadaires, voire quotidiennes, sur l'évolution de l'économie. Ces données infra-mensuelles présentent souvent un comportement saisonnier qui nécessite un ajustement. C'est pourquoi JDemetra+, le logiciel officiel de désaisonnalisation des données mensuelles et trimestrielles du Système statistique européen et du Système européen de banques centrales, a été récemment enrichi d'un modèle de pré-ajustement de type regArima et de versions étendues des algorithmes de décomposition basées sur les modèles Arima, STL et X-11, adaptées aux spécificités des données infra-mensuelles. Celles-ci sont accessibles par le biais d'un écosystème de packages R qui permet également d'accéder à une modélisation structurelle des séries temporelles, dans un cadre espace-état. Nous donnons un aperçu complet de ces packages et en illustrons les principales caractéristiques. Nous fournissons des extraits de code R utilisés pour désaisonnaliser les naissances quotidiennes en France, la consommation horaire d'électricité en Allemagne et les demandes initiales hebdomadaires d'assurance chômage aux États-Unis.

**Mots clés :** modèle Airline étendu, données haute-fréquence, statistique publique, extraction de signaux, décomposition en composantes inobservables

**Classification JEL :** C01, C02, C14, C18, C22, C40, C50.

---

\*Contact : Anna Smyk, Insee, Département des Méthodes Statistiques [anna.smyk@insee.fr](mailto:anna.smyk@insee.fr)

# 1 Introduction

JDemetra+ (JD+) est le logiciel recommandé pour la production harmonisée des séries désaisonnalisées dans le Système statistique européen et le Système européen des banques centrales. Il met en œuvre les algorithmes de désaisonnalisation les plus utilisés, X13-Arima et Tramo-Seats, qui ont des modules de pré-ajustement regArima très proches. Une version 3.0 de JD+, actuellement en cours de développement, a été principalement créée pour permettre l’ajout de méthodes de désaisonnalisation de séries infra-mensuelles. Ces séries temporelles sont devenues de plus en plus populaires dans au sein de la statistique publique au cours de la dernière décennie. Un facteur clé a été l’émergence de nouvelles sources de données numériques qui offrent un accès facile, parfois presque en temps réel à de telles données. En 2020, la demande s’est fortement accélérée juste après l’apparition de la pandémie de COVID-19, lorsque de nombreux utilisateurs institutionnels ont demandé des indicateurs plus précoces afin de suivre l’évolution de l’économie sur une base hebdomadaire voire quotidienne. Fortement saisonnières dans de nombreux cas, les données économiques infra-mensuelles ne se prêtent pas à la modélisation avec les approches classiques de Box-Jenkins et à la désaisonnalisation avec les méthodes traditionnelles en raison de particularités que l’on ne retrouve pas dans les données mensuelles et trimestrielles. Parmi les principaux exemples, on peut citer la superposition de multiples composantes saisonnières avec des périodicités potentiellement non entières, la mesurabilité directe des effets de calendrier et une forte susceptibilité à l’espacement irrégulier, aux valeurs aberrantes et aux valeurs manquantes (voir les discussions dans (Ollech, 2023; Proietti et Pedregal, 2022; Webel, 2022)). Ainsi, les stratégies possibles pour modéliser et désaisonnaliser les séries infra-mensuelles sont (1) une régularisation appropriée des données afin que les approches traditionnelles deviennent applicables aux données régularisées, (2) des modifications appropriées des approches traditionnelles afin qu’elles deviennent applicables aux données infra-mensuelles, et (3) la conception de nouvelles approches non utilisées dans les statistiques officielles jusqu’à présent.

Les solutions actuelles offertes par JD+ 3.0 suivent la deuxième branche puisqu’elles sont fondées sur la modification d’algorithmes établis, tels que TRAMO pour le pré-ajustement des données et des versions étendues de SEATS (AMB, Arima Model Based), X-11 et STL pour la décomposition. Les méthodes correspondantes ont été mises en œuvre en Java et sont facilement accessibles par le biais d’un écosystème de packages R (voir les [Annexes](#) pour plus de détails). Bien que le code Java soit encore expérimental, quelques aperçus méthodologiques et applications sont déjà disponibles (Ladiray, Palate, Mazzi et Proietti, 2018; Webel, 2022). Dans la suite de ce document, nous détaillons les principales modifications des algorithmes, y compris d’un point de vue théorique, discutons les principales propriétés des modèles et des méthodes de filtrage, fournissons des exemples et des extraits de code.

Nous commençons par une discussion sur le pré-ajustement des données au moyen d’un modèle de régression sur séries temporelles dans lequel les résidus suivent un modèle Arima de type Airline qui autorise les puissances non entières de l’opérateur retard ([section 2](#)). Ce modèle est ensuite utilisé pour introduire les modifications appropriées à l’approche AMB traditionnelle ([section 3](#)). Les principales extensions des approches traditionnelles X-11 et STL sont présentées respectivement dans la [section 4](#) et la [section 5](#), suivies d’une brève description de l’utilisation de modèles structurels ([section 6](#)). Nous

illustrons les principales caractéristiques des méthodes discutées en désaisonnant les naissances quotidiennes en France, la consommation horaire d'électricité en Allemagne et les demandes initiales hebdomadaires d'assurance chômage aux États-Unis (section 7) avant de conclure par quelques remarques et des suggestions de futurs développements (section 8).

## 2 Pré-ajustement

### 2.1 Modèle et propriétés

Soit  $\{y_t\}$  une série temporelle infra-mensuelle observée, décomposée de manière additive, éventuellement après un passage en logarithme, en une composante tendance-cycle  $\{t_t\}$ , une composante saisonnière  $\{s_t\}$ , une composante calendaire  $\{c_t\}$  et une composante irrégulière  $\{i_t\}$ . On obtient ainsi la décomposition standard en composantes inobservables (UC, Unobservable Components).

$$y_t = t_t + s_t + c_t + i_t. \quad (1)$$

Comme pour les données mensuelles et trimestrielles, les effets de calendrier et ceux des valeurs aberrantes (outliers) doivent être éliminées de  $\{y_t\}$  avant l'extraction des mouvements saisonniers. Cette linéarisation est réalisée par régression de la série temporelle

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

où  $\mathbf{x}_t$  est un vecteur de variables de régression associées aux effets de calendrier et aux valeurs aberrantes,  $\boldsymbol{\beta}$  est le vecteur des coefficients correspondants et  $\eta_t$  est un processus résiduel.

Les séries infra-mensuelles contiennent souvent des composantes saisonnières multiples, de sorte que la variation saisonnière de (1) peut être représentée comme suit

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

où  $\mathbb{S} = \{\tau_1, \dots, \tau_{|\mathbb{S}|}\} \subset \mathbb{R}$  est un ensemble de périodicités saisonnières et  $\{s_t^{(\tau)}\}$  est la composante saisonnière avec une périodicité potentiellement non-entière  $\tau$ . Les résidus de l'équation (2) contiennent naturellement les mêmes composantes saisonnières superposées et ne peuvent donc pas être modélisés comme un processus Arima standard. Une modélisation alternative avec un modèle Airline étendu permet une intégration saisonnière multiple du premier ordre, y compris pour les puissances non entières de l'opérateur retard. Ce modèle s'écrit

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

où  $B$  est l'opérateur retard,  $\delta_1(B) = 1 - B$  est l'opérateur de différenciation non saisonnière,  $\delta_\tau(B) = 1 - B^\tau$  est l'opérateur de différenciation saisonnière au retard  $\tau$ ,  $\theta_1(B) = 1 - \theta_1 B$  est un opérateur de moyenne mobile non saisonnière,  $\theta_\tau(B) = 1 - \theta_\tau B^\tau$  est un opérateur de moyenne mobile saisonnière au retard  $\tau$  et  $\{\varepsilon_t\}$  est un bruit blanc avec une

variance finie  $\sigma_\varepsilon^2 > 0$ . Les puissances non entières de  $B$  sont définies par l'approximation de Taylor du premier ordre au voisinage de 1, en utilisant le fait que  $B^k = B^{\lfloor k \rfloor} B^{\alpha_k}$  est valable pour tout  $k$  dans  $\mathbb{R}$ , où  $\lfloor k \rfloor$  est le plus grand entier ne dépassant pas  $k$  et  $\alpha_k = k - \lfloor k \rfloor$  dans  $[0, 1[$  est le reste décimal de  $k$ . Maintenant,  $B^{\alpha_k} \approx (1 - \alpha_k) + \alpha_k B$  et, par conséquent,  $B^k$  est approximé par une moyenne pondérée de  $B^{\lfloor k \rfloor}$  et  $B^{\lfloor k \rfloor + 1}$  :

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

L'approximation (5) peut être introduite dans les opérateurs de différenciation saisonnière et de moyenne mobile (MA) dans (4). Pour des données hebdomadaires, par exemple, nous avons  $|\mathbb{S}| = 1$  avec  $\tau_1 = 52, 18$ , ainsi la différenciation saisonnière devient

$$\delta_{52,18}(B) y_t = (1 - B^{52,18}) y_t = y_t - (0,82 y_{t-52} + 0,18 y_{t-53}). \quad (6)$$

De manière générale, le gain au carré de tout opérateur de différenciation saisonnière  $\delta_\tau(B)$  est donné par

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

où

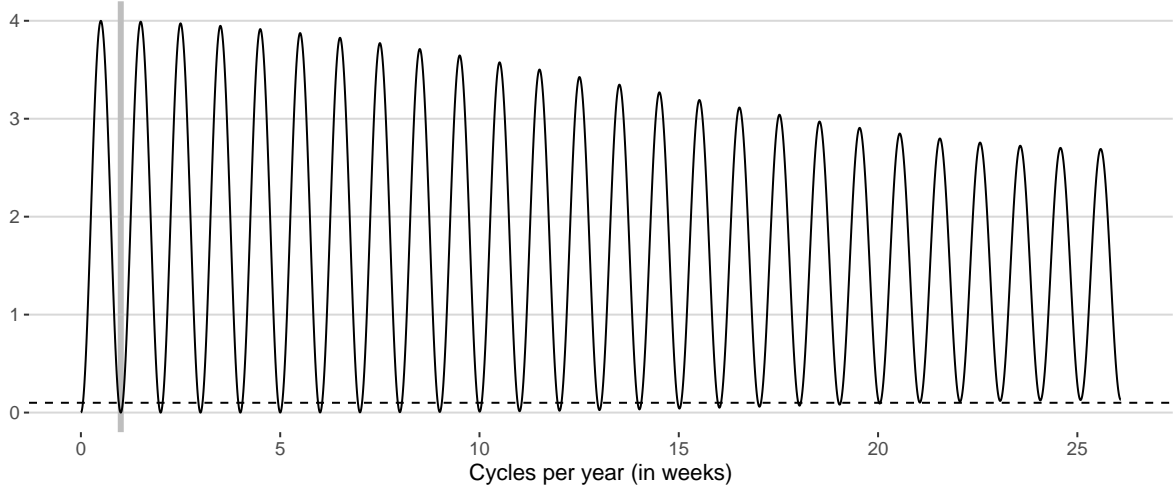
$$c_\tau(\omega) = (1 - \alpha_\tau) \cos(\lfloor \tau \rfloor \omega) + \alpha_\tau \cos([\lfloor \tau \rfloor + 1] \omega)$$

et  $s_\tau(\omega)$  est défini de façon analogue en utilisant la fonction sinus au lieu de la fonction cosinus. Notons que  $c_\tau(\omega) = \cos(\tau\omega)$  et  $s_\tau(\omega) = \sin(\tau\omega)$  si  $\tau \in \mathbb{N}$ , de sorte que (7) se simplifie à un cas standard bien connu. La [figure 1](#) montre le gain au carré (7) pour l'opérateur de différenciation saisonnière appliqué dans (6), révélant que cet opérateur n'annihile pas complètement la dynamique de la semaine-de-l'année (WOY, week-of-the-year). Bien que le gain au carré diminue aux harmoniques de la semaine-de-l'année, il n'atteint pas zéro (en fait, ces diminutions restent même supérieures à 0,1 aux harmoniques plus élevés). Ce défaut est commun à tous les opérateurs  $\delta_\tau(B)$  avec un  $\tau$  non entier. Il est une conséquence de la troncature (du premier ordre) dans le développement en série de Taylor de  $B^{\alpha_\tau}$ , qui fait que  $\delta_\tau(B)$  n'a qu'une seule racine unitaire. Pour s'en convaincre, il suffit de noter que cet opérateur se factorise sous la forme  $\delta_\tau(B) = \delta_1(B) S_\tau(B)$ , où

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

est l'opérateur d'agrégation fractionnaire associé à  $\tau$ . Lorsque  $\tau \in \mathbb{N}$  et donc  $\alpha_\tau = 0$ , alors  $S_\tau(B)$  porte  $\lfloor (\tau - 1)/2 \rfloor$  paires de racines unitaires complexes conjuguées, également espacées autour du cercle unité, aux fréquences saisonnières et accompagnées d'une racine unitaire réelle supplémentaire unique égale  $-1$  si  $\tau$  est pair. Cependant, lorsque  $\tau \notin \mathbb{N}$  et donc  $\alpha_\tau > 0$ ,  $S_\tau(B)$  doit contenir une racine supplémentaire puisque l'ordre polynomial de (8) a été augmenté d'une unité. Bien que les racines de  $S_\tau(B)$  reflètent toujours la dynamique saisonnière, elles ne sont plus exactement associées aux fréquences saisonnières respectives et se situent également en dehors du cercle unité. <sup>1</sup>

<sup>1</sup>En fait, la racine réelle à  $-1$  qui est présente dans (8) lorsque  $\tau \in \mathbb{N}$  est pair se transforme en une racine complexe avec une fréquence associée proche de  $\pi$  lorsque  $\tau \notin \mathbb{N}$  et  $\lfloor \tau \rfloor$  est pair. De même, une racine réelle négative supplémentaire est introduite lorsque  $\tau \notin \mathbb{N}$  et  $\lfloor \tau \rfloor$  est impair (par rapport au cas où  $\tau \in \mathbb{N}$  est impair). Les solutions numériques pour les périodicités non entières basées sur l'algorithme de



**Figure 1 :** Gain au carré (7) pour  $\tau = 52, 18$ . La verticale grise marque la fréquence principale de la semaine-de-l'année. L'horizontale en pointillé indique 0, 1.

Le principe de la moyenne pondérée (5) facilite également le calcul de la densité pseudo-spectrale des résidus de la régression  $\{\eta_t\}$  dans (2). Il découle directement de (4) et de (5) que

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

où les gains au carré des filtres de différenciation saisonnière sont donnés par (7) et ceux des autres filtres présents s'écrivent

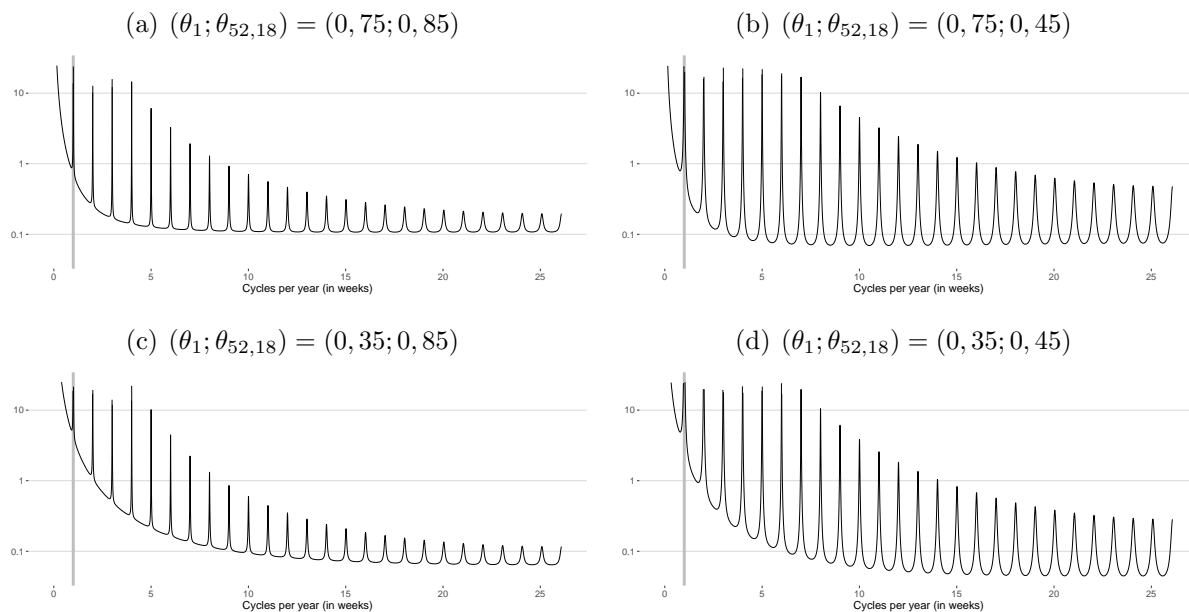
$$\begin{aligned} |\theta_1(e^{-i\omega})|^2 &= 1 - 2\theta_1 \cos(\omega) + \theta_1^2, \\ |\theta_\tau(e^{-i\omega})|^2 &= 1 - 2\theta_\tau c_\tau(\omega) + \theta_\tau^2 [c_\tau^2(\omega) + s_\tau^2(\omega)], \\ |\delta_1(e^{-i\omega})|^2 &= 2[1 - \cos(\omega)]. \end{aligned}$$

La figure 2 montre la densité pseudo-spectrale (9) pour le modèle Airline hebdomadaire étendu et quatre combinaisons de ses paramètres MA. On constate que, comme pour le modèle Airline classique, le paramètre MA non saisonnier régit la forme de (9) lorsque  $\omega \rightarrow 0$  (des valeurs plus élevées/plus faibles de  $\theta_1$  correspondent à des augmentations plus fortes/plus faibles), tandis que le paramètre MA saisonnier régit la forme de (9)

---

Jenkins-Traub (Jenkins et Traub, 1972) suggèrent que les racines de  $S_\tau(B)$  sont associées à des fréquences légèrement plus grandes que les fréquences saisonnières et ont des modules légèrement plus grands que 1. Par exemple, lorsque  $\tau = 52, 18$ , les différences en radians entre les fréquences des racines complexes dans (8) et les fréquences respectives de la WOY augmentent de  $5,2 \cdot 10^{-7}$  pour la fréquence principale de la WOY à  $9,8 \cdot 10^{-3}$  pour la 25-ième harmonique de la WOY, tandis que les modules correspondants augmentent progressivement de 1,000021 à 1,008634. Conceptuellement, lorsque  $\tau \notin \mathbb{N}$ , le principe (5) sacrifie les propriétés de racine unitaire saisonnière de l'opérateur d'agrégation fractionnaire  $S_\tau(B)$  qui apparaît dans la factorisation descendante de  $\delta_\tau(B)$ . Un autre opérateur de différenciation saisonnière qui conserve ces propriétés de racine unitaire saisonnière peut être construit via une sélection a priori des fréquences saisonnières pertinentes et une multiplication atomique ascendante des polynômes de degré minimal respectifs. Voir la discussion dans McElroy et Livsey, 2022, Annexe B).

au voisinage des fréquences WOY  $\omega_k = 2\pi k/52,18$  avec  $k \in \{1, \dots, 26\}$  (les valeurs supérieures/inférieures de  $\theta_{52,18}$  correspondent à des pics plus étroits/plus larges). Cependant, contrairement au modèle Airline classique, l'ampleur des pics saisonniers diminue considérablement pour les harmoniques WOY les plus élevés, en raison du fait susmentionné que  $\delta_{52,18}(B)$  n'annihile pas entièrement la dynamique WOY (rappelons que son gain au carré apparaît dans le dénominateur de (9)).



**Figure 2** : Densité pseudo-spectrale (9) avec  $\mathbb{S} = \{52, 18\}$ ,  $(\theta_1; \theta_{52,18}) \in \{0, 35; 0, 75\} \times \{0, 45; 0, 85\}$  et  $\sigma_\varepsilon^2 = 1$  (en décibels). La verticale grise indique la fréquence principale de la semaine-de-l'année.

## 2.2 Estimation du modèle

Les paramètres inconnus du modèle de pré-ajustement des équations (2) et (4) sont estimés par la procédure itérative schématisée dans le tableau [Algorithm 1](#). Essentiellement, à chaque itération, les paramètres de régression dans (2) sont estimés par les moindres carrés ordinaires (MCO), tandis que les paramètres ARMA dans (4) sont estimés par les techniques du maximum de vraisemblance (MV). En parcourant les itérations, les estimations optimales des paramètres de régression et ARMA sont obtenues par l'algorithme de Levenberg-Marquardt (Levenberg, 1944; Marquardt, 1963). La liste suivante contient des détails supplémentaires, les numéros se réfèrent aux lignes dans le tableau [Algorithm 1](#) :

- 3 : Pour éviter les inversions matricielles fastidieuses, la solution des MCO utilise la factorisation QR de la matrice des variables explicatives, qui est calculée *via* les transformations de Householder.
- 4–7 : La recherche automatique des valeurs aberrantes par addition et suppression est facultative. Les outliers pris en compte sont les Additive Outliers (AO), les Level

---

**Algorithm 1** Estimation itérative du modèle de pré-ajustement (2) et (4)

---

Soit  $\boldsymbol{\psi} = (\theta_1, \theta_{\tau_1}, \dots, \theta_{\tau_{|s|}}, \sigma_\varepsilon^2)^\top$  le vecteur des paramètres ARMA de (4), y compris la variance du bruit blanc.

- 1: Set  $\boldsymbol{\psi}^{(0)} = (0, 2; \dots; 0, 2; 1)^\top$ .
  - 2: **repeat**
  - 3: Étant donné  $\hat{\boldsymbol{\beta}}^{(i-1)}$  dans l'équation (4), estimer  $\hat{\boldsymbol{\beta}}^{(i)}$  dans (2) par MCO après avoir appliqué le filtre et lisseur de Kalman (KFS) à  $\{y_t\}$  et  $\{\mathbf{x}_t\}$ .
  - 4: **repeat**
  - 5: Estimer par régressions MCO successives sur les résidus KFS chaque type d'outlier potentiel à chaque date et ajouter l'outlier le plus significatif à (2).
  - 6: Enlever pas à pas les outliers non significatifs de (2) sans ré-estimer le modèle.
  - 7: **until** une estimation stable de (2) est trouvée ou un maximum de 100 boucles est atteint.
  - 8: Étant donné  $\hat{\boldsymbol{\beta}}^{(i)}$  dans (2), déterminer une représentation espace-état du modèle (4), appliquer le KFS et estimer  $\hat{\boldsymbol{\psi}}^{(i)}$  par maximum de vraisemblance, en utilisant le résultat du KFS.
  - 9: **until** convergence, ou un maximum de 200 itérations est atteint.
- 

Shifts (LS) et les interventions de commutation de lag-1 (WO), qui sont un cas particulier de valeurs aberrantes de réaffectation (Wu, Hosking et Ravishanker, 1993). Compte tenu d'une valeur critique préétablie, la significativité des valeurs aberrantes est déterminée au moyen d'une statistique semblable à celle calculée dans deJong et Penzer (1998).

- 8 : Le modèle de l'équation (4) est placé dans la représentation markovienne d'Akaike (espace-état minimal) (Gomez et Maravall, 1994). La mise en œuvre du KFS utilise les récursions rapides de Chandrasekhar (Morf, Sidhu et Kailath, 1974; Morf et Kailath, 1975) et un calcul approprié de la matrice de covariance de l'état initial (Jones, 1980). Les estimations MV sont obtenues en maximisant la fonction de log-vraisemblance concentrée (Francke, Koopman et deVos, 2010).
- 9 : La convergence est mesurée par les changements incrémentaux des trois éléments suivants : 1 - les paramètres de régression estimés dans l'équation (2), 2 - le gradient l'équation (2) de la fonction de vraisemblance utilisée pour obtenir les estimations MV des paramètres ARMA dans l'équation (4) et 3 - la somme des carrés résiduels de l'ensemble du modèle de pré-ajustement.

### 3 Approche basée sur le modèle Arima étendu

L'extension des algorithmes de désaisonnalisation des séries mensuelles et trimestrielles basés sur les modèles Arima (AMB) repose également sur le modèle Airline étendu de l'équation (4). Dans un premier temps, ce modèle est estimé (pour la série linéarisée) de la manière décrite dans la [section 2](#). Dans un deuxième temps, la décomposition canonique de ce modèle Airline étendu en composantes inobservables (UC) est trouvée selon l'approche classique de Burman (1980). (Voir également Gomez et Maravall (2001b) et Maravall

(1995) pour un exposé détaillé) Pour cela, les opérateurs de différenciation non saisonnière et saisonnière dans (4) sont factorisés au moyen d’une division polynomiale, ce qui, après substitution de (8) dans (4), se traduit par

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

de sorte que les racines unitaires proches 1 sont affectées à la tendance-cycle et que les racines non stationnaires ( $\alpha_\tau = 0$ ) ou stationnaires ( $\alpha_\tau > 0$ ) des opérateurs d’agrégation fractionnaire sont affectées aux composantes saisonnières correspondantes.

Compte tenu d’une telle factorisation, la décomposition de la partie MA du modèle est réalisée par le biais d’une expansion fractionnaire partielle. La combinaison des polynômes de différenciation et de MA des modèles UC selon la théorie classique de Wiener-Kolmogorov (WK) donne alors les estimateurs de l’erreur quadratique moyenne minimale (MMSE) pour chaque composante inobservable, y compris pour la série désaisonnalisée. Dans certains cas, cette décomposition peut être inadmissible dans le sens où le spectre de la composante irrégulière estimée est négatif pour au moins une fréquence. Une inflation de la variance est utilisée pour résoudre ce problème, c’est-à-dire qu’un bruit blanc artificiel est ajouté au modèle jusqu’à ce que le spectre de la composante irrégulière devienne non négatif. Ensuite, les filtres WK et les estimateurs MMSE sont recalculés à partir de la décomposition admissible du modèle Airline étendu gonflé par le bruit. Il convient toutefois de noter que l’estimateur MMSE pour tout modèle saisonnier avec une périodicité fractionnaire sera techniquement stationnaire puisqu’il hérite de la transformation induisant la stationnarité du modèle UC correspondant.

Bien que les filtres WK fournissent en théorie des estimateurs optimaux des composantes inobservables en termes d’erreur quadratique moyenne, des études empiriques préliminaires menées à la Banque nationale de Belgique suggèrent que la décomposition classique peut devenir assez rapidement numériquement instable pour les données infra-mensuelles et que, par conséquent, ces estimations optimales peuvent ne pas être fiables. Ainsi, des estimations affinées des composantes inobservables sont calculées dans une troisième étape *via* une application d’un KFS à la représentation espace-état du modèle Airline étendu décomposé. Le KFS (diffuse square-root) est utilisé si l’on souhaite obtenir les écarts-types des estimations des composantes inobservables, sinon un lisseur rapide des résidus (Koopman, 1993) est appliqué (également avec une initialisation diffuse). Quoi qu’il en soit, le processus de décomposition AMB produit des données rétrolées et des prévisions pour les estimations des composantes inobservables. Celles-ci peuvent ensuite être utilisées pour construire par simple agrégation une rétropolation et des prévisions pour les séries temporelles observées.

## 4 Approche X-11 étendue

La méthode classique de désaisonnalisation X-11 dérivée de Shiskin, Young et Musgrave (1967) et examinée en détail dans Ladiray et Quenneville (2001) s’applique aux séries temporelles mensuelles et trimestrielles. Elle est basée sur l’application itérative du principe suivant en 4 étapes. Premièrement, la tendance-cycle est extraite de la série brute par l’application de  $2 \times \tau$  moyennes mobiles ou de filtres de Henderson, en fonction

de la position de l'opération dans l'ensemble du processus de filtrage. Deuxièmement, la tendance-cycle estimée est supprimée de la série brute. Troisièmement, la série brute privée de la tendance est séparée en composantes saisonnières normalisées et irrégulières par l'application de moyennes mobiles saisonnières de  $3 \times k$ . À certains endroits, cette étape comprend également une détection automatique et une pondération à la baisse des valeurs extrêmes dans la composante irrégulière sur la base de  $\sigma$ -limites. Quatrièmement, la composante saisonnière estimée est supprimée de la série brute. Ce principe clé est appliqué deux fois dans chacune des itérations B à D de l'approche X-11. Ensuite, les estimations finales de la tendance-cycle et de la composante irrégulière sont obtenues en décomposant la série désaisonnalisée finale à l'aide d'un filtre d'Henderson, complétant ainsi une séquence d'estimations préliminaires, intermédiaires et finales des composantes inobservables.

L'approche X-11 étendue adopte ce processus de lissage itératif mais incorpore certaines généralisations et modifications adaptées aux particularités des données infra-mensuelles. Cette section décrit les principaux changements apportés à la méthode X-11 classique. La mise en œuvre actuelle permet l'extraction séquentielle des tendances saisonnières dans l'équation (3) sans sélection automatique des filtres de tendance-cycle et saisonniers basée sur les rapports  $I/C$  et  $I/S$  et sans aucune prévision, même naïve, de la composante saisonnière.

## 4.1 Filtres de tendance-cycle

### 4.1.1 Estimation préliminaire de la tendance-cycle

L'extraction préliminaire de la tendance-cycle dans la méthode classique X-11 est effectuée avec des moyennes mobiles symétriques centrées  $2 \times \tau$  pour  $\tau \in \{4, 12\}$ . L'approche X-11 étendue utilise une version généralisée de ces filtres bruts permettant de prendre en compte n'importe quelle périodicité saisonnière  $\tau$ . Soit  $\mathbb{N}_{\text{odd}}$  l'ensemble des entiers impairs et  $\lceil k \rceil^{\text{odd}}$  le plus petit entier impair non inférieur à  $k$ , c'est-à-dire  $\lceil k \rceil^{\text{odd}} = \min \{l \geq k \mid l \in \mathbb{N}_{\text{odd}}\}$ . Alors, la moyenne mobile généralisée  $2 \times \tau$  a une longueur  $l_\tau = \lceil \tau \rceil^{\text{odd}}$  et ses poids sont donnés par

$$w_i = \begin{cases} 1/\tau, & i \in \{1, l_\tau\} \text{ with } \tau \in \mathbb{N}_{\text{odd}} \\ (\mathbb{I}\{\lceil \tau \rceil \text{ pair}\} + \alpha_\tau)/(2\tau), & i \in \{1, l_\tau\} \text{ with } \tau \notin \mathbb{N}_{\text{odd}}, \\ 1/\tau, & i \in \{2, \dots, l_\tau - 1\} \end{cases} \quad (10)$$

où  $\mathbb{I}\{\cdot\}$  est la fonction indicatrice. Il est facile de voir que  $\sum_{i=1}^{l_\tau} w_i = 1$  est valable pour tout choix de  $\tau$ . Les variantes asymétriques ne sont pas utilisées pendant l'extraction préliminaire de la tendance-cycle, de sorte que  $(l_\tau - 1)/2$  observations sont temporairement perdues au début et à la fin de la série en sortie du filtre.

### 4.1.2 Estimation affinée et finale de la tendance-cycle

Les filtres intermédiaires et finaux d'extraction de la tendance-cycle dans la méthode classique X-11 sont un ensemble de poids pré-spécifiés pour les filtres de Henderson symétriques à  $m$  termes avec  $m \in \{3, \dots, 101\} \cap \mathbb{N}_{\text{odd}}$  et leurs substituts asymétriques

de Musgrave. L'approche X-11 étendue met en œuvre la méthode généralisée d'estimation de la tendance-cycle dérivée de Proietti et Luati (2008), qui est proche de celle de Gray et Thomson (2002), ainsi que trois façons très différentes de calculer les variantes asymétriques.

L'estimation de la tendance-cycle selon Proietti et Luati (2008) est fondée sur l'application de régressions polynomiales locales à la série en entrée. Soit  $h$  l'horizon local, ou demi-longueur, du filtre tendance-cycle et  $d$  le degré polynomial que le filtre doit pouvoir préserver. De même,  $q \in 0, \dots, h$  est le nombre d'observations futures disponibles après le point courant  $t$ . Le modèle de régression polynomiale locale est alors donné par

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

où  $\mathbf{z}_t = (z_{t-h}, \dots, z_{t+q})^\top$  est le span local de la série en entrée,

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

est la matrice des variables explicatives constante dans le temps,  $\boldsymbol{\zeta} = (\zeta_0, \dots, \zeta_d)^\top$  est le vecteur des coefficients de régression inconnus de dimension  $(d+1)$  et  $\boldsymbol{\varepsilon}_t = (\varepsilon_{t-h}, \dots, \varepsilon_{t+q})^\top$  est un vecteur local représentant un bruit blanc gaussien de moyenne nulle, dont les éléments sont à la fois mutuellement et sériellement non corrélés. Un estimateur des moindres carrés pondérés (WLS) de  $\boldsymbol{\zeta}$  est obtenu par minimisation de la fonction objectif

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

où  $\{\kappa_j\}$  est un ensemble de poids non négatifs définis selon un noyau déterminé.<sup>2</sup> En utilisant la théorie standard des moindres carrés, l'estimateur WLS (Weighted Least Squares) de l'équation (12) est trouvé comme suit

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

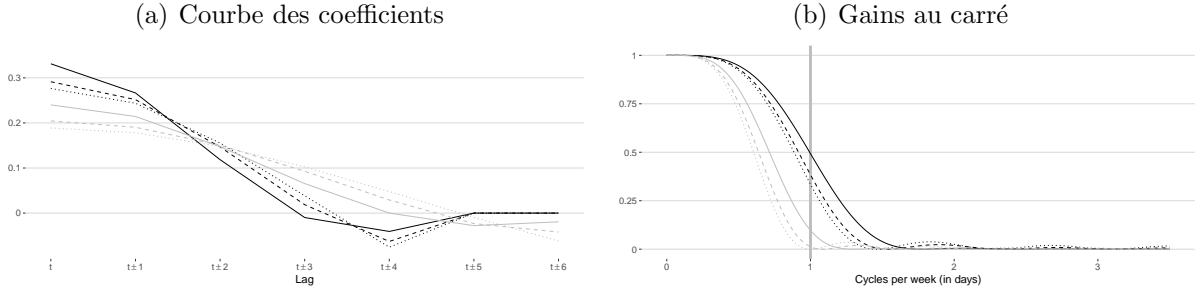
avec  $\mathbf{K}_q = \text{diag}(\kappa_{-h}, \dots, \kappa_q)$ . Puisque l'estimateur local de la tendance-cycle est donné par  $\hat{\zeta}_0 = \mathbf{e}_1^\top \hat{\boldsymbol{\zeta}}$ , où  $\mathbf{e}_1$  est le premier vecteur unitaire de dimension  $(d+1)$ , nous avons finalement  $\hat{\zeta}_0 = \mathbf{w}_q^\top \mathbf{z}_t$  avec

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

En supposant que  $2h \geq d$ , les poids des filtres d'extraction de la tendance-cycle basés sur des noyaux symétriques résultent de la définition de  $q = h$  dans l'équation (13). La [figure 3](#) montre les gains au carré des filtres cubiques dérivés des noyaux de Henderson, d'Epanechnikov et gaussiens à variance unitaire pour les horizons  $h \in \{4, 6\}$  (ce qui

<sup>2</sup>Le package R implémente actuellement les noyaux biweight, Epanechnikov, gaussien, Henderson, trapézoïdal, triangulaire, tricube, triweight et uniforme.

correspond aux longueurs des filtres classiques d'extraction de tendances et de cycles de 9 et 13). Chacun de ces filtres peut être utilisé pour l'extraction de la tendance-cycle lors de l'ajustement saisonnier des données quotidiennes avec une dynamique prononcée du jour de la semaine. Quelque soit l'horizon, les gains au carré des trois filtres ont une forme très similaire, bien que la bande passante devienne légèrement plus étroite lorsque les noyaux d'Epanechnikov et gaussien sont utilisés à la place du noyau d'Henderson.



**Figure 3 :** Filtres d'extraction symétriques de la tendance-cycle basés sur un noyau de Henderson (*trait plein*), Epanechnikov (*pointillés*) et Gaussien de variance unitaire (*points*) (13) avec  $q = h = 4$  (*noir*) et  $q = h = 6$  (*gris*) et  $d = 3$ . La ligne verticale grise sur le graphique (b) indique la fréquence principale day-of-the-week (DOW).

Les poids des filtres asymétriques correspondants peuvent être obtenus pour tout choix de  $q \in \{0, \dots, h - 1\}$  dans l'équation (13). Par exemple, les filtres d'extraction de la tendance-cycle basés sur des noyaux contemporains sont donnés par  $\mathbf{w}_0$ . Ces variantes sont conceptuellement cohérentes avec les filtres symétriques  $\mathbf{w}_h$  - et sont donc appelées filtres asymétriques "directs" - mais produisent souvent des estimations de tendance-cycle fortement localisées et donc volatiles. L'approche "cut and normalize" (Gasser et Mueller, 1979) est une alternative assez simple permettant de trouver les filtres asymétriques correspondants en supprimant les poids inutiles du filtre symétrique  $\mathbf{w}_h$  et en divisant les poids restants par leur somme, c'est à dire

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

où  $\mathbf{w}_h^{(q)}$  ne contient que les premiers  $h + q + 1$  éléments de  $\mathbf{w}_h$  et  $\mathbf{1}$  est un vecteur unitaire de dimension  $(h + q + 1)$ . Une alternative plus sophistiquée est l'approche "erreur de révision quadratique moyenne minimale" (MMSRE), également examinée dans Proietti et Luati (2008). Visant un compromis biais-variance, son idée principale est d'introduire délibérément un léger biais dans les estimateurs de tendance-cycle en supposant que le degré polynomial est d'ordre inférieur à la frontière et en recherchant des filtres asymétriques de tendance-cycle qui préservent les polynômes de degré encore plus faible  $\underline{d} < \bar{d}$  sous réserve des contraintes de reproduction imposées par le filtre symétrique  $\mathbf{w}_h$ . Les substituts de Musgrave, par exemple, peuvent être reproduits dans le cas linéaire-constant  $(\bar{d}, \underline{d}) = (1, 0)$  si le noyau de Henderson est utilisé pour calculer  $\mathbf{w}_h$  conformément à l'équation (13). La fonction objectif correspondante, qui est l'analogue de la représentation matricielle de l'équation (12), est alors composée de la variance et du biais quadratique de l'erreur de révision de la tendance-cycle plus un vecteur de multiplicateurs de Lagrange liés aux contraintes de reproduction. La mise en œuvre actuelle

de l'approche MMSRE utilise un critère de minimisation généralisé dérivé dans Grun-Rehomme, Guggemos et Ladiray (2018), qui traduit la fonction objectif susmentionnée dans le domaine des fréquences où elle est décomposable en effets liés aux propriétés de gain et de déphasage des filtres de tendance-cycle utilisés et donc capable d'équilibrer certains critères de fidélité, de lissage et de rapidité.

Notons que le cadre de Proietti et Luati (2008) permet également d'identifier l'horizon  $h$  par validation croisée pour une série donnée, un degré polynomial  $d$  et un ensemble de poids de noyau  $\{\kappa_j\}$  prédéfinis. Toutefois, cette procédure n'a pas encore été mise en œuvre dans le package R.

## 4.2 Filtres saisonniers

La méthode classique X-11 met en œuvre un ensemble de moyennes mobiles saisonnières symétriques de  $3 \times k$ , où  $k \in \{1, 3, 5, 9, 15\}$ , ainsi que des variantes asymétriques afin d'extraire des estimations des facteurs saisonniers de la composante saisonnière des données expurgées de la tendance-cycle (i.e de la composante saisonnière-irrégulière). L'approche X-11 étendue adopte exactement les mêmes filtres saisonniers mais permet davantage de combinaisons de filtres lors de l'extraction des facteurs saisonniers initiaux et finaux dans les itérations B à D. Les estimations saisonnières sont normalisées en utilisant la moyenne mobile symétrique généralisée  $2 \times \tau$  avec des poids donnés dans l'équation (10) et ses variantes asymétriques définies implicitement en remplissant le début (la fin) d'une estimation saisonnière avec  $(l_\tau - 1)/2$  copies de son premier (dernier) élément. Le principe de l'équation (5) est utilisé en plus dans le cas de périodicités non entières, c'est-à-dire que les moyennes pondérées des deux valeurs entières encadrant la périodicité non entière de la composante saisonnière irrégulière constituent le filtre saisonnier. Si l'on considère le filtre saisonnier symétrique  $3 \times 3$ , par exemple, l'estimation des facteurs saisonniers est obtenue à partir des données infra-mensuelles expurgées de la tendance selon la méthode suivante

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

La [figure 4](#) montre le gain au carré du filtre d'extraction saisonnière (14) pour  $\tau = 30, 44$ , c'est-à-dire du filtre qui extrait la composante du jour-du-mois (DOM) des données quotidiennes expurgées de la tendance (d'autres exemples sont présentés sur la [figure 14](#) en [Annexes](#)<sup>3</sup>). Son gain au carré est presque égal à un pour la fréquence principale DOM,

<sup>3</sup>En général, le gain au carré de tout filtre d'extraction saisonnier symétrique de  $3 \times k$  est calculé comme suit :  $|\mathcal{S}_{3 \times k}(e^{-i\omega})|^2$  avec  $\mathcal{S}_{3 \times k}(B)$  la représentation polynomiale du filtre donnée par

mais à partir de la première harmonique, les pics des DOM commencent à diminuer visiblement de façon presque linéaire pour atteindre 0,25 environ. En même temps, le niveau moyen du gain au carré aux fréquences infra-DOM s'élève visiblement au-dessus de zéro entre les harmoniques DOM les plus élevées (avec des pics infra-DOM initialement faibles atteignant presque 0,25 entre la treizième et la quatorzième harmonique). Cet effet de divergence a déjà été observé pour les gains au carré des opérateurs de différenciation saisonnière avec des périodicités non entières (voir par exemple [figure 1](#)). Les filtres saisonniers asymétriques sont dérivés en conséquence. Par exemple, la variante contemporaine de (14) se lit comme suit

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

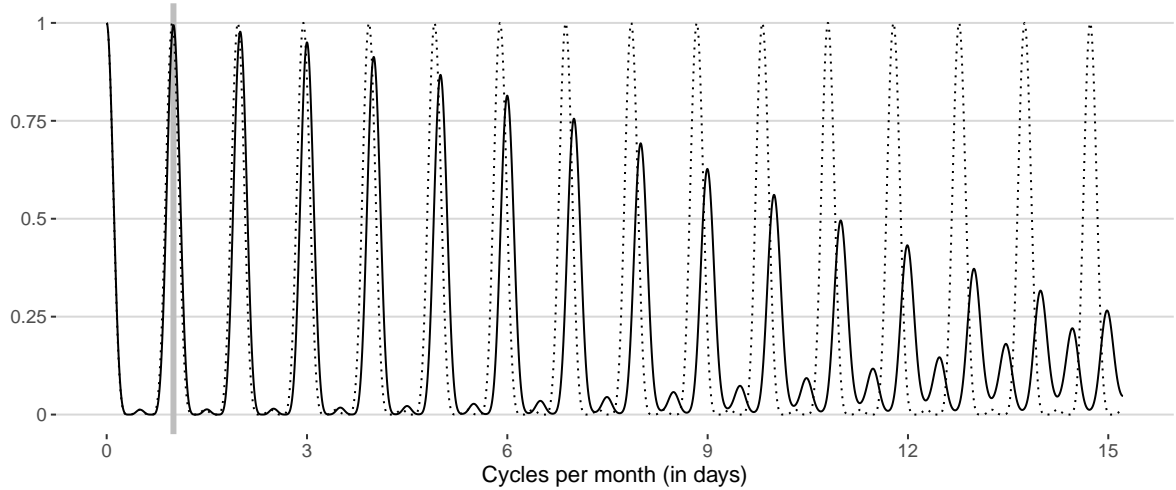
Son gain au carré est représenté sur la [figure 5 \(a\)](#) pour  $\tau = 30, 44$ . Par rapport au filtre symétrique, ce dernier a une forme similaire mais présente des pics légèrement plus élevés et plus larges aux fréquences DOM et "entre-DOM". Le graphique (b) montre que le filtre d'extraction saisonnière contemporain  $3 \times 3$  (15) introduit un retard de phase allant jusqu'à 24 jours dans la variation à basse fréquence de la série filtrée, ce qui peut être négligé puisque le filtre est appliqué à des données expurgées de la tendance. En dehors de cela, aucun retard ou avance de phase notable n'est introduit au-delà de la fréquence DOM principale.

Les filtres saisonniers basés sur le principe d'approximation (5) opèrent aux périodicités saisonnières non entières exactes mais doivent inévitablement compromettre une certaine puissance d'extraction aux harmoniques saisonnières supérieures. D'autres approches de désaisonnalisation pour les séries temporelles infra-mensuelles possèdent un pouvoir d'extraction élevé pour toutes les harmoniques saisonnières au prix d'un fonctionnement avec des périodicités entières légèrement inexactes, ce qui est souvent obtenu par une régularisation temporaire des données. Par exemple, l'approche basée sur STL pour les données quotidiennes mise en œuvre dans le package R `{dsa}` (Ollech, 2021) utilise l'interpolation avec des splines afin d'étirer temporairement chaque mois à 31 jours. Le réglage de  $\tau = 31$  au lieu de  $\tau = 30, 44$  dans (14) imite le lisseur saisonnier LOESS à 5 termes correspondant, et l'effet de compromis susmentionné est visualisé par le gain quadratique en pointillés sur la [figure 4](#).

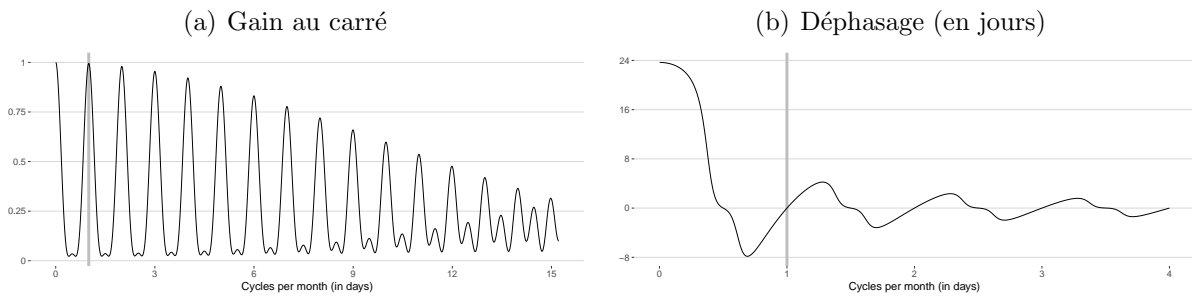
---


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

où les puissances non entières de  $B$  sont définies dans (5).



**Figure 4 :** Gains au carré du filtre saisonnier symétrique  $3 \times 3$  selon l'équation (14) pour  $\tau = 30,44$  (*trait plein*) et  $\tau = 31$  (*pointillés*). La ligne grise verticale indique la fréquence principale jour-du-mois (DOM).



**Figure 5 :** Représentation spectrale du filtre saisonnier contemporain  $3 \times 3$  (15) pour  $\tau = 30,44$ . Les lignes verticales grises indiquent la fréquence fondamentale jour-du-mois (DOM).

## 5 Approche STL

La méthode STL classique mise au point par Cleveland, McRae et Terpenning (1990) adopte schématiquement le concept de filtrage linéaire séquentiel de X-11 dans sa boucle interne (*inner loop*), mais incorpore quelques modifications. La plus notable d'entre elles est probablement l'utilisation de filtres tendance-cycle et saisonniers dont les poids ne sont plus prédéfinis mais dérivés de régressions univariées non paramétriques localement pondérées (LOESS). La fonction objectif LOESS est similaire à (12), les poids du noyau étant remplacés par les poids dits de voisinage obtenus grâce au noyau tri-cube et généralement avec  $d \in \{1, 2\}$ . Une autre modification est l'application d'un filtre passe-bas supplémentaire lors de l'extraction saisonnière. Comme dans X-11, le lissage par période des données expurgées de la tendance produit une estimation de la composante saisonnière, mais dans STL, cette estimation est temporaire car sa variation à basse fréquence est encore éliminée par convolution d'un filtre MA et LOESS afin d'empêcher les lisseurs de tendance-cycle et saisonnier finaux d'être en concurrence pour les mêmes varia-

tions dans les données <sup>4</sup>. Il convient également de noter que les variantes asymétriques du filtre tendance-cycle et du filtre saisonnier ne sont pas nécessaires dans STL, étant donné que le lissage à proximité des points finaux est géré par des poids de voisinage modifiés de manière appropriée dans la régression LOESS. La longueur de la fenêtre, ou du filtre, de cette dernière reste fixée à  $2h + 1$  dans la terminologie de (12), alors que dans X-11 la longueur du filtre diminue progressivement de  $2h + 1$  dans le cas symétrique à  $h + 1$  dans le cas contemporain. Enfin, une boucle externe (*outer loop*) optionnelle fournit des poids de robustesse basés sur le noyau biweight pour la composante irrégulière, ce qui est conceptuellement proche de la détection des valeurs extrêmes de X-11. Les poids de voisinage sont ensuite multipliés par les poids de robustesse lors de l'exécution suivante de la boucle interne, c'est-à-dire que le principe de l'estimation itérée des MCO est appliqué dans (12).

La mise en œuvre actuelle est pour l'essentiel conforme aux recommandations formulées dans Cleveland, McRae et Terpenning (1990) et vise à obtenir une "quasi-certitude de convergence" des estimations STL de la tendance et de la saisonnalité. Par conséquent, nous nous contentons de mentionner à ce stade les principales différences. Deux passages dans la boucle interne sont effectués lorsqu'aucune pondération de robustesse n'est nécessaire et 15 passages dans la boucle externe, avec un passage dans la boucle interne à chaque itération, sont effectués dans le cas contraire. Les polynômes de la fonction objectif LOESS sont d'ordre  $d = 1$  pour les lisseurs de tendance-cycle et  $d = 0$  pour les lisseurs LOESS saisonniers, respectivement. Si elle n'est pas spécifiée, la longueur du lisseur LOESS tendance-cycle pour l'extraction du modèle saisonnier avec une périodicité  $\tau$  est automatiquement fixée à  $\lceil l_t(\tau) \rceil^{\text{odd}}$  avec

$$l_t(\tau) = \left\lfloor \frac{1,5 \lfloor \tau \rfloor}{1 - 1,5 \lfloor l_s(\tau) \rfloor^{-1}} \right\rfloor, \quad (16)$$

où  $l_s(\tau)$  est la longueur du lisseur LOESS saisonnier correspondant. Ce dernier affecte également la longueur du lisseur LOESS dans le filtre passe-bas convolué, qui est donnée par  $\lceil l_s(\tau) \rceil^{\text{odd}}$ . Il est également à noter que (16) fait déjà allusion à la manière dont les périodicités non entières sont traitées dans STL, car elles sont arrondies à l'entier inférieur le plus proche, c'est-à-dire que  $\lfloor \tau \rfloor$  est toujours utilisé, ce qui contraste fortement avec le principe de la moyenne pondérée (5) utilisé dans les approches AMB et X-11 étendues.

## 6 Modèles structurels

Les méthodes de pré-ajustement et de filtrage examinées jusqu'à présent permettent la linéarisation des données et l'extraction du signal de manière séquentielle. Les modèles structurels pour les séries temporelles (STS), quant à eux, visent l'estimation simultanée de toutes les composantes inobservables définies dans (1) grâce à une spécification ascendante de la dynamique pertinente, dont l'agrégat constitue alors naturellement le modèle pour les observations dans l'équation (1). (Comme référence standard voir Harvey (1989).)

<sup>4</sup>La convolution consiste en un filtre MA de  $3 \times \tau \times \tau$  (selon la terminologie X-11) suivi d'un lisseur LOESS avec une longueur de fenêtre recommandée donnée par  $\lceil \tau \rceil^{\text{odd}}$ . Ce filtrage passe-bas supplémentaire se traduit par un gain au carré du lisseur LOESS saisonnier final qui reste proche de zéro au voisinage de  $\omega = 0$ , contrairement à ce que prévoit la norme X-11 (voir figure 4).

La spécification la plus générale de la tendance-cycle dans le cadre des STS est le modèle linéaire local donné par

$$t_t = t_{t-1} + \nu_{t-1} + \xi_t, \quad (17)$$

$$\nu_t = \nu_{t-1} + \chi_t, \quad (18)$$

où  $\{\xi_t\}$  et  $\{\chi_t\}$  sont des bruits blancs gaussiens mutuellement non corrélés avec des moyennes nulles et des variances finies  $\sigma_\xi^2$  et  $\sigma_\chi^2$ , respectivement. D'autres modèles de tendance-cycle couramment utilisés peuvent être dérivés directement de cette forme générale. Par exemple, le modèle de niveau local avec dérive, caractérisé par une pente constante  $\nu$  dans (17), est obtenu en fixant  $\sigma_\chi^2 = 0$  dans (18). De même, la tendance dite lisse est obtenue en fixant  $\sigma_\xi^2 = 0$  dans (17).

La spécification de la composante saisonnière de l'équation (1) est conforme à la représentation de West-Harrison (1997). En supposant  $|\mathbb{S}| = 1$  et  $\tau \in \mathbb{N}$  dans (3) pour faciliter l'exposé (les notations suivantes s'étendent naturellement au cas de composantes saisonnières multiples avec des périodicités potentiellement non entières, voir Proietti et Pedregal (2022) pour plus de détails), la forme générale est la suivante

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

où  $\mathbf{e}_{k,n}$  est le  $k$ -ième vecteur unitaire de longueur  $n$  et  $\{\tilde{\mathbf{s}}_t\}$  est la forme réduite à  $(\tau - 1)$  dimensions des effets saisonniers  $\tau$  donnée par

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

où  $\mathbf{s}_t^\top = (s_{1,t}, \dots, s_{\tau,t})$  est le vecteur des effets saisonniers  $\tau$  au temps  $t$ ,  $\mathbf{P}$  et  $\mathbf{D}$  sont des matrices de permutation et de réduction de dimension données par

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

avec  $\mathbf{0}_n$  et  $\mathbf{1}_n$  des vecteurs colonnes à  $n$  dimensions de zéros et de uns, respectivement,  $\mathbf{I}_n$  la matrice d'identité à  $n$  dimensions et  $\mathbf{D}^- = (\mathbf{D}^\top \mathbf{D})^{-1} \mathbf{D}^\top$  l'inverse de Moore-Penrose de  $\mathbf{D}$ . La forme réduite est alors supposée suivre le modèle vectoriel autorégressif du premier ordre

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

où  $\{\boldsymbol{\omega}_t\}$  est un bruit blanc gaussien vectoriel de moyenne nulle et de matrice de covariance non singulière  $\boldsymbol{\Sigma}_\omega$ . Cette dernière peut être reparamétrée comme  $\boldsymbol{\Sigma}_\omega = \sigma_\omega^2 \boldsymbol{\Omega} \boldsymbol{\Omega}^\top$  pour certaines composantes saisonnières, où  $\boldsymbol{\Omega}$  est une matrice spécifique au modèle qui incorpore explicitement une restriction stochastique à somme nulle dans les effets saisonniers de  $\tau$ . Sa forme exacte est donnée dans Proietti (2000) pour les représentations brute, *dummy*, Harrison-Stevens et trigonométrique.

Le modèle STS complet est finalement présenté sous la forme d'un modèle espace-état gaussien linéaire univarié, qui permet également d'inclure les effets de calendrier de l'équation (1) de manière directe en ajoutant les variables de régression de l'équation (2) à la matrice du système correspondant à l'équation d'observation et les effets associés conte-

nus dans  $\beta$  dans le vecteur d'état. Les valeurs aberrantes peuvent être traitées de la même manière, et une procédure de détection automatique (Grassi, Mazzi et Proietti, 2018) est également disponible. Celle-ci est assez similaire à la procédure utilisée dans le modèle de pré-ajustement (2) et (4), il s'agit schématiquement d'un algorithme de "forward-addition-backward-deletion" basé sur le maximum ponctuel de la  $\tau_t^{*2}$ -statistique dérivée dans deJong et Penzer (1998). Le modèle espace-état complet, y compris les hyperparamètres, est finalement estimé avec le KFS et à partir de la maximisation de sa fonction de vraisemblance. La mise en œuvre suit de près les algorithmes décrits dans Durbin et Koopman (2012).

## 7 Applications

Nous considérons trois exemples de données réelles pour mettre en évidence les caractéristiques clé des méthodes discutées dans les sections précédentes : les naissances quotidiennes en France, la consommation horaire d'électricité en Allemagne et les demandes initiales hebdomadaires d'assurance chômage aux États-Unis. La première série est utilisée pour illustrer le pré-ajustement des données avec les modèles (2) et (4) et l'approche de décomposition AMB étendue (sous-section 7.1). La deuxième série sert à illustrer les approches X-11 et STL étendues (sous-section 7.2), tandis que la troisième série permet de détailler l'estimation simultanée des effets du pré-ajustement et des composantes saisonnières avec les modèles structurels (Structural Time Series, STS) (sous-section 7.3). Pour tous ces exemples, des extraits de code R sont fournis.

### 7.1 Naissances quotidiennes en France

Nous considérons le nombre quotidien des naissances en France Métropolitaine (BIRTHS) entre le 1er janvier 1968 et le 31 décembre 2020, soit 19 359 observations <sup>5</sup> (figure 6 (a)).

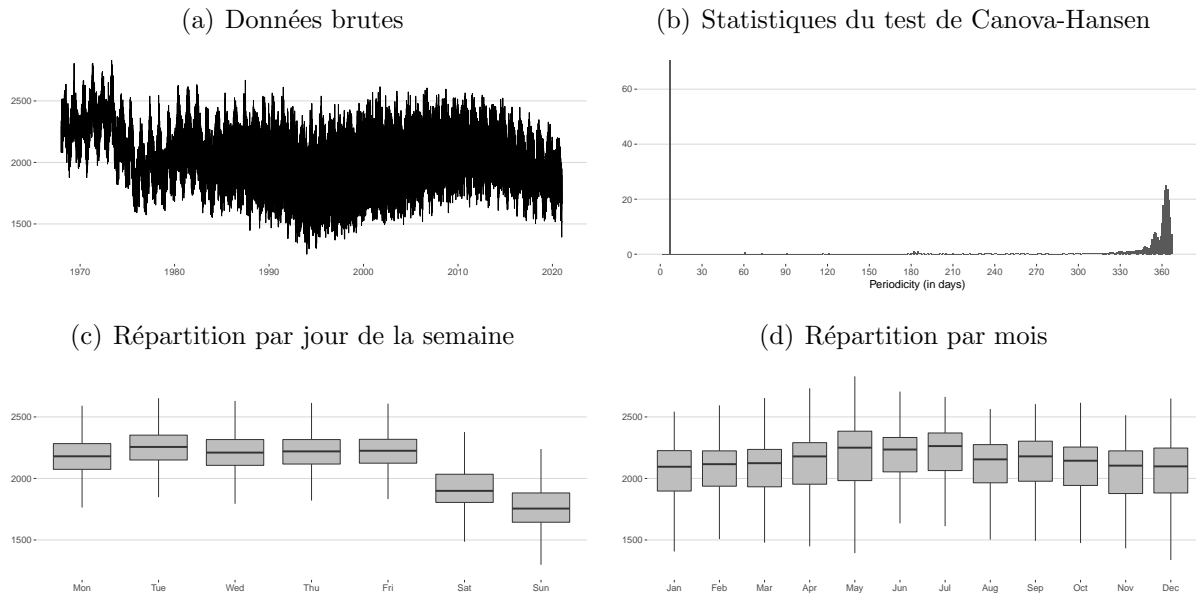
#### 7.1.1 Identification des composantes saisonnières

Le profil saisonnier des naissances quotidiennes est représenté sur la figure 6 (b)–(d). Pour toutes les périodicités entières  $\tau \in \{2, \dots, 367\}$ , nous calculons la statistique du test de Canova-Hansen (CH) généralisé (Canova et Hansen, 1995 ; Busetti et Harvey, 2003) pour l'hypothèse  $H_0$  d'absence de saisonnalité contre l'hypothèse alternative de présence de saisonnalité, soit déterministe soit stochastique-non stationnaire (trigonométrique), en utilisant la commande suivante :

```
R> rjd3toolkit::seasonality.canovahansen(births,
+   p0 = 2, p1 = 367, np = 366, original = TRUE)
```

Sous l'hypothèse nulle, la distribution de la statistique de test CH généralisée appartient à la famille de Cramer-von Mises et les valeurs critiques ont été tabulées dans Harvey (2001). Sur le graphique (b), les statistiques CH présentent des pics à 7 jours et autour

<sup>5</sup>Les données sont disponibles gratuitement auprès de l'Insee sous l'URL <https://www.insee.fr/fr/statistiques/5414759?sommaire=5414771> (Tableau T79JNAIS).



**Figure 6** : Profil saisonnier des naissances quotidiennes en France.

de 365 jours, ce qui suggère la présence de saisonnalité hebdomadaire (day-of-the-week, DOW) et annuelle (day-of-the-year, DOY) dans la série.

Les graphiques (c) et (d) mettent en évidence ces mouvements à l'aide de box-plots. Dans l'ensemble, le nombre de naissances est relativement stable du lundi au vendredi, mais il est visiblement plus faible le week-end, en particulier le dimanche. Il semble également que les naissances soient plus fréquentes au printemps et en été qu'en automne et en hiver. Sur la base de ces éléments, nous avons fixé  $S = \{7; 365, 2425\}$  dans l'équation (4) pour la suite de l'analyse.

### 7.1.2 Pré-ajustement

Pour corriger la série des effets de calendrier, des variables de régression (*dummies*) sont créées avec le package `{rjd3toolkit}` pour les jours fériés fixes et mobiles du calendrier français : Jour de l'An (1er janvier), lundi de Pâques, fête du travail (1er mai), fête de la Victoire (8 mai, célébrée à partir de 1982), Ascension, lundi de Pentecôte, fête Nationale (14 juillet), Assomption (15 août), Toussaint (1er novembre), Armistice (11 novembre) et Noël (25 décembre). En stockant ces variables dans un objet matriciel nommé `hol.FR` et en supposant une décomposition multiplicative en composantes inobservables dans l'équation (1), l'estimation du modèle de pré-ajustement (2) et (4) –y compris la détection automatique des valeurs aberrantes avec des valeurs critiques déterminées en fonction des données– peut être réalisée avec la fonction `fractionalAirlineEstimation()` (voir [tableau 4](#) en [Annexes](#)), à l'aide de la commande suivante :

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

```

+         outliers = c("ao", "wo"),
+         criticalValue = 0)

```

Le [tableau 1](#) présente les effets de calendrier estimés, ainsi que les  $t$ -statistiques et les P-values. Dans l'ensemble, chaque jour férié a un fort effet modérateur sur le nombre de naissances, s'élevant, par exemple, à environ  $-20\%$  le jour de Noël. La composante correspondant aux effets de calendrier estimés peut être extraite de l'objet contenant l'ensemble des résultats de la modélisation, avec la commande

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

où  $n = \text{ncol}(\text{hol.FR})$  est le nombre de régresseurs de calendrier définis par l'utilisateur. En outre, un total de 52 outliers a été détecté automatiquement, mais les effets estimés ne sont pas rapportés ici par souci de concision. Le modèle Airline étendu estimé est donné par

$$\delta(B) w_t = \begin{pmatrix} 1 + 0,257 B \\ (0,007) \end{pmatrix} \begin{pmatrix} 1 - 0,865 B^7 \\ (0,006) \end{pmatrix} \begin{pmatrix} 1 - 0,818 B^{365,2425} \\ (0,005) \end{pmatrix} \varepsilon_t, \quad (20)$$

où  $\delta(B) = \delta_1(B) \delta_7(B) \delta_{365,2425}(B)$ ,  $\{w_t\}$  représente la série linéarisée, les écarts-type sont indiqués entre parenthèses sous les estimations des paramètres.

**Table 1 :** Effets de calendrier estimés pour la série des naissances quotidiennes

Jour férié	Coefficient	Ecart-type	$t$ -value
Jour de l'an	-0,129	0,0063	-20,3479
Lundi de Pâques	-0,123	0,0056	-22,0609
1er Mai	-0,162	0,0043	-37,3471
Armistice 1945	-0,163	0,0058	-28,1360
Ascension	-0,179	0,0043	-41,8757
Lundi de Pentecôte	-0,113	0,0056	-20,2993
14 Juillet	-0,175	0,0043	-40,7100
Assomption	-0,111	0,0056	-19,8978
Toussaint	-0,136	0,0056	-24,2569
Armistice 1918	-0,140	0,0056	-24,9068
Noël	-0,201	0,0057	-35,0796

### 7.1.3 Désaisonnalisation

Les composantes hebdomadaire (DOW) et annuelle (DOY) sont extraites de la série linéarisée avec l'approche basée sur la modélisation Arima (AMB) étendue, dans l'ordre croissant de leurs périodicités afin de minimiser le risque de confusion entre les périodicités. Cela peut être réalisé par une application en deux étapes de la fonction `fractionalAirlineDecomposition()` (voir [tableau 5](#) en [Annexes](#)). L'extraction de la composante saisonnière DOW (sans estimation des écarts-types, sans rétropolation ni prévisions) se fait avec la commande suivante :

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

L'objet `amb.dow` contient toutes les estimations et les paramètres spécifiés par l'utilisateur. Par exemple, la composante saisonnière hebdomadaire (DOW) estimée peut être obtenue *via*

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

Dans un deuxième temps, la composante saisonnière annuelle (DOY) peut être extraite de la série linéarisée et corrigée de la composante saisonnière hebdomadaire (DOW) en utilisant les mêmes options que précédemment, à l'exception de la périodicité saisonnière :

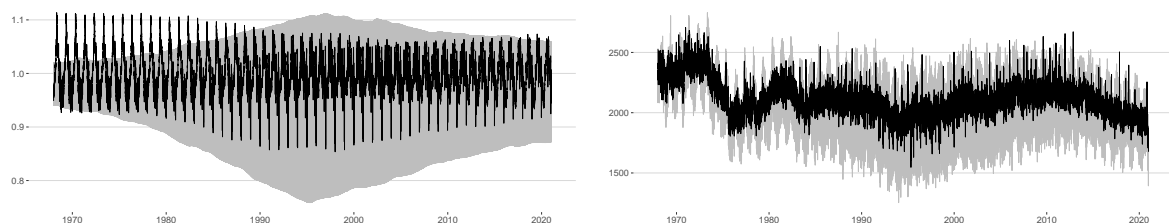
```
R> amb.doy <- rjd3highfreq::fractionalAirlineDecomposition(
+       amb.dow$decomposition$sa,
+       period = 365.2425, ...)
R> births.doy <- exp(amb.doy$decomposition$s)
```

La [figure 7 \(a\)](#) montre les composantes saisonnières `births.dow` et `births.doy`. La première présente une amplitude distinctement variable dans le temps et plus forte au milieu des années 1990. La seconde, en revanche, a une amplitude relativement constante mais subit apparemment des changements structurels infra-annuels progressifs après 1990. Ces deux effets pourraient être une réaction à l'évolution de la volatilité de la série des naissances au cours de la période 1980–2010. Cette explication est étayée par le fait que la volatilité de la série corrigée des variations saisonnières, calculée comme suit

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

reste relativement constante tout au long de la période couverte par les données ([Graphique \(b\)](#)).

(a) Composante saisonnière hebdomadaire (DOW) (*gris*) et composante saisonnière annuelle (DOY) (*noir*) (b) Série brute (*gris*) et corrigée des variations saisonnières (*noir*) data



**Figure 7 :** Ajustement saisonnier (AMB) de la série des naissances quotidiennes.

## 7.2 Consommation horaire d'électricité en Allemagne

Nous nous intéressons à présent à la consommation d'électricité en Allemagne (ELCON) en gigawatts/heure (GWh) à partir du 1er janvier 2015 (00h00) jusqu'au 30 juin 2022 (23h00), ce qui donne 65 712 observations horaires. La série représente l'électricité fournie au réseau général, à l'exclusion de l'électricité fournie au réseau ferroviaire et aux réseaux de distribution industriels internes et fermés, ainsi que l'électricité consommée par les producteurs<sup>6</sup>.

### 7.2.1 Régularisation des données et profil saisonnier

L'heure d'été (DST) est en vigueur en Allemagne et rend les séries horaires d'ELCON légèrement irrégulièrement espacées. Plus précisément, 24 observations sont disponibles pour chaque jour, à l'exception du dernier dimanche de mars, où l'enregistrement de 02h00 est manquant et où l'on ne dispose donc que de 23 observations, et du dernier dimanche d'octobre, où la valeur correspondant à 02h00 est enregistrée deux fois et où l'on dispose donc de 25 observations. Une régularisation des données est donc nécessaire. Nous utilisons la technique simple suivante : une valeur artificielle de 02h00 est calculée pour chaque jour de début d'heure d'été comme la moyenne des valeurs de 01h00 et 03h00 de ce jour et les deux valeurs de 02h00 pour chaque jour de fin d'heure d'été sont également moyennées.

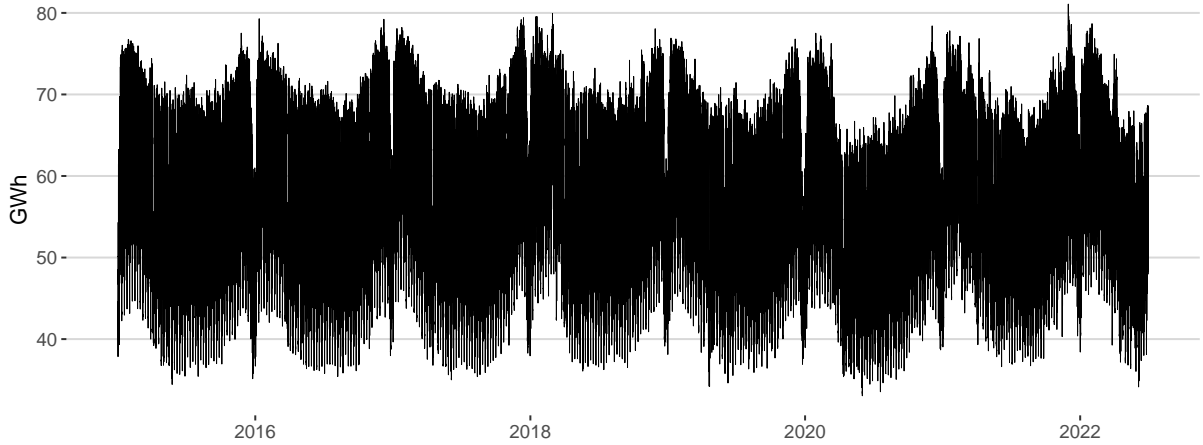
La [figure 8](#) montre le profil saisonnier de la série horaire ELCON régularisée. Comme la consommation d'électricité a tendance à être plus élevée en hiver qu'en été, le profil infra-annuel en forme de  $U$  observé sur le graphique (a) n'est pas une surprise. En outre, des creux en forme de  $V$  peuvent être facilement repérés en fin d'année. Le graphique (b) révèle que, du lundi au vendredi, la consommation horaire médiane s'établit entre 45 et 50 GWh avant 05h00, puis augmente jusqu'à 70 GWh jusqu'à l'heure du déjeuner, pour finalement diminuer progressivement au cours de l'après-midi et de manière plus importante après 20h00. Une tendance similaire peut être observée les samedis et dimanches, bien que le profil global soit plus faible en niveau et sensiblement plus plat. Le graphique (c) confirme l'existence de dynamiques infra-quotidiennes et infra-hebdomadaires, car le périodogramme de la série ELCON rendue journalière et différenciée présente des pics aux fréquences fondamentales correspondantes (mises en évidence par des lignes verticales grises) et à leurs harmoniques.

### 7.2.2 Pré-ajustement

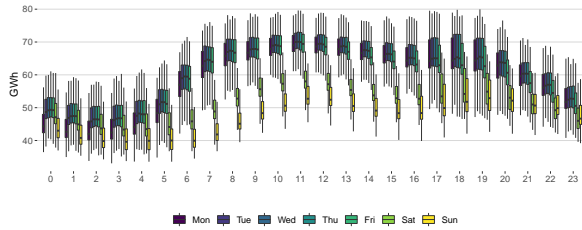
Des études empiriques préliminaires non discutées ici n'ont pas fourni de preuves suffisantes en faveur de l'hypothèse d'effets de calendrier horaires constants au sein de chaque jour dans la série ELCON horaire (régularisée et en logarithme). Pour tenir compte des effets variables du calendrier horaire, nous pourrions utiliser les modèles (2) et (4) avec un ensemble approprié de variables de régression horaires muettes et  $\mathbb{S} = \{24; 168; 8765, 82\}$ . Toutefois, le premier ensemble serait d'une taille presque prohibitive puisque près de 400 variables seraient nécessaires pour tenir compte des seuls effets de calendrier. Par conséquent, une telle approche serait inefficace sur le plan informatique et conduirait très probablement à des estimations instables. C'est pourquoi nous adoptons plutôt une

<sup>6</sup>Les données sont disponibles gratuitement auprès de l'Agence fédérale des réseaux ("Bundesnetzagentur | SMARD.de") sous l'URL <https://www.smard.de/en> et ont été téléchargées le 4 juillet 2022.

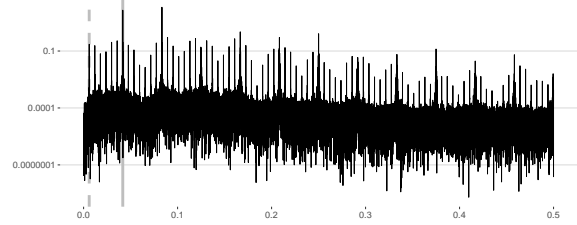
(a) Données brutes



(b) Composantes infra-quotidienne et infra-hebdomadaire : boxplots groupés



(c) Composantes infra-quotidienne et infra-hebdomadaire : périodogramme (en décibels)



**Figure 8** : Profil saisonnier de la série ELCON horaire. Les lignes verticales grises sur le graphique (c) indiquent les fréquences horaires fondamentales hour-of-the-day (*trait plein*) et hour-of-the-week (*pointillés*).

approche de régression périodique, conformément à plusieurs autres études sur séries horaires (voir Cancelo, Espasa et Gafe, 2008 ; Ramanathan et al., 1997). Pour cela, nous créons 24 sous-séries quotidiennes  $\{y_t^{(h)}\}$ , où la  $h$ -ième sous-série représente la consommation d'électricité (en logarithme) enregistrée pendant l'heure  $h \in \{0, \dots, 23\}$  au jour  $t$ , et utilisons pour chacune d'entre elles dans l'équation (2) les mêmes variables de régression modélisant les effets de calendrier quotidiens et les mêmes procédures de détection automatique des valeurs aberrantes déjà appliquées dans une étude antérieure de la série quotidienne ELCON (Webel, 2022), à l'exception des variables relatives à l'heure d'été <sup>7</sup>.

Les résidus de chaque sous-série quotidienne sont modélisés avec l'équation (4) et  $\mathbb{S} = \{7; 365, 2425\}$ , de sorte que l'ensemble du modèle saisonnier de pré-ajustement est donné par

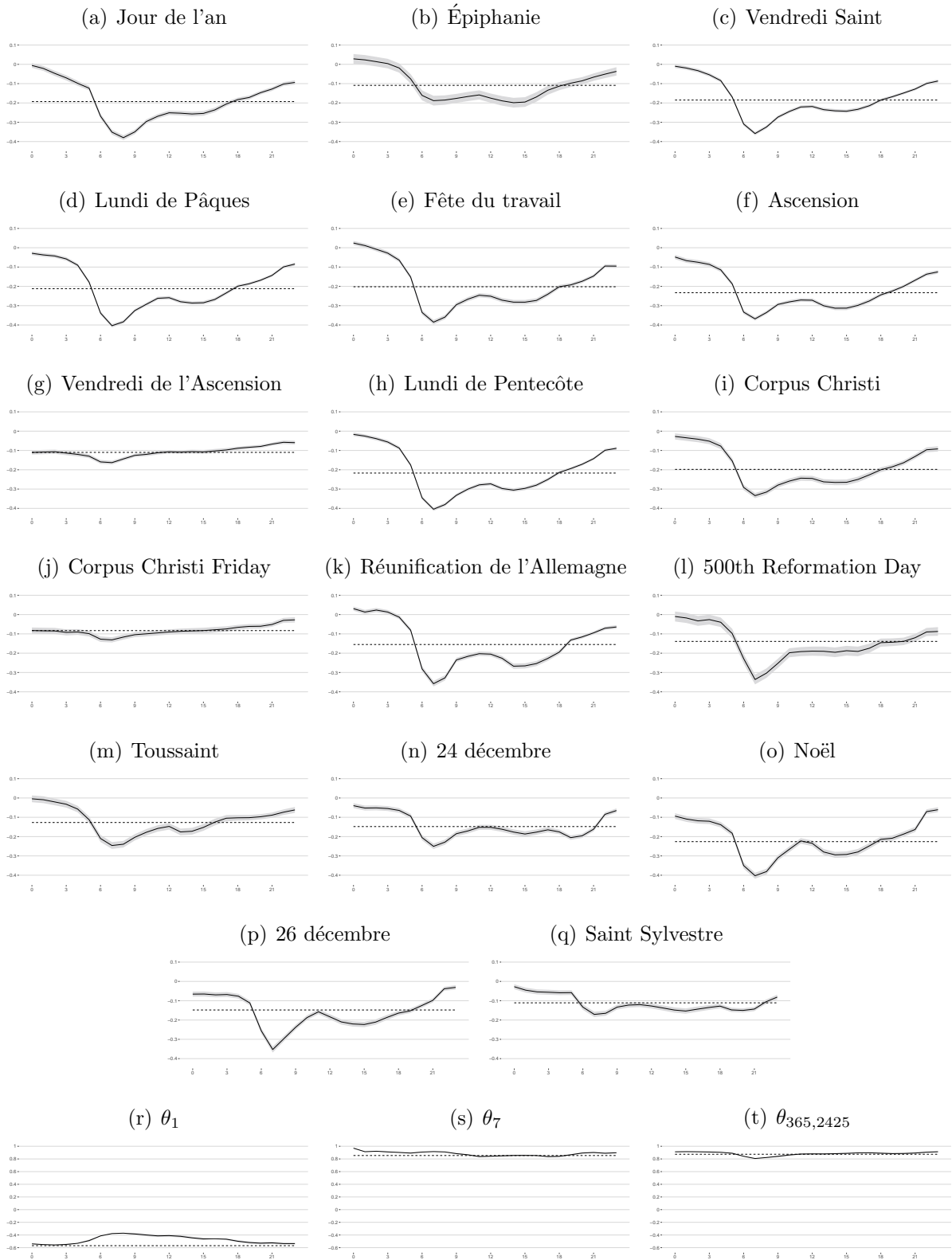
$$\delta(B) w_t^{(h)} = \left(1 - \theta_1^{(h)} B\right) \left(1 - \theta_7^{(h)} B^7\right) \left(1 - \theta_{365,2425}^{(h)} B^{365,2425}\right) \varepsilon_t^{(h)}, \quad (21)$$

où  $\delta(B)$  a été défini dans (20) et  $\{w_t^{(h)}\}$  la sous série numéro  $h$  linéarisée définie par

<sup>7</sup>Comme dans Webel (2022), les variables de régression pour les jours fériés mobiles ont été centrées en enlevant leur moyenne de long-terme, afin d'éviter toute confusion avec la saisonnalité annuelle.

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

Il convient de noter que dans (21)–(22), la saisonnalité infra-quotidienne est implicitement prise en compte par la variation des paramètres saisonniers dans les 24 sous-modèles quotidiens. Les effets de calendrier variables dans le temps estimés et les paramètres MA sont présentés sur la [figure 9](#) avec leurs écarts-type (les résultats détaillés sont disponibles auprès des auteurs sur demande). Les effets de calendrier estimés présentent des profils infra-quotidiens similaires qui, dans la plupart des cas, s'écartent visiblement des effets quotidiens constants estimés sur la série journalière ELCON, correspondant aux données horaires agrégées. Les effets les plus marqués sont observés entre 06h00 et 09h00 et - dans une moindre mesure - vers 15h00, tandis que, sans surprise, les effets les plus faibles sont enregistrés dans la soirée et pendant la nuit. Les paramètres MA estimés restent relativement constants tout au long de la journée. En particulier, les estimations des paramètres saisonniers sont supérieures à 0,8 pour chaque heure de la journée, ce qui indique la présence de fortes dynamiques infra-hebdomadaire (DOW) et infra-annuelles (DOY) dans les 24 sous-séries quotidiennes. Les versions linéarisées de ces dernières sont ensuite remises à l'échelle ELCON d'origine, c'est-à-dire horaire, puis agrégées pour former la série ELCON horaire pré-nettoyée, qui sera désignée par `elcon` dans les sections suivantes.



**Figure 9** : Effets de calendrier estimés et paramètres MA du modèle de pré-ajustement saisonnier (21) et (22). Les zones grisée délimitent les intervalles de confiance  $\pm 1$  écart-type. Les lignes horizontales en pointillés indiquent la valeur des paramètres estimés pour la série ELCON auxiliaire (voir Webel (2022)).

### 7.2.3 Désaisonnalisation

Sur la base de preuves visuelles ([figure 8](#)), nous fixons  $\mathbb{S} = \{24; 168; 8765, 82\}$  pour les séries horaires ELCON pré-nettoyées et nous montrons comment les composantes saisonnières correspondantes peuvent être extraites itérativement avec les approches X-11 et STL étendues. On supposera une décomposition multiplicative en composantes inobservables.

Les estimations X-11 peuvent être obtenues avec la fonction `x11()`, dont un aperçu est donné dans le [tableau 6](#) en [Annexes](#). En commençant à nouveau par la plus petite périodicité, c'est-à-dire  $\tau_1 = 24$ , la commande suivante permet d'obtenir une décomposition X-11 étendue avec un filtre tendance-cycle de Henderson cubique à 25 termes, des filtres saisonniers à  $3 \times 9$  et des  $\sigma$ -limites par défaut sur la série ELCON pré-nettoyée (en logarithme) :

```
R> elcon.x11.idp <- rjd3x11plus::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)
```

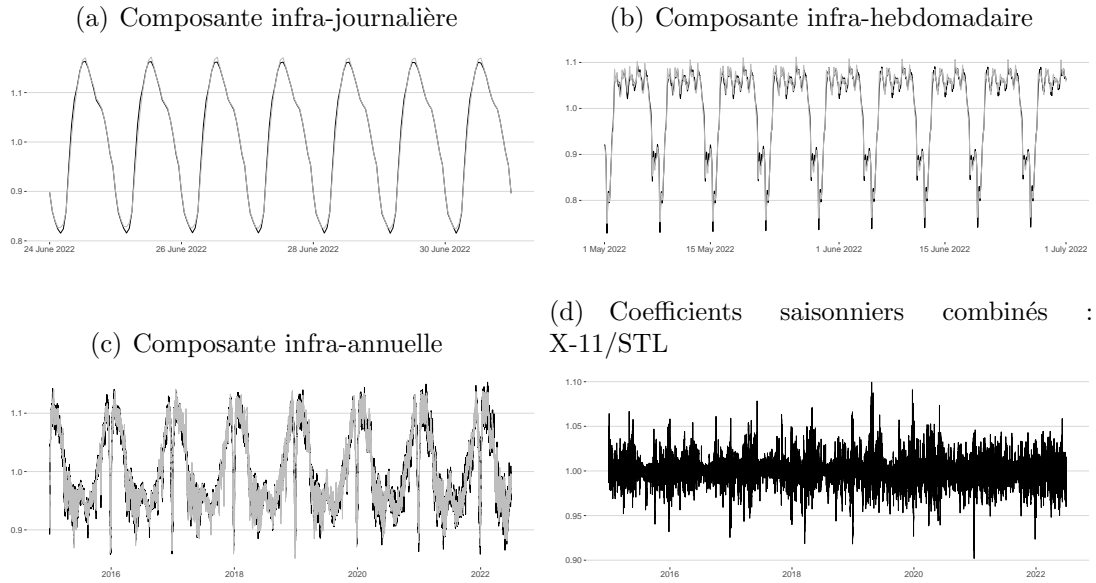
Les composantes infra-hebdomadaires et infra-annuelles peuvent être estimées de la même manière. En ciblant  $\tau_2 = 168$ , par exemple, la commande suivante permet d'obtenir une décomposition X-11 étendue sur la série ELCON privée de la composante infra-quotidienne estimée (avec une spécification inchangée des arguments omis) :

```
R> elcon.x11.iwp <- rjd3x11plus::x11(  
+       elcon.x11.idp$decomposition$sa,  
+       period = 168, trend.horizon = 84, ...)
```

Les trois composantes saisonnières – normalisées pour être centrées à 1 – sont tracées en noir sur la [figure 10 \(a\)–\(c\)](#).

Les estimations STL peuvent être obtenues avec la fonction `stl()`, voir [tableau 7](#) en [Annexes](#). En ciblant à nouveau  $\tau_1 = 24$  en premier lieu, la commande suivante applique l'algorithme STL sur la série ELCON pré-nettoyée sans poids de robustesse, un lisseur LOESS à 25 termes est utilisé pour l'extraction de la tendance-cycle et un lisseur LOESS à 11 termes pour l'extraction de la composante saisonnière, de sorte que les longueurs sont identiques aux filtres X-11 correspondants :

```
R> elcon.stl.idp <- rjd3stl::stl(  
+       log(elcon), multiplicative = FALSE,  
+       period = 24,  
+       swindow = 11, twindow = 25,  
+       robust = FALSE)
```



**Figure 10** : Coefficients saisonniers estimés pour la série ELCON horaire obtenus avec X-11 étendu (*noir*) et STL (*gris*).

Notons que la fonction `stl()` a moins d'arguments que la fonction `x11()`, bien que les deux méthodes suivent à peu près les mêmes étapes itératives. La raison en est que les poids des lisseurs LOESS symétriques et asymétriques pour la tendance-cycle et la composante saisonnière, ainsi que les poids robustes optionnels pour l'irrégulier, sont dérivés directement par des noyaux pré-spécifiés, alors que dans X-11, les noyaux pour l'extraction de la tendance-cycle, la dérivation des variantes asymétriques, les filtres saisonniers et les  $\sigma$ -limites pour la détection des valeurs extrêmes peuvent être sélectionnés par l'utilisateur à partir d'une multitude d'options pré-définies.

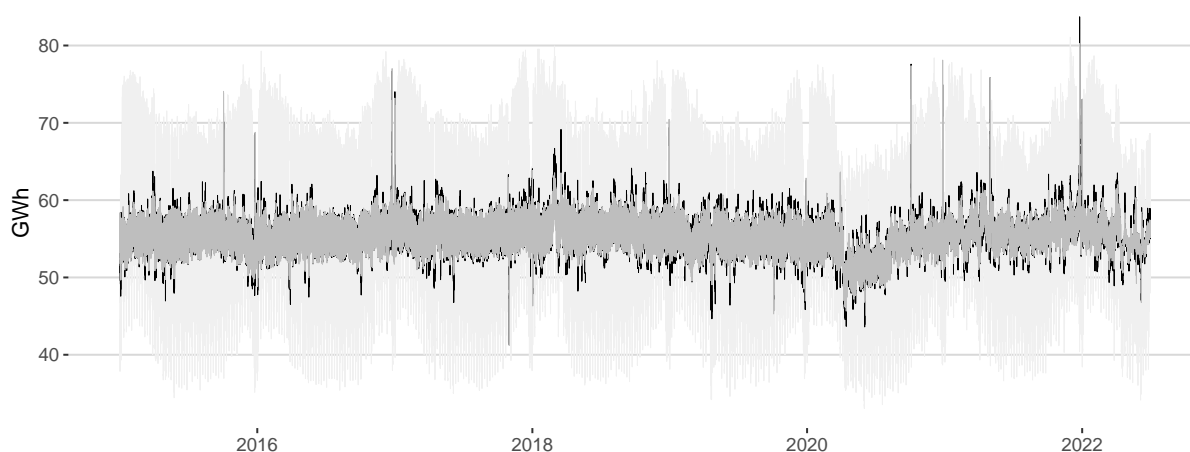
Les estimations STL des composantes infra-hebdomadaires et infra-annuelles peuvent être obtenues de la même manière en modifiant les arguments appropriés. Par exemple, l'argument `period` est à nouveau dicté par  $\tau \in \mathbb{S}$  et nous pourrions fixer `twindow` à  $\lceil \tau \rceil^{\text{odd}}$  et `swindow` à la longueur des filtres saisonniers  $3 \times k$  sélectionnés dans X-11, qui étaient les  $3 \times 9$  et  $3 \times 3$  pour les composantes infra-hebdomadaires et infra-annuelles, respectivement. Les facteurs saisonniers STL correspondants, qui sont représentés par des lignes grises sur la [figure 10 \(a\)–\(c\)](#), sont assez similaires à ceux de X-11, bien que les facteurs d'ajustement combinés diffèrent parfois de 10 % en valeur absolue. Ces différences importantes mais rares peuvent également être observées entre deux versions de la série horaire ELCON désaisonnalisée ([figure 11](#)).

Enfin, il convient de noter qu'une application répétée des lisseurs LOESS à la place des filtres linéaires avec des poids prédéfinis doit généralement dégrader le temps de calcul, en particulier pour les fenêtres de régression plus longues. Nous avons reproduit les extractions X-11 et STL des trois composantes saisonnières avec les spécifications ci-dessus un total de 25 fois - avec ou sans poids robustes dans STL - et nous indiquons les temps de calcul moyens dans le [tableau 2](#). X-11 est plus rapide en général, sa marge par rapport à STL est plutôt faible lors de l'extraction des composantes infra-quotidiennes et infra-hebdomadaires sans poids robustes, mais devient substantielle dans tous les autres

**Table 2 :** Temps moyen de calcul pour l'extraction des composantes saisonnières (25 répétitions avec un OS Windows 64-bit doté d'un processeur Intel Core i3-8100 CPU @ 3.60 GHz et 8.00 GB de RAM)

Méthode d'ajustement saisonnier		Composante Infra-quotidienne	Composante Infra-hebdomadaire	Composante Infra-annuelle
X-11 étendu	avec $\sigma$ -limites par défaut	0,19 sec	0,75 sec	34,42 sec
STL étendu	sans poids robustes	0,24 sec	1,30 sec	87,37 sec
	avec poids robustes	1,72 sec	9,41 sec	681,57 sec

cas. Il faut garder à l'esprit que le nombre total d'exécutions dans les boucles internes et externes augmente d'un facteur de 7,5 lorsque des poids robustes sont utilisés.



**Figure 11 :** Série horaire ELCON désaisonnalisée avec X-11 étendu (*noir*) et STL étendu (*gris*). La série brute correspond à la ligne en gris clair.

### 7.3 Demandes hebdomadaires initiales d'assurance chômage aux États-Unis

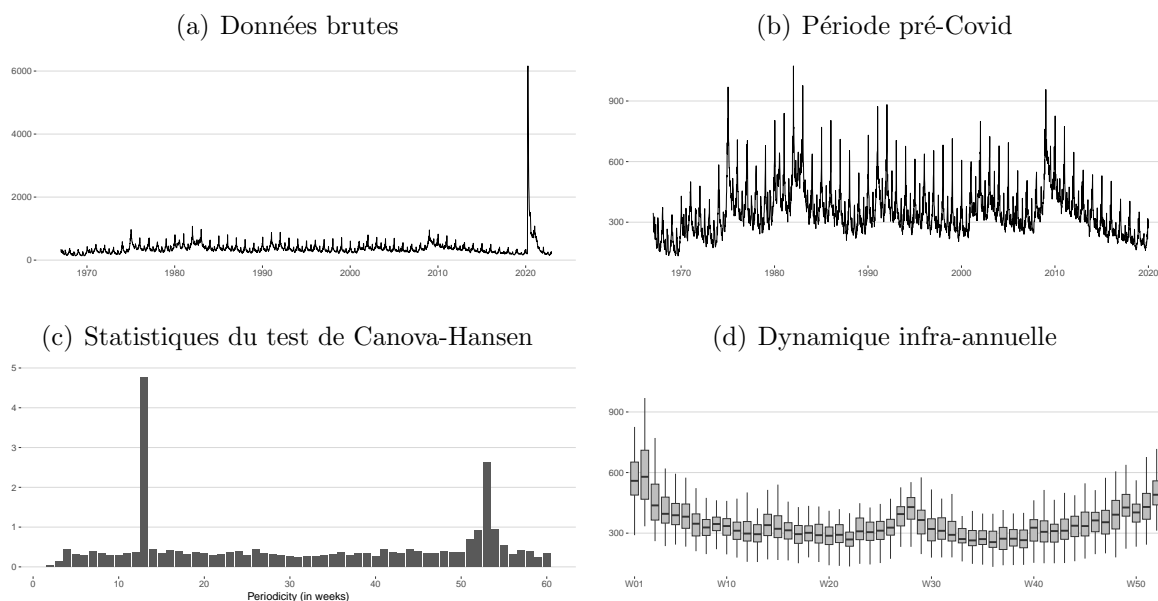
Nous allons maintenant illustrer comment les effets du pré-ajustement et les facteurs saisonniers peuvent être estimés conjointement dans le cadre des modèles structurels. Pour cela, nous examinons les demandes initiales hebdomadaires d'assurance chômage (CLAIMS) aux États-Unis de 1967 (semaine 01) à 2022 (semaine 48), soit 2 918 observations.<sup>8</sup> Des versions plus courtes de cette série ont également été analysées dans des études antérieures sur les données hebdomadaires (voir par exemple Cleveland et Scott (2007), Cleveland, Evans et Scott (2018), Proietti et Pedregal (2022)).

<sup>8</sup>Les données sont disponibles gratuitement auprès de l'U.S. Department of Labor's Employment and Training Administration sous l'URL <https://oui.doleta.gov/unemploy/claims.asp> et ont été téléchargées le 29 décembre 2022. Les chiffres hebdomadaires reflètent l'activité économique du dimanche au samedi.

### 7.3.1 Profil saisonnier

La série CLAIMS oscille entre 200 000 et 900 000 demandes, leur nombre est périodiquement plus élevé pendant les mois les plus froids (de novembre à février). Elle présente également quelques augmentations brutales associées à des périodes économiques tumultueuses, telles que la crise financière (du troisième trimestre 2008 au deuxième trimestre 2009), et des chiffres sans précédent immédiatement après l'apparition de la pandémie de COVID-19 en mars 2020. Ces derniers culminent légèrement au-dessus de 6,1 millions de demandes en 2020 (semaine 14), plafonnent temporairement autour de 1 million de demandes à partir de de 2020 (semaine 31), et reviennent à des niveaux d'avant Covid après une baisse progressive au cours de l'année 2021 (figure 12 (a)–(b)).

Les statistiques du test de Canova-Hansen généralisé présentent des pics à 13 semaines et autour de 53 semaines, ce qui indique la présence de fortes fluctuations liées à l'effet semaine-de-l'année (Week-Of-the-Year, WOY) (figure 12 (c)). Les box-plots hebdomadaires, qui ont été calculés à partir des données d'avant Covid, confirment cette impression (figure 12 (d)). Leur disposition en forme de  $W$  révèle également de légères augmentations en milieu d'année s'expliquant en partie par un changement de pratiques dans l'industrie automobile (Cleveland et Scott, 2007). Dans l'ensemble, le profil saisonnier de la série CLAIMS est caractérisé par une forte composante WOY avec une dynamique dominante de l'effet semaine-dans-le trimestre, de sorte que nous pourrions définir soit  $\mathbb{S} = \{13; 52, 18\}$  soit  $\mathbb{S} = \{52, 18\}$  dans l'équation (4). Pour les analyses ultérieures nous choisissons cette dernière définition, dans un souci de parcimonie.



**Figure 12 :** Profil saisonnier de la série hebdomadaire CLAIMS (en milliers).

### 7.3.2 Modèle structurel et désaisonnalisation

Nous spécifions un modèle structurel additif (Basic Structural Model, BSM) pour la série CLAIMS non transformée<sup>9</sup>. La tendance-cycle est modélisée comme une tendance linéaire locale selon (17)–(18), la composante WOY est modélisée par des termes trigonométriques, et la dynamique liée au calendrier est capturée par un ensemble de 11 variables de régression hebdomadaires (indicatrices), une pour chacun des jours fériés fixes et mobiles suivants : Jour de l'An (1er janvier) – dont l'effet est généralement attribué à la deuxième semaine comme dans Cleveland, Evans et Scott (2018) –, Martin Luther King Day (célébré à partir de 1986), President's Day, Pâques, Memorial Day, Independence Day (4 juillet), Labor Day, Columbus Day, Veterans Day (11 novembre), Thanksgiving et Noël (25 décembre)<sup>10</sup>. En outre, deux Level-Shifts (LS) ont été inclus pour modéliser le caractère atypique de la phase Covid en 2020 (Semaines 12 à 24 et 28 à 32).

Après avoir stocké les variables de régression hebdomadaires dans un objet matriciel nommé `claims.reg`, l'estimation du BSM spécifié - y compris une recherche automatique d'additive outliers (AO) et level-shift (LS) supplémentaires, avec des valeurs critiques calculées en fonction du nombre d'observations - est effectuée avec la commande suivante (voir [tableau 8](#) en [Annexes](#)). Il est à noter que les périodicités non entières sont automatiquement arrondies à l'entier inférieur le plus proche :

```
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")
```

L'objet `claims.bsm` contient l'ensemble des séries, diagnostics et paramètres relatifs au processus d'estimation. Par exemple, les estimations lissées du niveau stochastique, de la pente, de la composante saisonnière WOY trigonométrique et des composantes irrégulières peuvent être récupérées dans cet objet avec la commande suivante :

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

Les composantes déterministes des effets de calendrier et outliers peuvent, quant à elles, être calculées à partir de l'objet `claims.bsm` en utilisant les mêmes commandes que celles illustrées pour la série journalière BIRTHS.

Le [tableau 3](#) présente les effets de calendrier estimés pour la série hebdomadaire CLAIMS. Thanksgiving et le jour de l'An ont les effets les plus importants, tandis que

<sup>9</sup>Un modèle additif pour la série CLAIMS est conforme aux pratiques officielles actuelles de désaisonnalisation du U.S. Bureau of Labor Statistics. En fait, le schéma de décomposition est passé de multiplicatif à additif en réaction à l'épidémie de Covid-19.

<sup>10</sup>À l'exception de Pâques, les dates des jours fériés ont été obtenues à partir du package `{tis}`, qui répertorie les dates auxquelles les jours fériés fédéraux sont effectivement célébrés (c'est-à-dire, dans certains cas, le vendredi précédant ou le lundi suivant le jour férié si ce dernier tombe un samedi ou un dimanche, respectivement). Cela n'affecte toutefois pas la création ultérieure de variables muettes relatives aux jours fériés hebdomadaires.

**Table 3** : Effets de calendrier estimés pour la série hebdomadaire CLAIMS

Jour férié	Coefficient	Ecart-type	$t$ -value
Jour de l'An	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
Pâques	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
Noël	-6,221	4,6013	-1,3520

l'effet de l'Independence Day est à peine significatif au niveau de 5 % et que celui du jour de Noël est même non significatif. La composante calendaire estimée ainsi que la composante WOY trigonométrique lissée sont présentées sur la [figure 13 \(a\)](#).

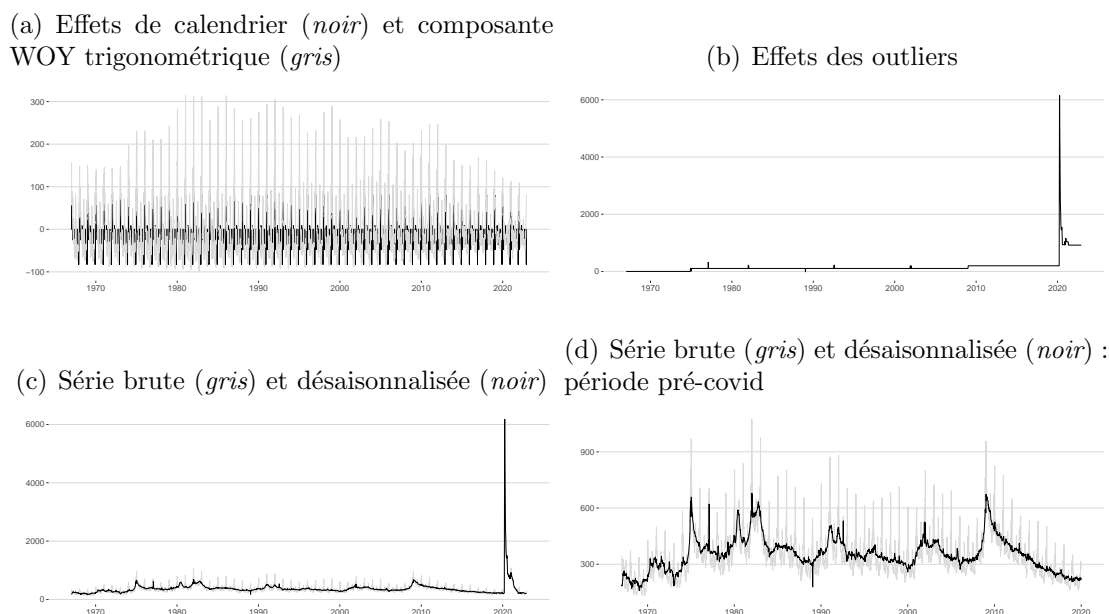
La détection automatique d'outliers a permis d'identifier 11 Additive Outliers (AO) et 4 Level-Shifts (LS) supplémentaires. Sept de ces valeurs aberrantes se situent dans les mois les plus froids de la période antérieure à 1990 et pourraient être liées à des hivers exceptionnellement froids ou chauds. Un des changements de niveau coïncide avec les valeurs élevées de la série CLAIMS pendant la crise financière mondiale (2009, semaine 2), et quatre valeurs aberrantes peuvent être associées à la phase de reprise post-pandémie. La composante liée aux outliers, y compris les deux Level-Shifts pré-spécifiées, est présentée sur la [figure 13 \(b\)](#).

Les ratios  $q$ , c'est-à-dire les variances d'innovation estimées des composantes inobservables (UC) exprimées en fraction de leur valeur la plus élevée, peuvent être extraits de l'objet `claims.bsm` avec la commande suivante

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

pour la composante trigonométrique WOY, et de la même manière pour toute autre composante stochastique. Les innovations du niveau  $\{\xi_t\}$  dans l'équation (17) se sont avérées avoir la plus grande variance estimée, les autres  $q$ -ratios sont de  $6,69 \times 10^{-5}$  pour les innovations de la pente  $\{\chi_t\}$  dans (18), de  $1,75 \times 10^{-3}$  pour les innovations  $\{\omega_t\}$  de la composante WOY dans (19) et de 0,206 pour la composante irrégulière (bruit blanc) dans (1).

La suppression des composantes de calendrier et saisonnière (WOY trigonométrique) réduit sensiblement la volatilité de la série hebdomadaire CLAIMS ([figure 13 \(c\)–\(d\)](#)). Ainsi, les données corrigées des variations saisonnières donnent une image beaucoup plus claire des mouvements sous-jacents de la série. En outre, elles révèlent certaines variations qui étaient complètement occultées dans les données non corrigées. Un bon exemple est l'augmentation temporaire à environ 370 000 des demandes d'indemnisation désaisonnalisées en moyenne au cours de l'année 2005 (semaines 35 à 40), c'est-à-dire une augmentation d'environ 10 % par rapport à 2005 semaine 34 suite à l'ouragan Katrina



**Figure 13 :** Estimations lissées pour les différentes composantes inobservables de la série hebdomadaire CLAIMS (en milliers).

qui a frappé la côte américaine du Golfe du Mexique le 29 août 2005 (voir également la discussion dans Cleveland, Evans et Scott (2018)).

## 8 Résumé

Nous avons donné une description détaillée des algorithmes de JDemetra+ 3.0 pour la modélisation et la désaisonnalisation des séries temporelles infra-mensuelles. En particulier, nous avons mis en évidence les principales modifications à apporter aux approches de pré-ajustement et de décomposition couramment utilisées dans la statistique publique pour prendre en compte les particularités de ces données. L'estimation des valeurs aberrantes et des effets de calendrier peut être traitée par une régression dans laquelle les perturbations suivent une extension du modèle Airline classique, ce qui permet de prendre en compte des composantes saisonnières multiples ainsi que des puissances non entières de l'opérateur retard. Ces dernières sont approximées par un développement de Taylor du premier ordre, qui est une moyenne pondérée des deux puissances adjacentes à valeur entière. Ce même modèle constitue la base d'une approche étendue de décomposition basée sur une modélisation Arima et fondée sur la théorie classique de Wiener-Kolmogorov. Le principe de la moyenne pondérée est également adopté dans une approche étendue de la décomposition X-11 afin de maintenir l'applicabilité des filtres saisonniers classiques de  $3 \times k$  aux séries présentant des périodicités non entières. Cette dernière méthode contient également un riche ensemble de filtres d'extraction de la tendance-cycle dérivés de régressions polynomiales locales qui englobent les filtres originaux de Henderson et les substituts de Musgrave dans X-11. Il est également possible de procéder à un autre type de décomposition empirique en utilisant des lisseurs LOESS pour extraire la tendance-cycle et les composantes saisonnières, dont la mise en œuvre correspond à la méthode STL

classique, à l'exception de quelques modifications mineures. La désaisonnalisation en deux étapes (régression préalable et extraction des facteurs saisonniers) est donc une stratégie courante pour ces méthodes de désaisonnalisation étendues. Cependant, une estimation simultanée des effets de calendrier et de la composante saisonnière, avec détection automatique des valeurs aberrantes, peut également être réalisée au sein de l'écosystème R de JDemetra+ à l'aide de modèles structurels. Nous avons également illustré certaines caractéristiques de ces méthodes en utilisant près de 20 000 observations pour le nombre quotidien de naissances en France, plus de 65 000 observations pour la consommation horaire d'électricité en Allemagne et près de 3 000 observations pour les demandes initiales hebdomadaires d'assurance chômage aux États-Unis.

La boîte à outils méthodologique proposée dans l'écosystème R est déjà très riche, mais elle peut encore être améliorée. Actuellement, les prévisions basées sur la modélisation des séries ne peuvent être obtenues qu'indirectement par l'agrégation des estimations des composantes inobservables prédites fournies lors de la phase de décomposition sur des modèles étendus, mais pas directement à partir des modèles de pré-ajustement décrits dans les équations (2) et (4). Cependant, nous ne doutons pas que de nombreux utilisateurs apprécieraient cette dernière solution. Des procédures automatiques pour l'étalonnage des modèles, tels que les tests de log/niveau et de saisonnalité ainsi que des diagnostics de qualité adaptés, constitueraient un autre ajout utile. Les recherches futures pourraient également accorder une certaine attention à la sélection, basée sur les données, des filtres d'extraction de la tendance-cycle et des composantes saisonnières lors de la désaisonnalisation avec X-11 et STL étendus. De plus, les opérations de lissage par période inhérentes à ces deux méthodes supposent implicitement des effets saisonniers différents pour chaque point dans le temps et pour chaque périodicité saisonnière : cela pourrait être remis en question, en particulier pour des points adjacents ou proches. En outre, les modèles saturés, tels que (19), pourraient être sur-paramétrés et non estimables sur le plan informatique pour les périodicités saisonnières les plus longues. Pour ces raisons, les modèles épars basés sur des composantes de base fonctionnelles semblent être des alternatives intéressantes, mais nécessitent très probablement des investigations méthodologiques supplémentaires. Par exemple, une modélisation automatique dans un cadre espace-état avec une saisonnalité capturée par des splines cubiques périodiques (Harvey et Koopman, 1993 ; Harvey, Koopman et Riani, 1997) nécessiterait une détermination fondée sur les données à la fois du nombre et des positions des nœuds de la spline pour chaque composante saisonnière (une stratégie de sélection basée sur LOESS est discutée dans Proietti et 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 Initial 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.

# Annexes

Les méthodes décrites et illustrées dans le présent document sont accessibles *via* un écosystème de packages R disponibles sur GitHub à l'adresse <https://github.com/rjdverse>. Nous documentons ici les principales fonctions qui ont été utilisées dans les applications (section 7).

## A Le package {rjd3highfreq}

Ce package permet l'estimation du modèle de pré-ajustement (2) et (4), avec la fonction `fractionalAirlineEstimation()` et la décomposition avec l'approche AMB étendue, avec la fonction `fractionalAirlineDecomposition()`, voir tableaux ci-dessous.

**Table 4 :** The `fractionalAirlineEstimation()` function

Argument	Explanation
<code>y</code>	Observed time series $\{y_t\}$ in (1)
<code>criticalValue</code>	Critical value for automatic outlier detection in (2) <sup>1)</sup>
<code>mean</code> (Boolean)	Add constant mean (after differencing) to $\mathbf{x}_t$ in (2)
<code>outliers</code>	Types for automatic outlier detection in $\mathbf{x}_t$ in (2)
<code>x</code>	User-defined regression variables in $\mathbf{x}_t$ in (2)
<code>ar</code> (Boolean)	Replace $\delta_1(B)$ and $\theta_1(B)$ with a stationary non-seasonal AR(1) operator in (4) <sup>2)</sup>
<code>ndiff</code>	Total order of non-seasonal differencing in (4) <sup>3)</sup>
<code>periods</code>	Set $\mathbb{S}$ of seasonal periodicities in (4)
<code>approximateHessian</code> (Boolean)	Use fast computation of the Hessian in Algorithm 1 <sup>4)</sup>
<code>precision</code>	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
<code>y</code>	Input time series
<code>period</code>	Seasonal periodicity $\tau \in \mathbb{S}$
<code>sn</code> (Boolean)	Decompose data into seasonally adjusted data (signal) and seasonal component (noise)
<code>stde</code> (Boolean)	Calculate standard deviations of component estimates
<code>nbcasts</code>	Backcast horizon
<code>nfcasts</code>	Forecast horizon

## B Le package `{rjd3x11plus}`

Ce package permet une décomposition avec l'approche X-11 étendue en utilisant la fonction `x11()`, voir tableau ci-dessous.

**Table 6 :** The `x11()` function

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

## C Le package `{rjd3stl}`

Ce package permet une décomposition avec l'approche STL étendue en utilisant la fonction `stl()`, voir tableau ci-dessous.

**Table 7 :** The `stl()` function

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

<sup>1</sup> Setting `twindow = 0` calls the automatic specification (16) (see [section 5](#)).

## D Le package {rjd3sts}

Ce package contient les fonctions permettant la spécification et l'estimation de modèles structurels pour les séries temporelles. Un Basic Structural Model (BSM) avec détection automatique des valeurs aberrantes peut être spécifié avec la fonction `sts.outliers()`, voir tableau ci-dessous.

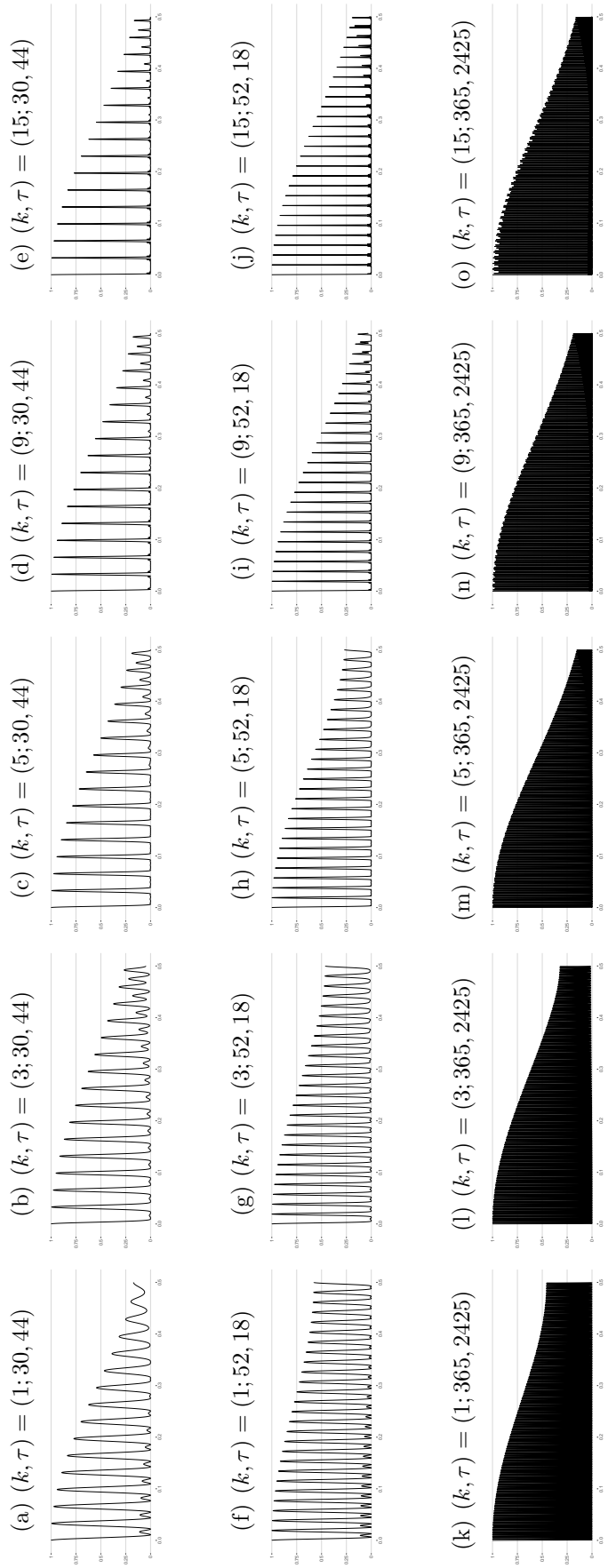
**Table 8 :** The `sts.outliers()` function

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

**1** The trading-day contrasts are internally calculated with the `rjd3toolkit::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".

## E Gain au carré des filtres saisonniers X-11 pour les périodicités non-entières les plus courantes

La [figure 14](#) montre le gain au carré de tous les filtres saisonniers  $3 \times k$  utilisés pour extraire de données haute-fréquence les composantes saisonnières les plus courantes à périodicités non-entières. Ces composantes sont l'effet du jour du mois (day-of-the-month pattern) et du jour de l'année (day-of-the-year pattern) susceptibles d'être rencontrés dans les séries journalières, ainsi que l'effet de la semaine de l'année (week-of-the-year pattern) pouvant être observé dans les séries hebdomadaires. Les calculs sous-jacents sont exposés dans la note de bas de page [3](#), et les poids des filtres saisonniers symétriques  $3 \times 3$  sont explicitement donnés dans l'équation [\(14\)](#).



**Figure 14 :** Carré du gain des filtres saisonniers symétriques  $3 \times k$  pour  $k \in \{1; 3; 5; 9; 15\}$  et  $\tau \in \{30, 44; 52, 18; 365, 2425\}$ .

## Série des Documents de Travail « Méthodologie Statistique »

- 9601** : Une méthode synthétique, robuste et efficace pour réaliser des estimations locales de population.  
**G. DECAUDIN, J.-C. LABAT**
- 9602** : Estimation de la précision d'un solde dans les enquêtes de conjoncture auprès des entreprises.  
**N. CARON, P. RAVALET, O. SAUTORY**
- 9603** : La procédure **FREQ** de **SAS** - Tests d'indépendance et mesures d'association dans un tableau de contingence.  
**J. CONFAIS, Y. GRELET, M. LE GUEN**
- 9604** : Les principales techniques de correction de la non-réponse et les modèles associés.  
**N. CARON**
- 9605** : L'estimation du taux d'évolution des dépenses d'équipement dans l'enquête de conjoncture : analyse et voies d'amélioration.  
**P. RAVALET**
- 9606** : L'économétrie et l'étude des comportements. Présentation et mise en œuvre de modèles de régression qualitatifs. Les modèles univariés à résidus logistiques ou normaux (**LOGIT**, **PROBIT**).  
**S. LOLLIVIER, M. MARPSAT, D. VERGER**
- 9607** : Enquêtes régionales sur les déplacements des ménages : l'expérience de Rhône-Alpes.  
**N. CARON, D. LE BLANC**
- 9701** : Une bonne petite enquête vaut-elle mieux qu'un mauvais recensement ?  
**J.-C. DEVILLE**
- 9702** : Modèles univariés et modèles de durée sur données individuelles.  
**S. LOLLIVIER**
- 9703** : Comparaison de deux estimateurs par le ratio stratifiés et application aux enquêtes auprès des entreprises.  
**N. CARON, J.-C. DEVILLE**
- 9704** : La faisabilité d'une enquête auprès des ménages.  
1. au mois d'août.  
2. à un rythme hebdomadaire  
**C. LAGARENNE, C. THIESSET**
- 9705** : Méthodologie de l'enquête sur les déplacements dans l'agglomération toulousaine.  
**P. GIRARD**
- 9801** : Les logiciels de désaisonnalisation **TRAMO & SEATS** : philosophie, principes et mise en œuvre sous **SAS**.  
**K. ATTAL-TOUBERT, D. LADIRAY**
- 9802** : Estimation de variance pour des statistiques complexes : technique des résidus et de linéarisation.  
**J.-C. DEVILLE**
- 9803** : Pour essayer d'en finir avec l'individu Kish.  
**J.-C. DEVILLE**
- 9804** : Une nouvelle (encore une !) méthode de tirage à probabilités inégales.  
**J.-C. DEVILLE**
- 9805** : Variance et estimation de variance en cas d'erreurs de mesure non corrélées ou de l'intrusion d'un individu Kish.  
**J.-C. DEVILLE**
- 9806** : Estimation de précision de données issues d'enquêtes : document méthodologique sur le logiciel **POULPE**.  
**N. CARON, J.-C. DEVILLE, O. SAUTORY**
- 9807** : Estimation de données régionales à l'aide de techniques d'analyse multidimensionnelle.  
**K. ATTAL-TOUBERT, O. SAUTORY**
- 9808** : Matrices de mobilité et calcul de la précision associée.  
**N. CARON, C. CHAMBAZ**
- 9809** : Échantillonnage et stratification : une étude empirique des gains de précision.  
**J. LE GUENNEC**
- 9810** : Le Kish : les problèmes de réalisation du tirage et de son extrapolation.  
**C. BERTHIER, N. CARON, B. NEROS**
- 9901** : Perte de précision liée au tirage d'un ou plusieurs individus Kish.  
**N. CARON**
- 9902** : Estimation de variance en présence de données imputées : un exemple à partir de l'enquête Panel Européen.  
**N. CARON**
- 0001** : L'économétrie et l'étude des comportements. Présentation et mise en œuvre de modèles de régression qualitatifs. Les modèles univariés à résidus logistiques ou normaux (**LOGIT**, **PROBIT**) (version actualisée).  
**S. LOLLIVIER, M. MARPSAT, D. VERGER**
- 0002** : Modèles structurels et variables explicatives endogènes.  
**J.-M. ROBIN**
- 0003** : L'enquête 1997-1998 sur le devenir des personnes sorties du RMI - Une présentation de son déroulement.  
**D. ENEAU, D. GUILLEMOT**
- 0004** : Plus d'amis, plus proches ? Essai de comparaison de deux enquêtes peu comparables.  
**O. GODECHOT**
- 0005** : Estimation dans les enquêtes répétées : application à l'Enquête Emploi en Continu.  
**N. CARON, P. RAVALET**
- 0006** : Non-parametric approach to the cost-of-living index.  
**F. MAGNIEN, J. POUGNARD**
- 0101** : Diverses macros **SAS** : Analyse exploratoire des données, Analyse des séries temporelles.  
**D. LADIRAY**
- 0102** : Économétrie linéaire des panels : une introduction.  
**T. MAGNAC**
- 0201** : Application des méthodes de calages à l'enquête EAE-Commerce.  
**N. CARON**
- C 0201** : Comportement face au risque et à l'avenir et accumulation patrimoniale - Bilan d'une expérimentation.  
**L. ARRONDEL, A. MASSON, D. VERGER**
- C 0202** : Enquête Méthodologique Information et Vie Quotidienne - Tome 1 : bilan du test 1, novembre 2002.  
**J.-A. VALLET, G. BONNET, J.-C. EMIN, J. LEVASSEUR, T. ROCHER, P. VRIGNAUD, X. D'HAULTFOEUILLE, F. MURAT, D. VERGER, P. ZAMORA**
- 0203** : General principles for data editing in business surveys and how to optimise it.  
**P. RIVIERE**
- 0301** : Les modèles logit polytomiques non ordonnés : théories et applications.  
**C. AFSA ESSAFI**
- 0401** : Enquête sur le patrimoine des ménages - Synthèse des entretiens monographiques.  
**V. COHEN, C. DEMMER**
- 0402** : La macro **SAS** **CUBE** d'échantillonnage équilibré  
**S. ROUSSEAU, F. TARDIEU**
- 0501** : Correction de la non-réponse et calage de l'enquêtes Santé 2002  
**N. CARON, S. ROUSSEAU**

**0502** : Correction de la non-réponse par ré pondération et par imputation  
**N. CARON**

**0503** : Introduction à la pratique des indices statistiques - notes de cours  
**J-P BERTHIER**

**0601** : La difficile mesure des pratiques dans le domaine du sport et de la culture - bilan d'une opération méthodologique  
**C. LANDRE, D. VERGER**

**0801** : Rapport du groupe de réflexion sur la qualité des enquêtes auprès des ménages  
**D. VERGER**

**M2013/01** : La régression quantile en pratique  
**P. GIVORD, X. D'HAULTFOEUILLE**

**M2014/01** : La microsimulation dynamique : principes généraux et exemples en langage R  
**D. BLANCHET**

**M2015/01** : la collecte multimode et le paradigme de l'erreur d'enquête totale  
**T. RAZAFINDROVONA**

**M2015/02** : Les méthodes de Pseudo-Panel  
**M. GUILLERM**

**M2015/03** : Les méthodes d'estimation de la précision pour les enquêtes ménages de l'Insee tirées dans Octopusse  
**E. GROS, K. MOUSSALAM**

**M2016/01** : Le modèle Logit Théorie et application.  
**C. AFSA**

**M2016/02** : Les méthodes d'estimation de la précision

de l'Enquête Emploi en Continu  
**E. GROS, K. MOUSSALAM**

**M2016/03** : Exploitation de l'enquête expérimentale Vols, violence et sécurité.  
**T. RAZAFINDROVONA**

**M2016/04** : Savoir compter, savoir coder. Bonnes pratiques du statisticien en programmation.  
**E. L' HOUR, R. LE SAOUT, B. ROUPPERT**


**M2016/05** : Les modèles multiniveaux  
**P. GIVORD, M. GUILLERM**


**M2016/06** : Econométrie spatiale : une introduction pratique  
**P. GIVORD, R. LE SAOUT**

**M2016/07** : La gestion de la confidentialité pour les données individuelles  
**M. BERGEAT**

**M2016/08** : Exploitation de l'enquête expérimentale Logement internet-papier  
**T. RAZAFINDROVONA**

**M2017/01**: Exploitation de l'enquête expérimentale Qualité de vie au travail  
**T. RAZAFINDROVONA**

**M2018/01**: Estimation avec le score de propension sous   
**S. QUANTIN**

**M2018/02**: Modèles semi-paramétriques de survie en temps continu sous   
**S. QUANTIN**

**M2019/01** : Les méthodes de décomposition appliquées à l'analyse des inégalités

**B. BOUTCHENIK, E. COUDIN, S. MAILLARD**

**M2020/01** : L'économétrie en grande dimension  
**J. L' HOUR**

**M2021/01** : R Tools for JDemetra+ - Seasonal adjustment made easier  
**A. SMYK, A. TCHANG**

**M2021/02** : Le traitement du biais de sélection endogène dans les enquêtes auprès des ménages par modèle de Heckman  
**L. CASTELL, P. SILLARD**

**M2021/03** : Conception de questionnaires auto-administrés  
**H. KOUMARIANOS, A. SCHREIBER**

**M2022/01** : Introduction à la géomatique pour le statisticien : quelques concepts et outils innovants de gestion, traitement et diffusion de l'information spatiale  
**F. SEMECURBE, E. COUDIN**

**M2022/02** : Le zonage en unites urbaines 2020  
**V. COSTEMALLE, S. OUJIA, C. GUILLO, A. CHAUVET**

**M2023/01** : Les réseaux de neurones appliqués à la statistique publique : méthodes et cas d'usages  
**D. BABET, Q. DELTOUR, T. FARIA, S. HIMPENS**

**M2023/02** : Redressements de la première vague de l'enquête epicov : un

exemple de correction des effets de sélection dans les enquêtes multimodes  
**L. CASTELL, C. FAVRE-MARTINOZ, N. PALIOD, P. SILLARD**

**M2023/03** : Appariements de données individuelles : concepts, méthodes, conseils  
**L. MALHERBE**

**M2023/04** : Victimations déclarées et effets de mode : enseignements de l'expérimentation panel multimode de l'enquête cadre de vie et sécurité  
**L. CASTELL, M. CLERC, D. CROZE, S. LEGLEYE, A. NOUGARET**

**M2024/01** : Estimation en temps réel de la tendance-cycle : apport de l'utilisation des filtres asymétriques dans la détection des points de retournement  
**A. QUARTIER-LA-TENTE**

**M2024/02** : La disponibilité des coordonnées de contact dans fidéli-nautille - quels enseignements pour les protocoles de collecte ?  
**G. CHARRANCE (INED)**

**M2024/03** : Discuter l'existence d'un effet de sélection dans un cadre multimode grâce à une analyse de sensibilité - Application aux enquêtes annuelles de recensement  
**L. COURT, S. QUANTIN**

**M2024/04** : Vers une désaisonnalisation des séries temporelles infra-mensuelles avec JDemetra+  
**A. SMYK, K. WEBEL**