Office fédéral de la statistique
Bundesamt für Statistik
Ufficio federale di statistica
Uffizi federal da statistica
OFS BFS UST

Rapport de méthodes
Methodenbericht
Rapporto sui metodi

Neuchâtel, février 2001

Désaisonnalisation

Aspects méthodologiques et application la statistique de l'emploi

Monique Graf

Service de méthodes statistiques
Dienst Statistische Methoden
Servizio metodi statistici

Renseignements:
Monique Graf
Office fédéral de la statistique
Service de méthodes statistiques
Espace de l'Europe 10, CH-2010 Neuchâtel
Tel.:(+41 32) 71 36615
Fax:(+41 32) 71 36375
Monique.Graf@bfs.admin.ch
Numéro de commande: 338-0003
© BFS/OFS

CH-2010 Neuchâtel
Espace de l'Europe 10

La désaisonnalisation s'impose de plus en plus comme une étape importante de l'analyse des séries chronologiques saisonnières, aussi l'Office fédéral de la statistique envisage d'introduire ces méthodes pour l'analyse d'un certain nombre de ses séries. Ce rapport, fruit de nombreux contacts internes et externes, se veut à la fois une introduction générale à la désaisonnalisation et une discussion plus mathématique de certains points difficiles à trouver dans la littérature. Comme illustration de la méthode, nous avons choisi une série présentant une saisonnalité marquée : l'emploi total dans la construction (branche 45 de la statistique de l'emploi, STATEM).

L'impressionnant travail de synthèse effectué par Eurostat dans le domaine de l'ajustement saisonnier a été une aide importante pour le choix de la méthode, de même que les travaux du U.S. Census Bureau et de Statistique Canada. Sur le plan interne, les nombreuses discussions avec Nicolas Fasel, de la Section des prix et de la consommation (PREIS), et avec Francis Saucy, de la section de la vie active et du marché du travail (EWL), ont contribué de manière décisive à l'élaboration de ce travail. Mes remerciements vont aussi à Paul-André Salamin, du Service de méthodes statistiques (METH) pour sa relecture attentive grâce à laquelle la lisibilité du présent document a été grandement améliorée.

Table des matières

1 Aspects méthodologiques **1**

1.1 Introduction 1
1.2 Problème de l'extraction du signal 2
1.3 Méthodologie basée sur un modèle ARIMA 8
1.3.1 Opérateurs de base 8
1.3.2 Le modèle par défaut (airline model) 10
1.3.3 Le modèle ARIMA 13
1.4 Méthodologie par filtre ad hoc 15
1.4.1 Etude de la statistique M7 (table D 8.A) 28

2 Désaisonnalisation de la statistique de l'emploi **31**

2.1 Adaptations aux recensements et aux changements d'échantillon 31
2.1.1 Correction d'un saut lors d'un changement d'échantillon 31
2.1.2 Interpolation rétrospective entre deux BZ 32
2.1.3 Adaptation au nouvel échantillon du deuxième trimestre 2000 35
2.1.4 Modèle global d'adaptation 35
2.2 Désaisonnalisation de la branche 45 39
2.2.1 Identification de l'ordre du modèle ARIMA et pré-ajustement 39
2.2.2 Décomposition de la série 41
2.2.3 Contrôle de la qualité de l'ajustement saisonnier 45
2.2.4 Evaluation de l'impact des révisions 50
2.3 Discussion 53

Annexes **57**

Bibliographie **71**

Chapitre 1

Aspects méthodologiques

1.1 Introduction

Depuis plus d'un siècle, les statisticiens et les économistes s'occupent de l'analyse et de la décomposition des séries chronologiques. La décomposition usuelle des séries chronologiques en tendance, cycle, composantes saisonnière et irrégulière a surtout été appliquée dans le cadre de l'analyse conjoncturelle. Le but est d'obtenir des informations sur l'état actuel et l'histoire d'une statistique conjoncturelle. Une part importante de cette analyse est l'extraction de la composante saisonnière, qui permet d'obtenir une image plus claire de l'état de l'économie.
Un consensus international émerge actuellement autour de deux méthodes comme étant les meilleures pour la désaisonnalisation. Ces méthodes sont fort bien décrites et comparées à l'ensemble des techniques disponibles dans la publication d'EUROSTAT [2]. Suite à cette étude (voir aussi Planas [7]) et à plusieurs autres que l'on peut trouver sur le site CIRCA[1], EUROSTAT recommande l'usage de X-12-ARIMA (méthode améliorant X-11 sur de nombreux points) ou de TRAMO/SEATS, développée par A. Maravall. La Banque centrale européenne a récemment mis sur pied un groupe de projet, dont le rapport [1] aboutit aux mêmes conclusions.
Dans les deux cas, le point de départ est un modèle de régression linéaire avec erreurs ARIMA (RegARIMA). La principale différence entre TRAMO/SEATS et X-12-ARIMA est que dans la première méthode, le filtre servant à l'extraction de la composante saisonnière est déduit des propriétés du modèle (voir Kaiser et Maravall [6] pour une introduction, à paraître) ; l'unicité de la composante saisonnière est alors garantie par la condition supplémentaire qu'elle doit être de variance minimum. Dans la seconde méthode, le modèle RegARIMA est utilisé pour prolonger la série aux deux extrémités, puis des filtres symétriques sont appliqués en plusieurs itérations. La longueur de ces filtres est adaptée itérativement à la variabilité des composantes de la série (voir Findley et al.[3] et le manuel [10]). La composante saisonnière ainsi extraite est donc déterminée par les filtres utilisés.
Dans un premier temps, nous avons choisi d'appliquer X-12-ARIMA aux séries de l'OFS, pour les raisons suivantes :

[1] http ://forum.europa.eu.int/Public/irc/dsis/Home/main

1. X-12-ARIMA est plus facile à expliquer aux utilisateurs.
2. Le programme, mis à disposition par le U.S. Census Bureau, semble avoir été plus largement testé.
3. La documentation est aisément accessible et très complète.
4. Il y a une interface graphique avec SAS très facile à utiliser.
5. Le fichier d'entrée (specification file) est beaucoup plus clair (plus mnémotechnique) que dans le cas de TRAMO/SEATS.
6. D'autres institutions en Suisse ont déjà une certaine expérience de X-12-ARIMA. Le choix de cette méthode améliore donc la comparabilité.
7. TRAMO/SEATS est une méthode purement basée sur un modèle. Notons que X-12-ARIMA utilise également des modèles dans la phase de détection des valeurs aberrantes, régression et ajustement du modèle ARIMA, mais la désaisonnalisation proprement dite n'en fait pas usage. Or la puissance des tests d'ajustement des modèles ARIMA dépend largement de la longueur de la série chronologique. Les séries à traiter à l'OFS étant souvent très courtes : une quarantaine d'observations dans le cas de la BESTA (statistique de l'emploi), par exemple, il nous semble donc plus prudent d'utiliser une méthode qui ne s'appuie pas exclusivement sur un modèle.

Le texte qui suit est un complément au manuel de l'utilisateur de X-12-ARIMA et ne le remplace pas. Après une introduction à la problématique générale de l'extraction des signaux, on présente les propriétés de la méthode, illustrées par des exemples synthétiques. Ces exemples permettent de jouer avec les caractéristiques de la série et d'étudier leur influence sur les étapes de la désaisonnalisation. Au chapitre 2, on verra une application de la méthode à un exemple réel, complétée par les différentes étapes du contrôle de la qualité du résultat.

1.2 Problème de l'extraction du signal

Considérons une série chronologique $\{x_t\}, t = 1, 2, ..., N$, où x_t est la valeur de la série au temps t, et N le nombre d'observations. Nous utiliserons comme unité de temps, soit le mois, soit le trimestre. Les accolades indiquent que l'on considère la série tout entière ; tandis que x_t est l'élément de la série au temps t.

Extraire un signal de cette série revient à la décomposer en la somme de deux termes

$$x_t = s_t + n_t \tag{1.1}$$

où $\{s_t\}$ représente le signal, et $\{n_t\}$ représente le reste (ou non-signal). Il y a bien sûr une infinité de façons de décomposer une série en la somme de deux termes, et pour que le problème soit bien posé, il faut définir ce que l'on entend par signal. Un signal est un phénomème qui se répète avec une certaine périodicité connue. Dans le contexte de la statistique publique, le signal, en général

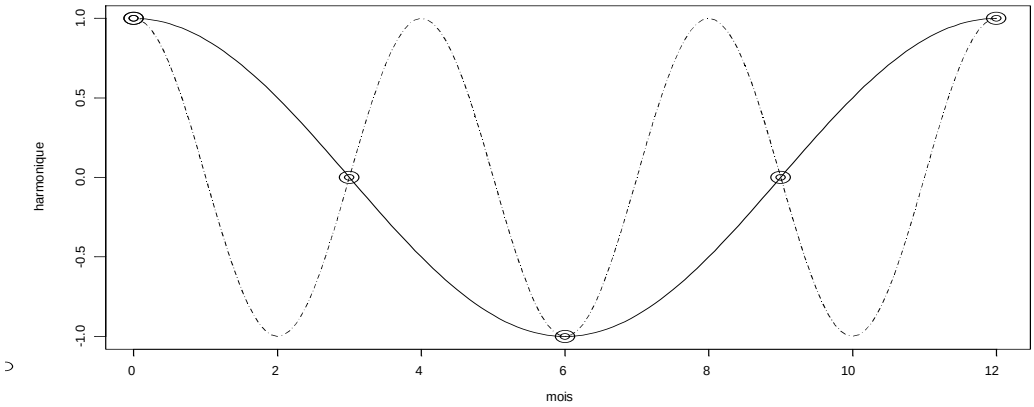


FIG. 1.1: Signal d'une périodicité d'une année (trait plein) et d'une périodicité de 4 mois (traitillé)

période	fréquence en multiples de τ	fréquence en radians	fréquence pour données mensuelles	fréquence pour données trimestrielles
1 trimestre	4 (fois par an)	4ω	$\pi 2/3$	–
1 semestre	2 (fois par an)	2ω	$\pi/3$	π
1 an	1 (fois par an)	1ω	$\pi/6$	$\pi/2$
2 ans	$\frac{1}{2}$ (1 fois tous les 2 ans)	$\frac{1}{2}\omega$	$\pi/12$	$\pi/4$
k ans	$\frac{1}{k}$ (1 fois tous les k ans)	$\frac{1}{k}\omega$	$\pi/6k$	$\pi/2k$
infinie ($k=\infty$)	nulle	0	0	0

TAB. 1.1: Relations entre période et fréquence

d'une périodicité annuelle, s'appelle la composante saisonnière. Ainsi, si le signal représente la composante saisonnière, la périodicité sera de 12 mois pour une série mensuelle, et de 4 trimestres pour une série trimestrielle. Si l'on veut extraire la tendance, on choisira un signal d'une périodicité très longue, par exemple de plus de 10 ans. Un cycle économique est caractérisé comme une composante dont la périodicité est de 2 à 8 ans environ. Remarquons enfin que l'on ne pourra pas extraire de signal d'une périodicité inférieure au semestre, pour les données trimestrielles (ou inférieure à deux mois pour les données mensuelles). De plus, la fig. 1.1 montre l'exemple de deux signaux périodiques totalement déterministes. On voit que si la série est trimestrielle, un signal d'une périodicité de 4 mois est indiscernable d'un signal d'une périodicité annuelle, car les points observés (mesurés tous les 3 mois) sont à l'intersection des deux courbes.

En résumé, la caractéristique principale du signal est sa périodicité.
- La plus petite périodicité observable est de 2 unités de temps (2 trimestres pour des données trimestrielles, ou 2 mois pour des données mensuelles).

- La plus grande périodicité décelable est de $N/2$ unités de temps, où N est la longueur de la série. Si, par exemple, une série est observée pendant 12 ans, la plus grande périodicité observable est de 6 ans ; ainsi dans ce cas, tous les cycles d'une période plus longue que 6 ans ne peuvent pas être dissociés de la tendance.
La plage de fréquence attribuée à la tendance dépend donc en partie de la longueur de la série à disposition.
Soit τ le nombre d'unités de temps par année ($\tau = 4$ pour une série trimestrielle et $\tau = 12$ pour une série mensuelle). Le tableau 1.1 résume la relation entre la période et la fréquence fondamentale $\omega = \frac{2\pi}{\tau}$.

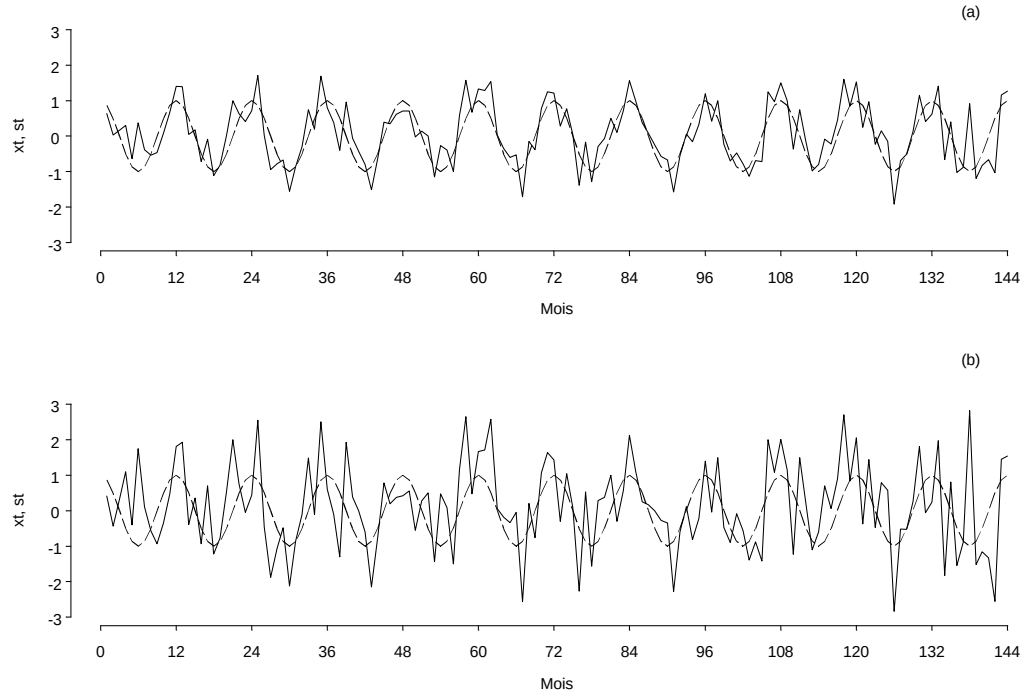


FIG. 1.2: Fonction purement périodique s_t (traitillé) et série $x_t = s_t + n_t$ (trait plein), où n_t est un bruit de variance $\frac{1}{4}$ (a) et 1 (b).

Exemple 1 : saisonnalité déterministe Considérons des données mensuelles avec N =144 mois. Représentons la composante saisonnière s_t au mois t par une fonction purement périodique, par exemple : $s_t = \cos(\frac{2\pi t}{12})$, et la composante irrégulière par une suite d'erreurs n_t non corrélées, d'espérance nulle et de même variance. Alors $\{s_t\}$ est purement déterministe, et $\{n_t\}$ est ce qu'on

appelle un *bruit blanc*. La fig. 1.2 est une représentation de

$$\begin{aligned} s_t &= \cos(\frac{2\pi t}{12}) \\ x_t &= s_t + n_t \quad t = 1, 2, ..., 144 \end{aligned}$$

où les n_t suivent une loi normale, de variance $\frac{1}{4}$ (partie a) et 1 (partie b).
La fonction purement périodique - ici un simple cosinus - est choisie uniquement pour illustrer le principe ; elle peut être donnée par n'importe quelle suite de 12 valeurs, une par mois, qui se répète chaque année.

Dans la pratique, les signaux ne sont en général pas totalement déterministes. Ils peuvent avoir une périodicité et une amplitude plus ou moins variable. Le but est d'extraire de la série originale le signal périodique perturbé, et non d'estimer une fonction purement périodique.

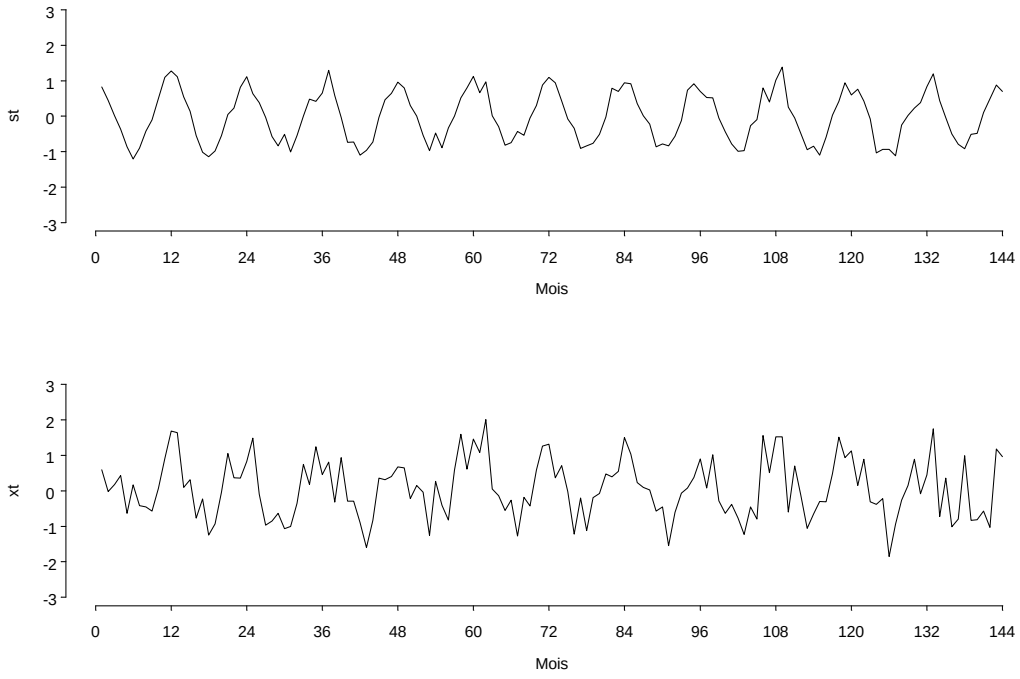


FIG. 1.3: Composante saisonnière (s_t) et série originale ($x_t = s_t + n_t$)

Exemple 2 : saisonnalité aléatoire Reprenons notre simple cosinus, mais donnons-lui une amplitude a_t qui fluctue aléatoirement autour de 1 et une phase qui fluctue aléatoirement d'une quantité b_t autour de la fréquence $\frac{2\pi t}{12}$, alors :

$$\begin{aligned} s_t &= a_t \cos(\frac{2\pi t}{12} + b_t) \\ x_t &= s_t + n_t \end{aligned}$$

Supposons que les $\{a_t\}$ soient non corrélées, normales, d'espérance 1 et de variance $\frac{1}{16}$, et les $\{b_t\}$ non corrélées, normales, d'espérance 0 et de variance $\frac{1}{100}$, et que les $\{n_t\}$ forment la même composante irrégulière que dans la fig. 1.2(a). Les séries $\{a_t\}$, $\{b_t\}$ et $\{n_t\}$ sont non corrélées entre elles. A la fig. 1.3 sont représentées la composante saisonnière $\{s_t\}$ ci-dessus, ainsi que la série $\{x_t\}$. La périodicité de $\{s_t\}$ n'est plus parfaite : c'est une composante saisonnière contenant une part aléatoire, mais la série originale $\{x_t\}$ contient plus de bruit que $\{s_t\}$.
Dans ces deux exemples, la série désaisonnalisée est tout simplement $\{n_t\}$.

On pourra bien sûr extraire plusieurs signaux de périodicité différente :

$$x_t = p_t + s_t + n_t \tag{1.2}$$

où p_t représente, par exemple, la tendance+le cycle, s_t la composante saisonnière, et n_t le reste, c'est-à-dire la composante irrégulière. Dans ce cas, la série désaisonnalisée est donnée par

$$p_t + n_t$$

Exemple 3 : Série avec saisonnalité et tendance aléatoires Ajoutons à la série de l'exemple précédent une tendance aléatoire, donnée par

$$\begin{aligned} p_t &= c_t + p_{t-1} \quad t = 2, 3, ..., 144 \\ p_1 &= c_1 \end{aligned}$$

où $\{c_t\}$ est une suite d'incréments (c'est-à-dire $p_t - p_{t-1}$) aléatoires non corrélés suivant une loi normale d'espérance $\frac{1}{20}$ et de variance $\frac{1}{400}$. La série originale x_t dans la formule (1.2) est la somme de cette tendance p_t, de la composante saisonnière s_t et du bruit blanc n_t de l'exemple 2. La série et la tendance sont représentées à la fig.1.4. Remarquons que si la variance des incréments était nulle, la tendance serait déterministe et donnée par une droite de pente $\frac{1}{20}$ (égale à l'espérance de c_t).

Exemple 4 : Série avec tendance aléatoire quadratique Pour obtenir une tendance aléatoire quadratique, il faut poser que la suite $\{c_t\}$ est la suite des incréments des incréments

$$c_t = (p_t - p_{t-1}) - (p_{t-1} - p_{t-2}) = p_t - 2p_{t-1} + p_{t-2}$$

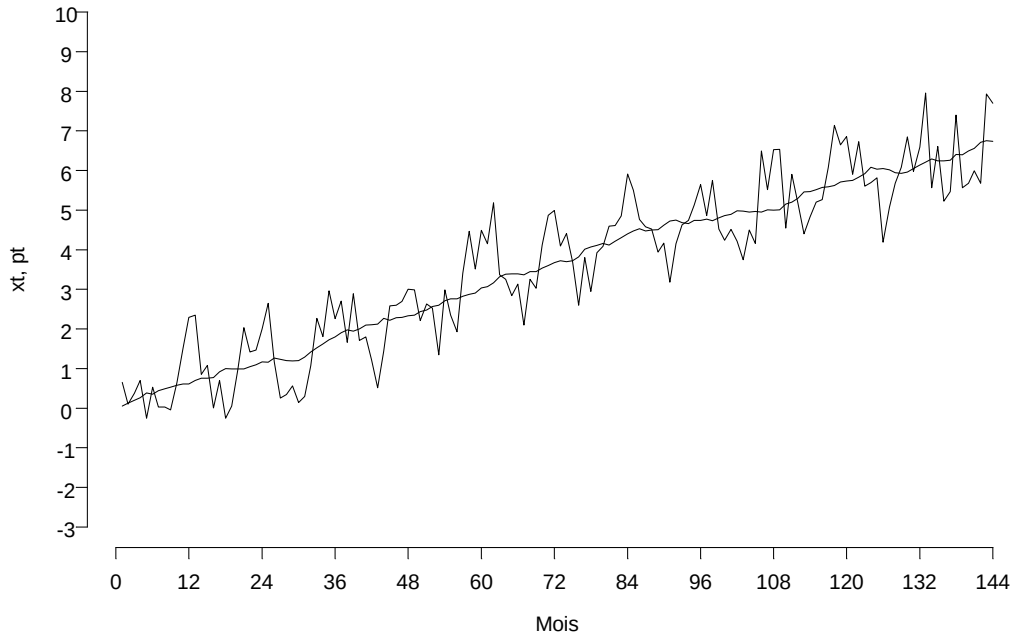


FIG. 1.4: Série originale x_t et sa tendance p_t

donc :

$$\begin{aligned} p_t &= c_t + 2p_{t-1} - p_{t-2} \\ p_1 &= c_1 \\ p_2 &= c_2 + 2c_1 \end{aligned}$$

Choisissons pour $\{c_t\}$ est une suite de variables aléatoires non corrélées suivant une loi normale d'espérance $\frac{1}{1000}$ et de variance $\frac{1}{10'000}$. La série originale x_t dans la formule (1.2) est la somme de la tendance ainsi obtenue p_t, de la composante saisonnière s_t et du bruit blanc n_t de l'exemple 2. La série et la tendance sont représentées à la fig.1.5(a). Remarquons que si la variance des c_t était nulle, la tendance serait déterministe et donnée par une parabole. Cette parabole est représentée à la fig.1.5(a) par la courbe en pointillé. On constate que la très petite variance attribuée aux c_t suffit pour faire diverger appréciablement la tendance aléatoire de cette parabole.
La série désaisonnalisée est la somme de la tendance aléatoire quadratique p_t et de la composante irrégulière n_t. Cette série est représentée à la fig.1.5(b).

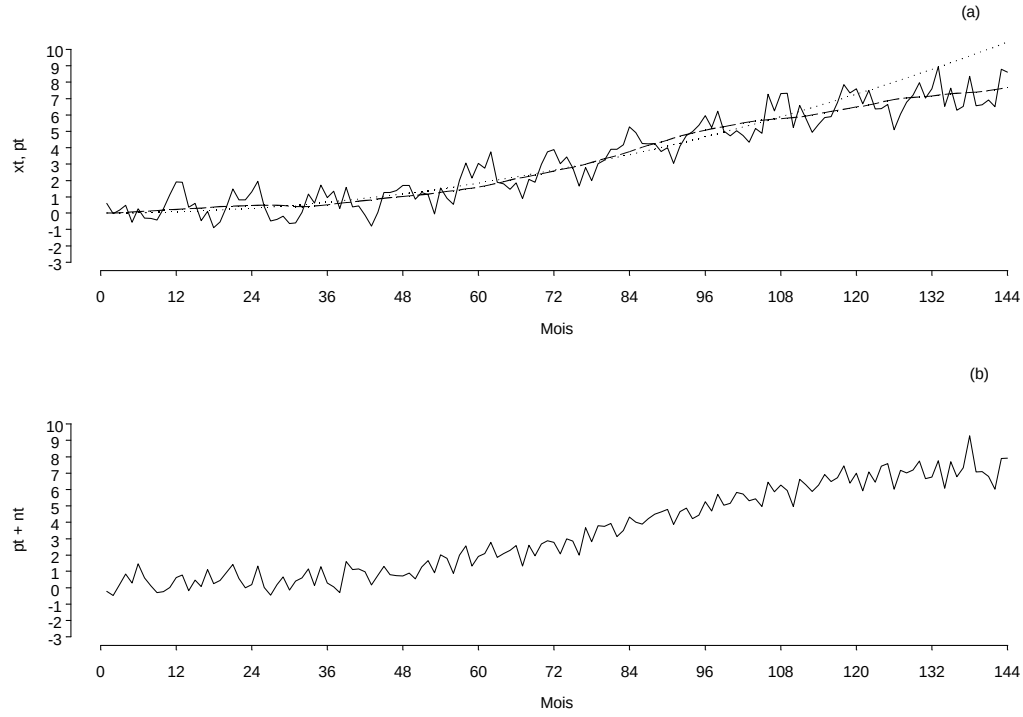


FIG. 1.5: (a) Tendance aléatoire de type quadratique (p_t en traitillé) et série originale ($x_t = p_t + s_t + n_t$ en trait plein). La courbe en pointillé est la tendance déterministe obtenue si c_t était de variance nulle. (b) Série désaisonnalisée.

1.3 Méthodologie basée sur un modèle ARIMA

1.3.1 Opérateurs de base

Il y a deux opérateurs de base : l'opérateur de décalage et l'opérateur de différence.
- L'opérateur de décalage B (pour backward) permet de reculer d'un pas dans la succession des éléments de la série $\{x_t\}$:

$$\begin{aligned} Bx_t &= x_{t-1} \\ B^2x_t &= B(Bx_t) = Bx_{t-1} = x_{t-2} \\ &\dots \\ B^{12}x_t &= x_{t-12} \end{aligned}$$

- L'opérateur de différence ∇ donne la différence d'un terme de la série avec le terme précédent :

$$\begin{aligned} \nabla x_t &= x_t - x_{t-1} \\ \nabla^2 x_t &= \nabla\left(x_t - x_{t-1}\right) = x_t - x_{t-1} - \left(x_{t-1} - x_{t-2}\right) \\ &= x_t - 2x_{t-1} + x_{t-2} \end{aligned}$$

- L'opérateur ∇_k donne la différence d'un terme de la série avec le k-ième terme précédent, par exemple :

$$\begin{aligned} \nabla_2 x_t &= x_t - x_{t-2} \\ \nabla_4 x_t &= x_t - x_{t-4} \\ \nabla_{12} x_t &= x_t - x_{t-12} \end{aligned}$$

Propriétés :
1) ∇ filtre une tendance linéaire déterministe (la transforme en une constante) :

$$\nabla(a + bt) = a + bt - (a + b(t-1)) = b$$

2) ∇^2 filtre une tendance quadratique déterministe (et annule une tendance linéaire déterministe) :

$$\nabla^2(a + bt + ct^2) = 2c$$

3) ∇_4 annule une périodicité annuelle déterministe, pour une série trimestrielle, car dans ce cas $s_t = s_{t-4}$:

$$\nabla_4 s_t = 0 \quad \text{si} \quad s_t = s_{t-4} \quad \text{pour tout } t$$

4) ∇_{12} annule une périodicité annuelle déterministe, pour une série mensuelle, car dans ce cas $s_t = s_{t-12}$:

$$\nabla_{12} s_t = 0 \quad \text{si} \quad s_t = s_{t-12} \quad \text{pour tout } t$$

5) On peut écrire les opérateurs du type différence comme des polynômes en B

$$\begin{aligned} \nabla &= 1 - B \\ \nabla^2 &= (1-B)^2 = 1 - 2B + B^2 \\ \nabla_2 &= 1 - B^2 = (1-B)(1+B) \\ \nabla_4 &= 1 - B^4 = (1-B)\left(1 + B + B^2 + B^3\right) \\ \nabla_{12} &= 1 - B^{12} = (1-B)\left(1 + B + B^2 + B^3 + ... + B^{11}\right) \end{aligned}$$

Ainsi, ∇_4, par exemple, peut être décomposé en deux opérations successives :

1. effectuer la somme des quatres derniers termes à partir du terme courant :

$$\left(1+B+B^2+B^3\right)x_t = x_t + x_{t-1} + x_{t-2} + x_{t-3}$$

2. effectuer la différence de deux sommes successives :

$$\begin{aligned} & \left(1+B+B^2+B^3\right)x_t - B\left(1+B+B^2+B^3\right)x_t \\ = \; & (x_t + x_{t-1} + x_{t-2} + x_{t-3}) - (x_{t-1} + x_{t-2} + x_{t-3} + x_{t-4}) \\ = \; & x_t - x_{t-4} = \left(1-B^4\right)x_t \end{aligned}$$

Application à des données trimestrielles : la différence entre deux trimestres identiques de deux années successives est égale à la différence des sommes consécutives sur 4 trimestres. Ces décompositions d'opérateurs interviennent dans la désaisonnalisation des séries.

Autre exemple utile de décomposition :

$$\begin{aligned} \nabla\nabla_{12}x_t &= (1-B)\left(1-B^{12}\right)x_t \\ &= \left[(1-B) - \left(B^{12}-B^{13}\right)\right]x_t \\ &= (x_t - x_{t-1}) - (x_{t-12} - x_{t-13}) \end{aligned}$$

.

6) Comme l'opérateur qui donne la somme sur les 4 trimestres ou les 12 mois précédents est très souvent utilisé, on le représente par le symbole S :

$$\begin{aligned} S_4 &= 1+B+B^2+B^3 \\ S_{12} &= 1+B+B^2+B^3+...+B^{11} \end{aligned}$$

7) S filtre une périodicité déterministe : Pour une période τ ($\tau = 4, 12$), on a $S_\tau s_t = c$ (somme sur une période).

En résumé :

∇ agit sur la tendance et S agit sur la composante saisonnière.

1.3.2 Le modèle par défaut (airline model)

Le modèle par défaut - dans le cadre de la problématique de la désaisonnalisation - pour une série présentant une tendance et des fluctuations aléatoires d'une périodicité d'un an a été suggéré par Box et Jenkins dans les années 70 pour modéliser le nombre mensuel de passagers d'une ligne d'avion, d'où son nom anglais de "airline model". Ce modèle est extrêmement simple, puisqu'il n'a que 3 paramètres ; il est pourtant d'une étonnante souplesse. Le modèle ARIMA est une généralisation de ce modèle.

$$\nabla\nabla_{12}z_t = (1-\theta_1 B)(1-\theta_{12}B^{12})e_t \tag{1.3}$$

$$\{e_t\} \quad : \quad \text{bruit blanc} \tag{1.4}$$

$$Var(e_t) = \sigma^2 \tag{1.5}$$

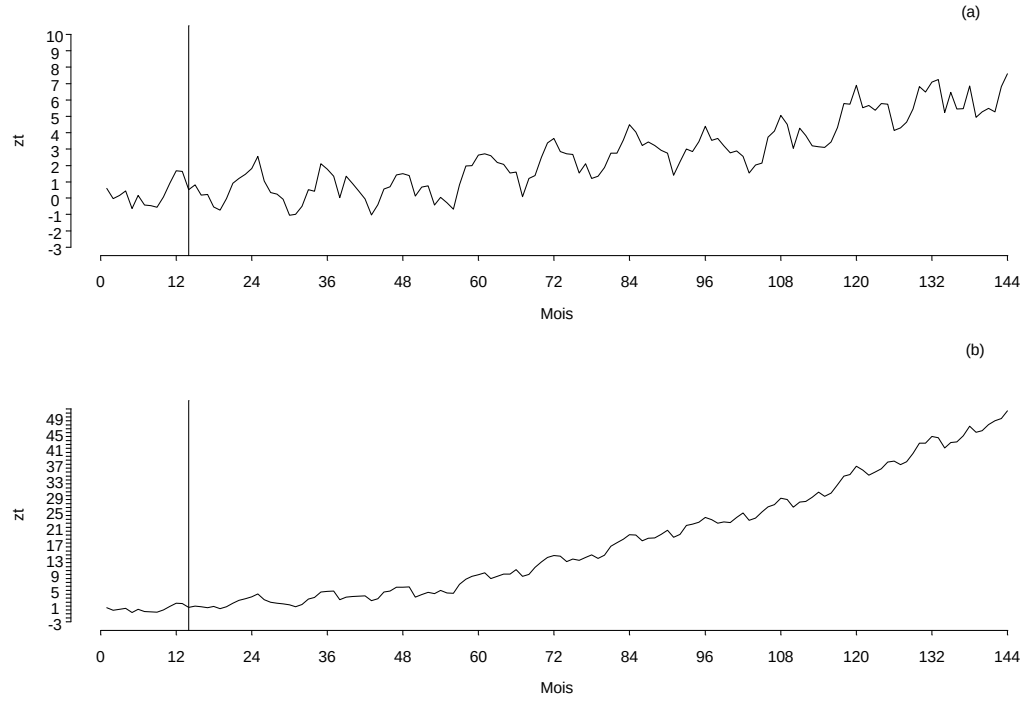


FIG. 1.6: Deux réalisation du modèle par défaut : (a) $\theta_1 = 0.6 \quad \theta_{12} = 0.8$ (b) $\theta_1 = 0.1 \quad \theta_{12} = 0.2$

Remarque *La notation est celle (classique) de Box-Jenkins, reprise par X-12-Arima et S-Plus ; pour Tramo-Seats et Demetra, il faut changer les signes des paramètres θ.*

Definition 1 *Un **bruit blanc** $\{e_t\}$ est une suite de variables aléatoires non corrélées d'espérance nulle et de même variance.*

On pose en général la condition supplémentaire que les e_t suivent une loi normale. Ce bruit blanc représente une suite de chocs aléatoires qui viennent perturber l'évolution de $\{z_t\}$.

Pour pouvoir calculer z_t à l'aide de (1.3), il faut poser des conditions initiales pour les 13 premières valeurs de z_t et de e_t :

$$\begin{aligned} z_1, z_2, ...z_{13} &: \text{ valeurs initiales de la série observée} \\ e_1 &= e_2 = ... = e_{13} = 0 \\ z_t &= z_{t-1} + z_{t-12} - z_{t-13} + e_t - \theta_1 e_{t-1} - \theta_{12} e_{t-12} + \theta_1 \theta_{12} e_{t-13} \qquad t \geq 14 \end{aligned}$$

La fig. 1.6 montre deux réalisations de ce modèle. Dans les deux cas de figure (a) et (b), les valeurs initiales de la série $\{z_t\}$ (avant la barre verticale) sont les 13 premières valeurs de $\{x_t\}$ de l'exemple 4, et la série $\{e_t\}$ est identique à la série $\{n_t\}$ de l'exemple 4.

Interprétation des paramètres θ_1 et θ_{12} On voit l'effet des paramètres θ_1 et θ_{12} : plus ils sont grands, et plus la valeur z_t est influencée par les chocs aléatoires du mois précédent (e_{t-1}) et des mêmes mois de l'année précédente (e_{t-12} et e_{t-13}). La série est alors plus stable (voir les échelles très différentes), car elle dépend dans une plus grande mesure de son passé.

Analyse des exemples précédents

- Exemple 1 : $x_t = \cos(2\pi t/12) + n_t$, on a $\nabla_{12}\, x_t = \nabla_{12}\, n_t$. L'application du modèle par défaut donne

$$\nabla\nabla_{12}x_t = \nabla\left[\cos(2\pi t/12) - \cos(2\pi\,(t-12)\,/12)\right] + \nabla\nabla_{12}n_t = \nabla\nabla_{12}n_t$$

  Dans ce cas, on a $\theta_1 = \theta_{12} = 1$. Pour éliminer une composante saisonnière déterministe, l'opérateur ∇_{12} suffit. On a donc "surdifférentié" la série. Ceci est en général à éviter, car les résultats dépendent de plus de valeurs initiales et sont moins stables.
- Exemple 2 : $x_t = a_t \cos(2\pi t/12 + b_t) + n_t$. Le modèle par défaut donne de nouveau $\widehat{\theta}_1 \simeq 1$ et $\widehat{\theta}_{12} \simeq 1$. La variabilité de la composante saisonnière n'est donc pas détectée en tant que telle, mais absorbée pas la variance du bruit blanc, estimée à $\widehat{\sigma}^2 = 0.31$ (alors que le paramètre théorique valait 0.25). Un examen des carrés des résiduelles montrera la non-linéarité du processus générant les données ; cette non-linéarité est non significative.
- Exemple 3 : $x_t = p_t + a_t \cos(2\pi t/12 + b_t) + n_t$, avec $p_t = p_{t-1} + c_t$. Le modèle par défaut remplace la tendance de type linéaire $\{p_t\}$ par le bruit blanc $\{c_t\}$ et donne de nouveau $\widehat{\theta}_1 \simeq 1$ et $\widehat{\theta}_{12} \simeq 1$.
- Exemple 4 : $x_t = p_t + a_t \cos(2\pi t/12 + b_t) + n_t$, avec $p_t = 2p_{t-1} - p_{t-2} + c_t$. Le modèle par défaut ne s'applique pas, car il ne remplace pas une tendance de type quadratique par un bruit blanc. Il faut le généraliser à

$$\nabla^2\nabla_{12}x_t = \left(1 - \theta_1 B - \theta_2 B^2\right)\left(1 - \theta_{12}B^{12}\right) n_t \tag{1.6}$$

  Les paramètres obtenus sont : $\widehat{\theta}_1 = 1.89$, $\widehat{\theta}_2 = -0.91$, $\widehat{\theta}_{12} = 1.00$ et $\widehat{\sigma}^2 = 0.35$. La variance estimée est légèrement supérieure à celle de l'exemple 2, car elle absorbe une part de la variabilité de la tendance.
  Si $(1 - \theta_1 B - \theta_2 B^2)$ était égal à ∇^2, on aurait $\theta_1 = 2$ et $\theta_2 = -1$. Pour interpréter la différence avec les paramètres estimés, écrivons

$$1 - \widehat{\theta}_1 B - \widehat{\theta}_2 B^2 = 1 - 2r\cos\alpha B + r^2 B^2$$

  alors, $r^2 = 0.91$ et $2r\cos\alpha = 1.89$, donc $r \simeq 0.95$ et $\cos\alpha \simeq 0.99$, $\alpha \simeq 0.125$ rad. Cet opérateur filtre une oscillation déterministe qui s'amortit d'un facteur r à chaque pas, donc très lentement, puisque r est proche de 1. En posant $\alpha = \pi/6k$ (voir tableau 1.1), on trouve que la période de cette oscillation est de $k \simeq 4.2$ ans. Cette période est longue vis-à-vis de la longueur de la série x_t, qui est de 12 ans. Elle pourrait correspondre à un cycle, mais

est difficile à distinguer empiriquement de la tendance sur une période d'observation aussi courte. Un "vrai" cycle aurait d'ailleurs une valeur de r égale à 1. On voit sur la fig 1.5(a) deux petits changements de courbure sur la tendance $\{p_t\}$ à environ 4 ans d'intervalle. C'est peut-être ce que ces paramètres expriment.
A l'opérateur ∇^2 correspond $r = 1$ et $\cos\alpha = 1$, donc $\alpha = 0$ et $k = \infty$. Ainsi une tendance peut être interprétée comme une oscillation de période infinie qui ne s'amortit pas. En conclusion, on ne distinguera pas entre tendance et cycle, et la composante est souvent appelée "trend-cycle" ou tendance-cycle.

Remarque 1 Dans tous ces exemples, on a $\widehat{\theta}_{12} \simeq 1$. Cela provient de la construction de nos exemples. La perturbation de la saisonnalité par n_t au temps t ne dépend pas du passé (ni des valeurs précédentes de la série $\{x_t\}$, ni de celles de $\{n_t\}$). Ainsi, dans l'exemple 1, au lieu d'éliminer la saisonnalité par l'opérateur ∇_{12}, on pourrait la modéliser par la fonction déterministe $\cos(2\pi t/12)$, qu'on introduit alors comme variable de régression.
Dans les exemples suivants, on peut utiliser la formule $\cos(A+B) = \cos A\cos B - \sin A\sin B$:

$$\begin{aligned} a_t\cos(\frac{2\pi t}{12} + b_t) &= a_t\cos(b_t)\cos(\frac{2\pi t}{12}) - a_t\sin(b_t)\sin(\frac{2\pi t}{12}) \\ &= u_t\cos(\frac{2\pi t}{12}) + v_t\sin(\frac{2\pi t}{12}) \\ &\cong \overline{u}\cos(\frac{2\pi t}{12}) + \overline{v}\sin(\frac{2\pi t}{12}) + e_t \end{aligned}$$

et introduire deux variables de régression ($\cos(\frac{2\pi t}{12})$ et $\sin(\frac{2\pi t}{12})$). Dans l'exemple 4, ce modèle s'écrit :

$$\nabla^2\left[x_t - \left(\overline{u}\cos(\frac{2\pi t}{12}) + \overline{v}\sin(\frac{2\pi t}{12})\right)\right] = \left(1 - \theta_1 B - \theta_2 B^2\right) n'_t$$

Le bruit blanc $\{n'_t\}$ n'est pas exactement identique à $\{n_t\}$, car il absorbe en partie les erreurs $\{e_t\}$. Les paramètres estimés $\widehat{\theta}_1$, $\widehat{\theta}_2$ et $\widehat{\sigma}^2$ sont pratiquement les mêmes qu'avec le premier modèle.

Remarque 2 Si on ignorait la forme de la fonction périodique, on introduirait, dans chacun des exemples, les variables indicatrices des mois de l'année comme variables de régression.

1.3.3 Le modèle ARIMA

Posons quelques définitions mathématiques.

Definition 2 *Soit $\{x_t\}$ une série chronologique. Elle est dite **stationnaire (du 2e ordre)**, si l'espérance et la variance de x_t existent et sont indépendantes de t.*

Par exemple, un bruit blanc est une série stationnaire d'espérance nulle pour tout t. La tendance linéaire $\{p_t\}$ de l'exemple 2 n'est pas une série stationnaire, mais $\{\nabla p_t\} = \{c_t\}$ est stationnaire ; de même la tendance quadratique $\{p_t\}$ de l'exemple 4 n'est pas une série stationnaire, mais $\{\nabla^2 p_t\} = \{c_t\}$ est stationnaire.

Definition 3 *On dit que la série stationnaire $\{x_t\}$ d'espérance nulle suit un modèle* ***autorégressif à moyenne mobile d'ordre*** *(p, q), et on la note ARMA(p, q), si x_t peut être représenté par*

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + ... + \varphi_p x_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} ... - \theta_q \epsilon_{t-q} \tag{1.7}$$

où $\{\epsilon_t\}$ est un bruit blanc de variance σ^2.

Cela signifie donc que l'on suppose que x_t dépend des p valeurs précédentes $x_{t-1}, ..., x_{t-p}$ (la série est en *autorégression* sur ses valeurs passées), d'un choc aléatoire imprévisible ϵ_t, et des q chocs précédents $\epsilon_{t-1}, ..., \epsilon_{t-q}$. Si $\{x_t\}$ n'est pas d'espérance nulle, on applique le modèle à la série centrée. L'ordre (p, q) représente en quelque sorte la mémoire de la série. On peut écrire (1.7) de manière plus compacte :

$$(1 - \varphi_1 B - \varphi_2 B^2 - ...\varphi_p B^p) x_t = (1 - \theta_1 B - \theta_2 B^2 - ...\theta_q B^q) \epsilon_t \tag{1.8}$$

Definition 4 *On dit que la série* ***non*** *stationnaire $\{x_t\}$ suit un modèle* ***autorégressif intégré à moyenne mobile d'ordre*** *(p, d, q), et on la note ARIMA(p, d, q), si*

$$\nabla^d x_t = \varphi_1 \nabla^d x_{t-1} + \varphi_2 \nabla^d x_{t-2} + ... + \varphi_p \nabla^d x_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} ... - \theta_q \epsilon_{t-q} \tag{1.9}$$

où $\{\epsilon_t\}$ est un bruit blanc de variance σ^2.

On voit que si $\{x_t\}$ suit un modèle ARIMA(p, d, q), sa d^e différence $\{y_t\} = \{\nabla^d x_t\}$ suit un modèle ARMA(p, q).
On peut écrire (1.9) de manière plus compacte :

$$(1 - \varphi_1 B - \varphi_2 B^2 - ...\varphi_p B^p) \nabla^d x_t = (1 - \theta_1 B - \theta_2 B^2 - ...\theta_q B^q) \epsilon_t \tag{1.10}$$

On voit apparaître dans (1.8) et (1.10) deux polynômes en B : $\varphi(B) = (1 - \varphi_1 B - \varphi_2 B^2 - ...\varphi_p B^p)$ et $\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - ...\theta_q B^q)$.

Definition 5 *La série non stationnaire $\{x_t\}$ suit un* ***modèle ARIMA saisonnier****, noté ARIMA$(p, d, q)(P, D, Q)_s$, si :*

$$\varphi(B)\, \nabla^d \nabla_s^D\, x_t = \theta(B)\, \epsilon_t$$

ou{' $\varphi(B)$ et $\theta(B)$ peuvent chacun être factorisés en deux polynômes, dont l'un dépend de B, et l'autre ne dépend que de B^s, et si $\{\nabla^d \nabla_s^D x_t\}$ est stationnaire.

Pour les séries mensuelles, $s = 12$, et pour les séries trimestrielles, $s = 4$.
Le modèle par défaut (1.3) est donc un modèle ARIMA$(0, 1, 1)\ (0, 1, 1)_{12}$, c'est-à-dire

$$\begin{array}{ll} s = 12 & \\ p = P = 0 & \varphi(B) = 1 \\ q = Q = 1 & \theta(B) = (1 - \theta_1 B)(1 - \theta_{12} B^{12}) \end{array}$$

Le modèle (1.6) pour l'exemple 4 est un modèle ARIMA$(0, 2, 2)(0, 1, 1)_{12}$. On a pour ce modèle :

$$
\begin{array}{lll}
s = 12 & & \\
p = P = 0 & & \varphi(B) = 1 \\
q = 2 \quad Q = 1 & & \theta(B) = (1 - \theta_1 B - \theta_2 B^2)(1 - \theta_{12} B^{12})
\end{array}
$$

$\varphi(B)$ est de degré 0 (donc $p = 0$ et $P = 0$), et $\theta(B)$ est formé de 2 facteurs, l'un de degré 2 en B ($q = 2$), et l'autre de degré 1 en B^{12} ($Q = 1$). En général, le degré de $\varphi(B)$ est $p + sP$ et celui de $\theta(B)$ est $q + sQ$.

Les programmes de désaisonnalisation font usage de ce modèle ARIMA saisonnier, mais pas de la même manière.

- X-12-Arima n'utilise ce modèle que pour prolonger la série aux deux extrémités, de façon à stabiliser l'estimation de la composante saisonnière qui est obtenue par les filtres implémentés dans X-11.
- Tramo-Seats utilise, pour calculer un filtre, directement les propriétés du modèle estimé dans le domaine des fréquences - propriétés basées sur les racines des polynômes $\varphi(B)$ et $\theta(B)$, comme nous l'avons vu dans l'analyse de l'exemple 4. Ce filtre possède des propriétés d'optimalité asymptotique, pour autant que le modèle ARIMA ajusté soit bon.

La qualité de l'ajustement du modèle ARIMA est donc cruciale pour Tramo-Seats. On peut s'attendre à ce que l'ajustement saisonnier soit meilleur comparativement à X-12-Arima, si le modèle est adéquat. D'un autre côté, X-12-Arima devrait être plus robuste à des écarts au modèle. Remarquons que le modèle ARIMA est un modèle linéaire. Il y a actuellement de nombreuses recherches sur les modèles de séries chronologiques non-linéaires, comme les modèles de type ARCH (autorégressif à hétéroscédasticité conditionnelle), les modèles à volatilité stochastique, les modèles à longue mémoire, etc. (pour un survol, voir [4]).

1.4 Méthodologie par filtre ad hoc

X-12-Arima allonge la série à désaisonnaliser d'une année aux deux extrémités en utilisant les prévisions basées sur le modèle RegArima. Cela permet d'éviter les effets de bord dans le calcul du filtre - un désavantage reconnu de X-11, voir par exemple [9, chap. III]. Les filtres utilisés sont ceux de X-11. Nous ne les décrivons pas ici, car on en trouve une description très complète dans la publication de Ladiray et Quenneville [8]. Ce texte donne aussi des commentaires détaillés sur les tables imprimées par le module X11 de X-12-Arima. Le texte de Schips et Stier [9, chap. III] forme un complément utile, même si le passage de X-11 à X-12-Arima répond à certaines critiques. En allemand, on peut recommander la lecture du papier de discussion de Kirchner [5] qui présente la méthode X-12-Arima et l'applique à plusieurs séries économiques allemandes.

Posons
$O =$ série originale $C =$ tendance + cycle $S =$ composante saisonnière
$I =$ irrégulier $SA =$ série désaisonnalisée.
Les modes d'ajustement saisonnier suivants sont possibles :

mode	modèle pour O	modèle pour SA
multiplicatif	$O = CSI$	$SA = CI$
additif	$O = C + S + I$	$SA = C + I$
pseudo-additif	$O = C\,(S - 1 + I)$	$SA = CI$
log-additif	$\log(O) = C + S + I$	$SA = \exp(C + I)$

Etude expérimentale de X-12-Arima Le programme X-12-Arima possède un grand nombre d'options : à côté de la forme du modèle ARIMA, on peut fixer ou non la longueur du filtre saisonnier, changer les valeurs critiques pour la détection des valeurs aberrantes, ainsi que leur poids dans le calcul des composantes. Pour évaluer les interactions entre ces options, nous allons exécuter le programme avec des combinaisons d'options différentes. Nous allons aussi comparer l'ajustement du modèle ARIMA avec celui donné par S-Plus.
Notre exemple sera la série $\{x_t\}$ de l'exemple 4, représentée à la fig. 1.5. Rappelons la construction de cette série : on part de 4 bruits blancs $\{a_t\}$, $\{b_t\}$,$\{c_t\}$,$\{n_t\}$, où , $a_t \sim N(1, \frac{1}{16})$, $b_t \sim N(0, \frac{1}{100})$, $c_t \sim N(\frac{1}{1000}, \frac{1}{10000})$ et $n_t \sim N(0, \frac{1}{4})$. On pose :

$$\begin{aligned} s_t &= a_t \cos\left(\frac{2\pi t}{12} + b_t\right) \qquad \nabla^2 p_t = c_t \\ x_t &= p_t + s_t + n_t \end{aligned}$$

La table 1.2 résume les résultats de 11 exécutions de X-12-Arima pour l'exemple 4.
Le modèle théorique utilisé est dans tous les cas donné par (1.6).

- PARAMETRES sont les 3 paramètres θ_1, θ_2 et θ_{12} intervenant dans (1.6). Ils sont, selon les cas, fixés aux valeurs estimées par S-Plus, ou réestimés par X-12-Arima.
- AICC est un critère d'ajustement du modèle basé sur la vraisemblance et tenant compte du nombre de paramètres estimés. Ce critère n'a qu'une valeur relative : plus il est bas, meilleur est l'ajustement. On compare, à l'aide de AICC, des modèles emboîtés, c'est-à-dire que l'on regarde s'il vaut la peine d'envisager un modèle plus général (avec plus de paramètres) ou si un modèle plus simple suffit. Dans la pratique, on considère généralement qu'un écart inférieur à 2 entre deux valeurs de AICC n'est pas significatif. Dans la table 1.2, certains modèles ont leurs paramètres fixés aux valeurs estimées par S-Plus. Comme les paramètres fixés ne sont pas réellement connus, nous avons corrigé la valeur du critère pour pouvoir comparer les différents modèles. Ainsi, dans la table 1.2, seuls les appels 7 (a) et (b) sont nettement inférieurs du point de vue de la qualité de l'ajustement du modèle Arima. Pour les modèles 1 à 9, le modèle Arima est le même.

NO	PARAMETRES	AICC [1]	SEASONALMA	CRITICAL	SIGMALIM	N_c [2]	M_1	M_5	Q
1	fixés	*258.9*	défaut	3.5	1.25, 2.25	3	1.4	1.6	0.80
2	estimés	259.0	défaut	3.5	1.25, 2.25	4	1.4	1.6	0.81
3	fixés	*258.9*	s3x5	3.5	1.25, 2.25	4	1.9	1.8	0.97
4	fixés	*258.9*	défaut	défaut	1.25, 2.25	5	2.4	1.9	1.07
5	fixés	*258.9*	défaut	3.5	défaut	5	2.2	1.8	1.11
6	X11	—	défaut	défaut	défaut	5	2.2	1.9	1.14
7	estimés [3] (a) (b)	286.1 278.8	défaut	défaut	défaut	5	2.3	1.9	1.15
8	fixés	*258.9*	stable	3.5	1.25, 2.25	5	2.8	1.8	1.43
9	fixés	*258.9*	s3x15	3.5	1.25, 2.25	5	2.8	1.8	1.43
10	estimés + 11 régr. trigo	260.5	défaut	3.5	1.25, 2.25	4	1.4	1.6	0.80
11	estimés + 2 régr. trigo	258.8	défaut	3.5	1.25, 2.25	4	2.0	1.8	1.04

[1] valeurs en italique : le critère a été modifié, car les paramètres fixés ont en fait été estimés par S-Plus.

[2] nombre de critères >1 dans la construction de la statistique Q.

[3] valeurs intiales : (a) par défaut / (b) données.

TAB. 1.2: Modèles pour l'exemple 4 ; ajustement par vraisemblance exacte

Pour suivre l'idée de la remarque 1, on a introduit, dans les modèles 10 et 11, une saisonnalité déterministe à l'aide de régresseurs trigonométriques. Le modèle Arima est alors estimé en même temps que les paramètres de la régression. Il y a 11 régresseurs dans le modèle 10 ($\cos(k\,2\pi t/12)$, $k = 1, ..., 6$ et $\sin(k\,2\pi t/12)$, $k = 1, ..., 5$), et 2 dans le modèle 11 ($\cos(2\pi t/12)$ et $\sin(2\pi t/12)$). Du strict point de vue de l'ajustement global mesuré par le critère AICC, le modèle 11 est à préférer au modèle 10, pour des raisons de parcimonie. Cependant, ses performances pour la désaisonnalisation sont inférieures, comme le montre le critère Q (la zone de rejet de l'ajustement saisonnier est $Q > 1$).

– CRITICAL fixe la valeur critique pour la détection des valeurs aberrantes (outliers). X-12-Arima considère 3 types différents de valeurs aberrantes : additives (additive outliers, AO), changement de niveau (level shift, LS) et rampe (temporary change, TC). Le premier type indique une perturbation instantanée de la série, le second un changement de niveau sur un intervalle de temps dont la longueur est calculée, et le troisième un changement progressif, suivi d'un retour brusque au niveau initial. Lorsque CRITICAL est égal à sa valeur par défaut (qui vaut 3.8829 pour une série de longueur 144 ; voir X-12-Arima Quick reference, table 3, p.16), il n'y a pas détection de données aberrantes pour notre exemple. Lorsque CRITICAL=3.5, deux AO apparaissent. Ils sont ajustés par RegArima, en même temps que les paramètres du modèle Arima. On voit que l'introduction de CRITICAL=3.5, n'a pas d'influence sur l'ajustement global du modèle (critère AICC), mais diminue la valeur de Q (voir plus bas).

– Les paramètres SEASONALMA et SIGMALIM sont liés à l'extraction des composantes par X-11.

  SEASONALMA contrôle la longueur du filtre utilisé pour l'extraction de la composante saisonnière. Cette extraction se fait en 3 itérations. La valeur par défaut est MSR ou "moving seasonality ratio" : à la première itération, le programme extrait une composante saisonnière provisoire par un filtre 3x3, puis au deuxième passage, il utilise un filtre 3x5 ; enfin, au troisième passage, la longueur du filtre est adaptée à la variabilité relative des composantes irrégulières et saisonnières provisoires de l'itération 2 : plus l'irrégulier est variable, plus le filtre sera long. On constate à la table 1.2, qu'il n'y a pas d'avantage à fixer SEASONALMA.

  SIGMALIM contrôle la pondération à apporter aux extrêmes de la composante irrégulière. Les deux valeurs indiquées sont en unités d'écart-type de l'irrégulier. Si la valeur standardisée de la composante irrégulière au temps t est en valeur absolue inférieure à la première limite, la valeur I_t de l'irrégulier au temps t a un poids de 1, si elle est supérieure à la seconde limite, I_t a un poids nul, pour les valeurs intermédiaires, le poids est compris entre 0 et 1. On effectue donc une winsorisation des poids des I_t, qui entraîne une modification du filtre. La composante saisonnière finale est alors extraite en utilisant cet irrégulier pondéré. Dans notre exemple, on voit qu'en diminuant les limites de SIGMALIM (par défaut : 1.5 et 2.5), on obtient une valeur de Q acceptable.

– Toues les statistiques Q et M_1 à M_{11} ont une zone de rejet donnée par $[1, \infty]$. Q est une moyenne pondérée des 11 statistiques M_1 à M_{11} qui mesurent la qualité de l'ajustement saisonnier sous différents aspects, que l'on retrouve au tableau F3 de l'output, cf [8, p.158-162]. Le nombre de critères M_1 à M_{11} supérieurs à 1 est donné par N_c. Nous détaillons plus bas la statistique M_7 qui a le poids maximum dans le calcul de Q.

– M_1 et M_5 testent si la composante irrégulière a une variance acceptable relativement à la variance totale. Ce sont les plus critiques des 11 statistiques dans le cas de l'exemple 4, c'est pourquoi nous les avons reportées. Ces statistiques indiquent que l'ajustement saisonnier pourrait être instable, à cause de la trop grande variance de la composante irrégulière. Pour tester la sensibilité de M_1 et M_5, nous avons refait l'analyse dans le cas du modèle 2, mais en divisant par 2 la composante n_t de la série x_t. Nous obtenons alors $M_1 = 0.383$, $M_5 = 0.853$ et $Q = 0.43$. L'ajustement saisonnier est donc accepté sans problème dans ce cas, et cela confirme l'interprétation des statistiques M_1 et M_5.

Il y a deux options pour l'estimation de la vraisemblance des paramètres du modèle Arima. Par défaut, le programme calcule la vraisemblance exacte, et cette méthode est utilisée dans tous les exemples de la table 1.2. L'autre option est de calculer la vraisemblance conditionnelle (sur les premières observations). Pour comparer, nous donnons ci-dessous les résultats pour les deux méthodes de calcul de la vraisemblance dans le cas du modèle 2.

Ces résultats résumés sont fournis par le programme X12diag, appelé après X12a. Le fichier de diagnostics ci-dessous montre - dans le cas de notre exemple - qu'il y a très peu de différence au niveau des critères de désaisonnalisation entre les deux méthodes de calcul de la vraisemblance. Pourtant l'ajustement par la vraisemblance conditionnelle possède ici l'avantage que les autocorrélations et les autocorrélations partielles des résiduelles sont toutes non-significatives.

Dans le cas de la vraisemblance exacte, l'autocorrélation et autocorrélation partielle d'ordre 12 sont significatives (fig. 1.7). Cela implique qu'une certaine saisonnalité subsiste dans les résiduelles. Le dernier panneau présente les autocorrélations des carrés des résiduelles. Celles-ci sont toutes non significatives, ainsi il n'y a pas de non-linéarité dans les résiduelles.

Exemple 4, fichier de diagnostics pour le modèle No 2 de la table 1.2

Series name :	xt0
Span used :	1st month,2000 to 12th month,2011

X-11 Seasonal Adjustment

Seasonal filter length :	M.S.R.
Seasonal filter length selected by MSR :	3x5
Trend filter length :	default
X-11 Extreme adjustment :	Standard Error
Type of Adjustment :	additive seasonal adjustment

Diagnostics for direct seasonally adjusted series	vrais. exacte	vrais. cond.
F-test for Stable Seasonality (D8) :	24.677	25.610
F-test for Moving Seasonality (D8) :	0.436	0.465

Identifiable Seasonality present
No evidence of residual seasonality in D11.

Relative Contribution of the Irregular to the Variance (F2.F :)[2]

	I	C	S	P	TD	Total
vrais. exacte	9.51	45.98	47.25	6.55	0.00	109.29
vrais. cond.	8.50	46.63	48.32	8.94	0.00	112.39

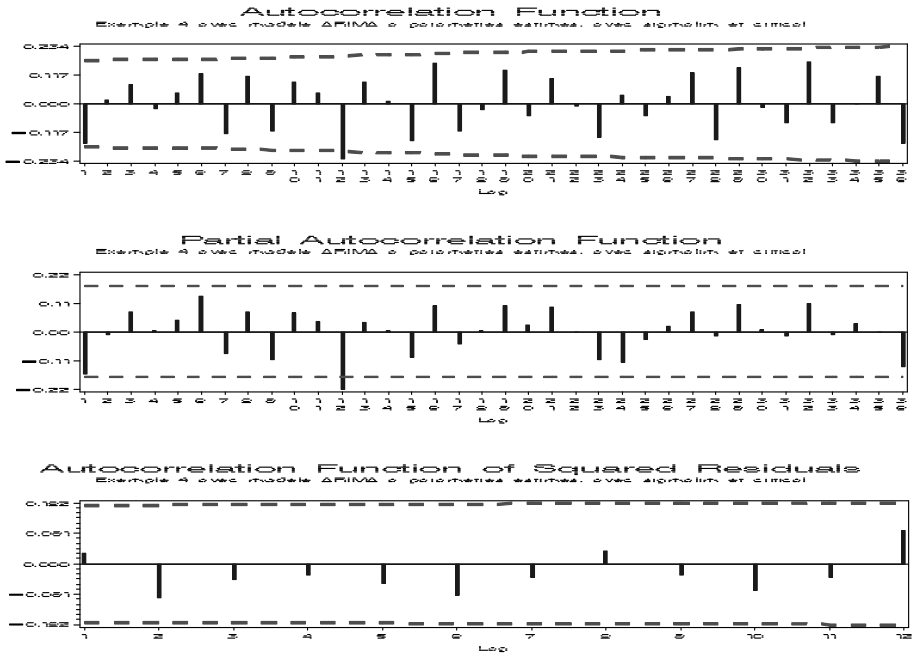
	I/C RATIO	I/S RATIO
vrais. exacte	3.19	5.94
vrais. cond.	3.25	5.63

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11
vrais. exacte	1.384	1.018	1.094	0.129	1.619	0.774	0.410	0.577	0.340	0.645	0.602
vrais. cond.	1.259	0.933	1.127	0.051	1.569	0.651	0.405	0.604	0.346	0.650	0.613

	Q	Q (without M2)
vrais. exacte	0.81	0.79
vrais. cond.	0.77	0.75

Visually distinct seasonal spectral peaks were not found.
Visually distinct trading day spectral peaks were not found.

[2]I=irregular component ; C=trend-cycle ; S=seasonal component ; P=preadjustment factors ; TD=trading days effect



Fonctions d'autocorrélation obtenues par la vraisemblance exacte

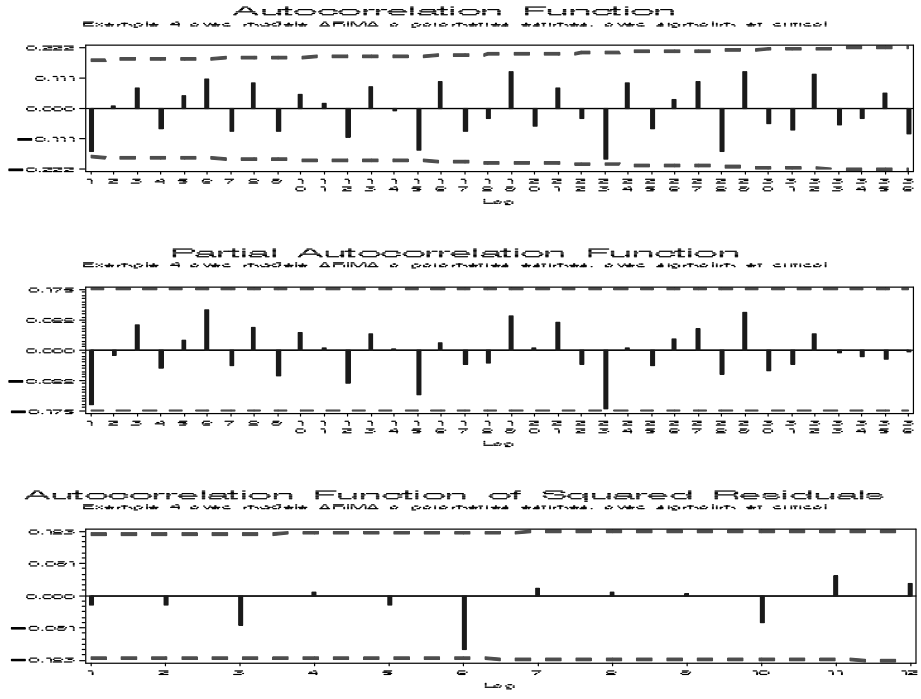


FIG. 1.7: Fonctions d'autocorrélation obtenues par la vraisemblance conditionnelle

L'analyse spectrale propose une autre façon d'évaluer la présence de la saisonnalité dans une série. On en trouve une bonne introduction dans Harvey [4, chap. 6].
La *densité spectrale* - ou *spectre de puissance* - est une fonction continue qui ne dépend que de la fonction d'autocovariance du processus (stationnaire). Elle est donnée par :

$$f(\lambda) = \frac{\sigma^2}{2\pi}\left[1 + 2\sum_{\tau=1}^{\infty}\rho(\tau)\cos\lambda\tau\right] \qquad -\pi \leq \lambda \leq \pi$$

où $\rho(\tau) = \mathbf{cor}\,(x_t, x_{t+\tau})$ est la fonction d'autocorrélation théorique et σ^2 la variance du processus en un instant donné, $\sigma^2 = \mathbf{Var}(x_t)$. On voit que

$$\sigma^2 = \int_{-\pi}^{\pi} f(\lambda)\,d\lambda = 2\int_{0}^{\pi} f(\lambda)\,d\lambda$$

La densité spectrale donc peut être interprétée comme une décomposition de la variance dans des composantes cycliques sur l'intervalle $[0, \pi]$.

Pour un processus ARMA, donné par $\phi(B)x_t = \theta(B)a_t$, où $\{a_t\}$ est un bruit blanc de variance σ_a^2, la densité spectrale peut s'exprimer par :

$$f(\lambda) = \frac{\sigma_a^2}{2\pi}\frac{\left|\theta(e^{-i\lambda})\right|^2}{\left|\phi(e^{-i\lambda})\right|^2} \tag{1.11}$$

où l'argument du polynôme est le nombre complexe $e^{-i\lambda} = \cos(-\lambda) + i\sin(-\lambda)$, et les deux barres verticales indiquent que l'on prend le module d'un nombre complexe.

Il y a deux façons d'estimer la densité spectrale d'un processus ARMA :

- par le spectre estimé $\widehat{f}$, où les paramètres θ_i, ϕ_j, et σ_a^2 sont remplacés dans la formule ci-dessus par leurs valeurs estimées.
- par le périodogramme. Soit N la longueur de la série. Les fréquences fondamentales sont données par $2\pi\ell/N,\ \ell = 1, .., n-1$, avec $n = [N/2]$. On peut alors représenter la série comme une combinaison linéaire de sinus et cosinus calculés à ces différentes fréquences fondamentales. C'est la représentation spectrale de la série x_t :

$$x_t = \frac{1}{\sqrt{N}}\left(a_0 + 2\sum_{\ell=1}^{n-1}\left(a_\ell \sin\frac{2\pi\ell}{N}t\ + b_\ell\cos\frac{2\pi\ell}{N}t\right) + a_n(-1)^t\right) \qquad t = 1, ..., N$$

Le périodogramme I est donné par

$$\begin{aligned} I\,(\ell/N) &= a_\ell^2 + b_\ell^2 \qquad \ell = 1, .., n \\ \sigma^2 &= \sum_{\ell=1}^{n} a_\ell^2 + b_\ell^2 \end{aligned}$$

Intuitivement, un périodogramme décompose la variance d'une série selon les contributions des différentes fréquences fondamentales.
Dans X-12-ARIMA (voir Findley et al [3]),

- le spectre estimé, exprimé en décibels, $10\log_{10}\left(\widehat{f}\left(2\pi\omega\right)\right)$ est représenté en fonction de $\omega,\ \omega \in \left[0, \frac{1}{2}\right]$.
- le périodogramme, exprimé en décibels, $10\log_{10}\left(I\left(\omega_\ell\right)\right)$ est représenté en fonction de $\omega_\ell = \frac{\ell}{N},\ \omega_\ell \in \left[0, \frac{1}{2}\right]$.

Le spectre n'existe que pour un processus stationnaire. Pour les processus ARIMA, on peut définir des pseudo-spectres, qui possèdent certaines propriétés des spectres. Tramo-Seats travaille avec ces pseudo-spectres, mais X-12-ARIMA ne les utilise pas. L'hypothèse de stationnarité est une hypothèse mathématique, qui n'a de sens que pour une série de longueur infinie. On ne peut donc pas la vérifier empiriquement. Une façon grossière d'évaluer l'existence d'une saisonnalité - qui est implémentée dans X-12-ARIMA - est d'estimer un modèle ARIMA(30,1,0),sur la série originale $\{x_t\}$ et de représenter le spectre de $\{\nabla x_t\}$ Autrement dit, on estime un modèle autorégressif d'ordre 30 sur les premières différences de la série originale $\{\nabla x_t\}$ et on représente le spectre de ce modèle [3]. Le spectre obtenu de cette façon n'a donc rien à voir avec le modèle RegARIMA estimé et utilisé pour la décomposition de la série.

La fig. 1.8 présente trois spectres. Le premier spectre de la figure d'en haut (trait plein) est basé sur un modèle AR(30) estimé pour la série $\{\nabla x_t\}$. C'est une façon rapide d'évaluer si la série est saisonnière, en ôtant par l'opérateur ∇ une tendance approximativement linéaire. Ce spectre ne permet que d'évaluer l'existence éventuelle de pics à des fréquence prédéfinies. Seuls les pics à des fréquences saisonnières ($k/12,\ k = 1, ..., 5$ pour des données mensuelles) et les fréquences correspondant à des jours ouvrables (trading day frequencies) doivent être interprétés. Ainsi, à la fig. 1.8, on voit un pic à la fréquence 1/12, ce qui indique la présence d'une saisonnalité. Pour comparer, le même spectre basé sur un modèle AR(30) et estimé sur les premières différences de la série désaisonnalisée est représenté en traitillé. On voit clairement que le pic à 1/12 a disparu. Le spectre des résiduelles du modèle RegArima se trouve à la fig 1.8, en bas[4]. Les fig. 1.8 et 1.9 permettent de comparer les résultats lorsque l'estimation des paramètres du modèle RegArima est faite par la vraisemblance exacte et conditionnelle. Le spectre des premières différences de la série originale est évidemment le même. On voit que le spectre de la série désaisonnalisée ne change pratiquement pas d'une figure à l'autre. Il y a quelques différences entre les spectres des résiduelles du modèle RegArima, en particulier à la fréquence 4/12. Comme X-12-Arima n'utilise ce modèle que pour prolonger la série, et non pour la décomposition , ces différences n'ont que peu d'effets sur le spectre de la série désaisonnalisée.

[3] Mathématiquement, il s'agit d'une approximation rationnelle de la densité spectrale, si on admet que $\{\nabla x_t\}$ est stationnaire ; voir Brockwell, P. J. & Davis, R.A. (1991) *Time Series : Theory and Methods (2nd ed)*, p.365-7.

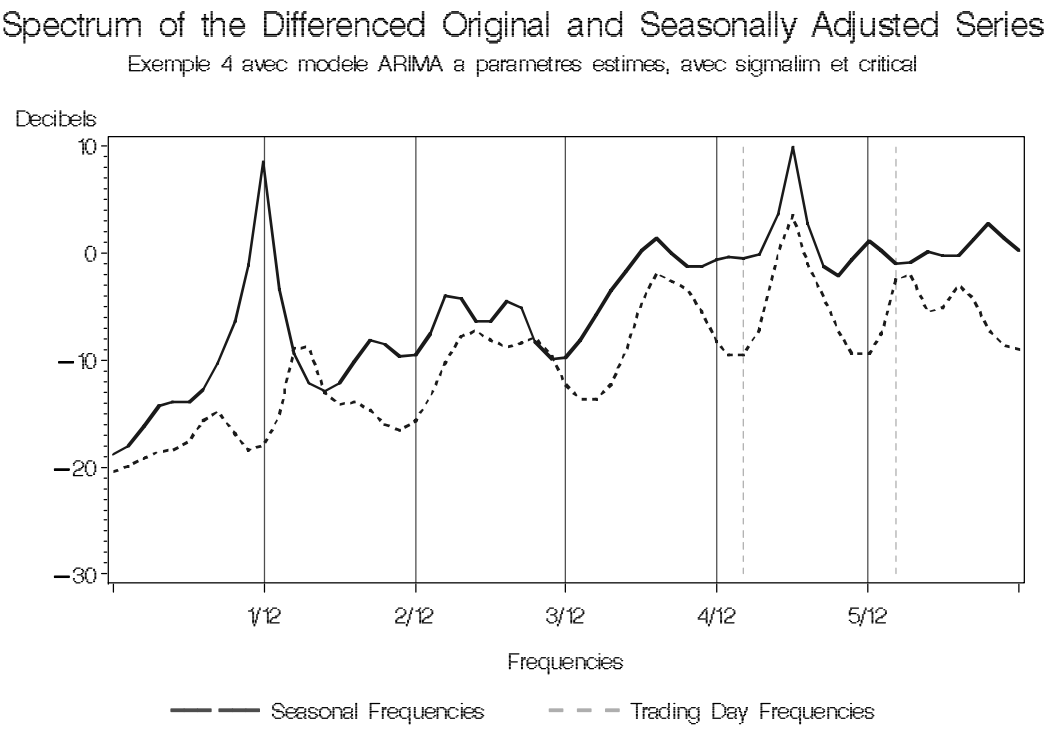
[4] Comme ce spectre ne présente pas de pics aux fréquences prédéfinies, on voit que le modèle estimé (ARIMA(0,2,2)$(0,1,1)_{12}$) contient implicitement la désaisonnalisation. La méthode Tramo-Seats est basée sur la décomposition du spectre du modèle estimé.

La fig. 1.10 montre les résultats des meilleurs modèles. On voit que la série désaisonnalisée est très semblable, que l'on introduise la saisonnalité par un modèle ARIMA saisonnier (comme dans le modèle 2), ou par des régresseurs trigonométriques (modèle 10). La partie du bas est subdivisée en 4 panneaux montrant la série désaisonnalisée, puis au 2e panneau, la série originale ajustée pour les 4 valeurs extrêmes (outliers additifs) trouvées lors de l'estimation par la vraisemblance conditionnelle. Sur le 3e panneau les valeurs extrêmes sont représentées sur l'axe du temps avec en ordonnée, le coeffcient de régression estimé par RegArima, et enfin, sur le 4e panneau, on voit la composante irrégulière.

La fig. 1.11 compare les performances du modèle 2, lorsque la variance du bruit blanc $\{n_t\}$ est de $\frac{1}{4}$ (en haut) ou de $\frac{1}{16}$ (en bas). La composante saisonnière $\{s_t\}$ et la tendance $\{p_t\}$ de la série $\{x_t\}$ ne sont pas modifiées. Dans le panneau 1, on voit la composante saisonnière extraite par X-12-Arima, en fonction de $\{s_t\}$. Les points $(s_t, \widehat{s}_t)_{t=1,\ldots,144}$ sont approximativement répartis autour de la première bissectrice, cela indique qu'il n'y a pas de biais dans l'estimation de la composante saisonnière. Au panneau 2 (intitulé "serie : zt0" et "serie : zt111" respectivement) se trouve la fonction d'autocorrélation (ACF) de $\{s_t - \widehat{s}_t\}$. On voit que la désaisonnalisation donne des résultats beaucoup moins satisfaisants lorsque la variance du bruit blanc est de $\frac{1}{4}$, plutôt que de $\frac{1}{16}$. En effet l'autocorrélation d'ordre 12, c'est-à-dire la corrélation de $s_t - \widehat{s}_t$ avec $s_{t-12} - \widehat{s}_{t-12}$, est significative dans le premier cas, mais pas dans le second. La même remarque s'applique pour les autocorrélations partielles (PACF au panneau 4)[5]. Une autocorrélation ou une autocorrélation partielle significative à des multiples de la fréquence indique qu'il reste une saisonnalité dans $\{s_t - \widehat{s}_t\}$ et implique donc que la désaisonnalisation n'est pas parfaitement réalisée.

Cette vérification n'est naturellement possible que pour un exemple construit, où la composante saisonnière est connue à l'avance. Lors de la désaisonnalisation par X-12-Arima, les statistiques M_1 et M_5 comparent la variance relative de la composante irrégulière par rapport à la variance totale. Dans l'exemple que nous avons présenté, où la variance du bruit blanc vaut $\frac{1}{4}$, ces statistiques dépassent la valeur critique de 1. Le programme a donc pu détecter le problème.

[5]Rappelons que l'autocorrélation partielle entre x_t et $x_{t-\tau}$ est l'autocorrélation conditionnée par les valeurs du processus aux instants intermédiaires. Une autocorrélation partielle d'ordre τ nulle signifie que la mémoire du processus est plus courte que τ.



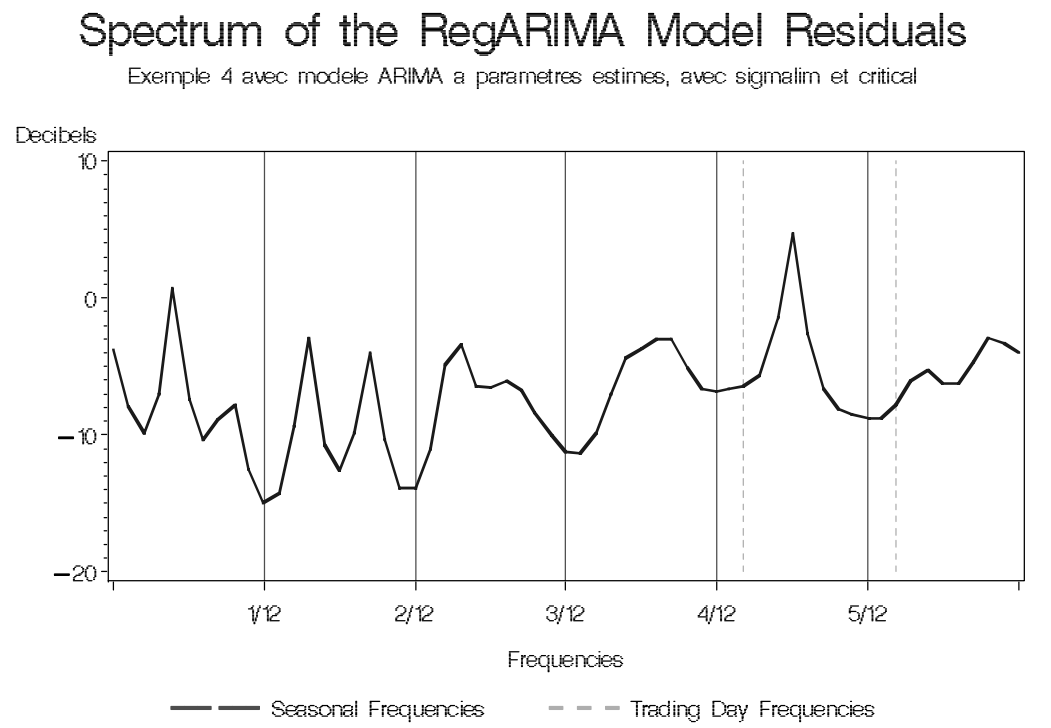
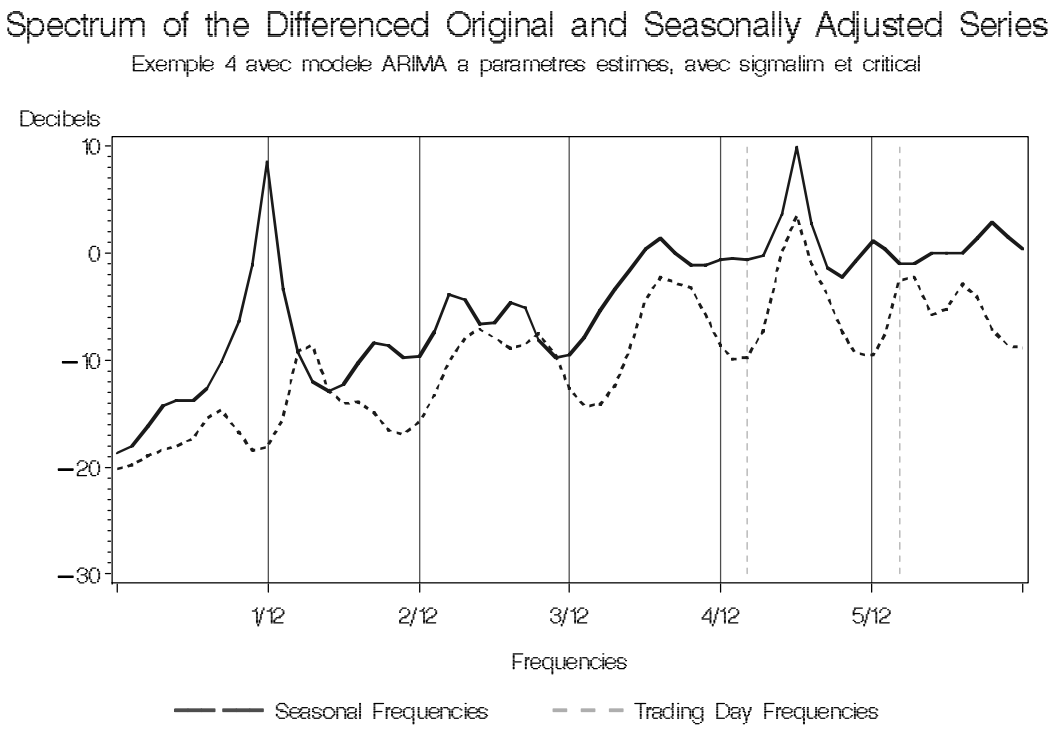


FIG. 1.8: Spectres du modèle 2 pour vraisemblance exacte



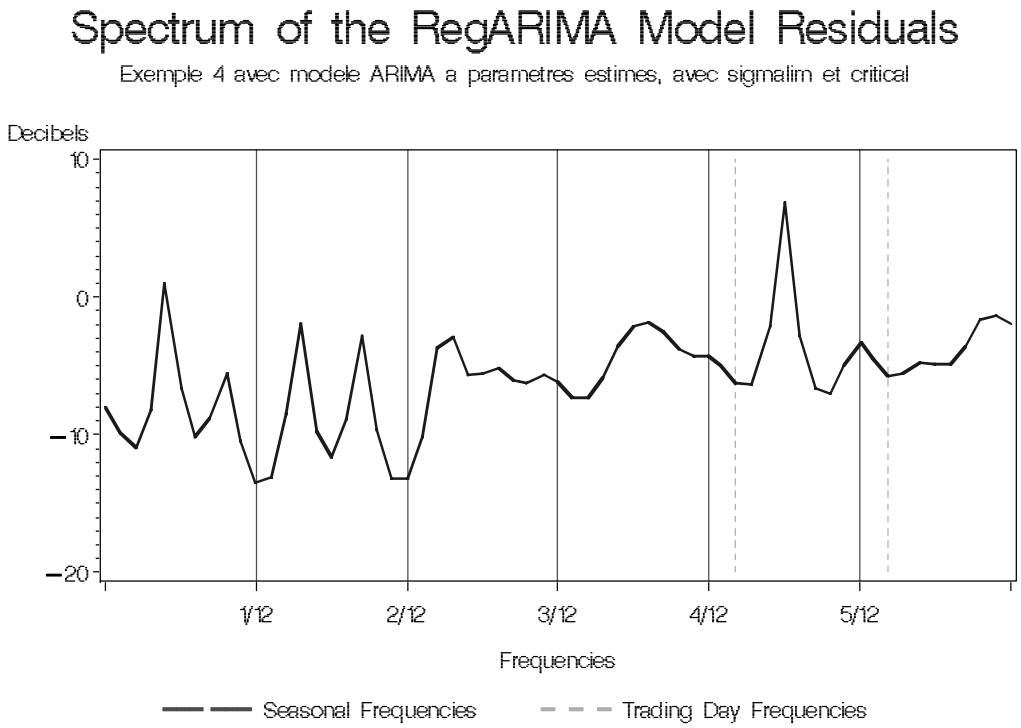
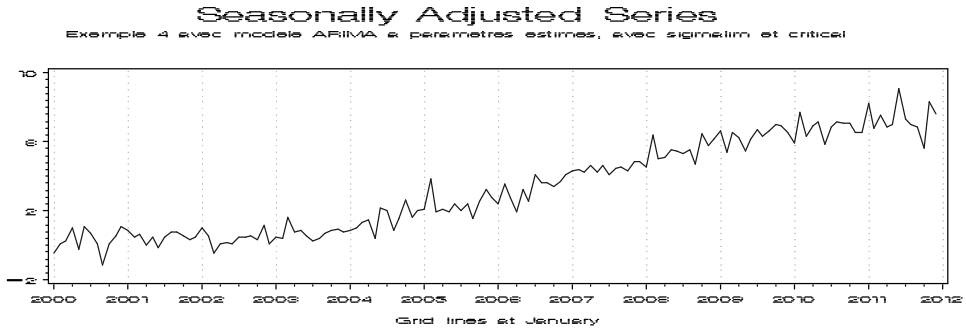
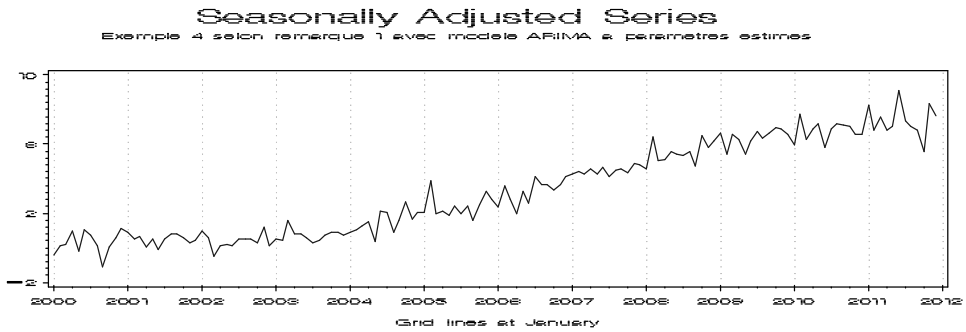


FIG. 1.9: Spectres du modèle 2 pour vraisemblance conditionnelle





Résultats avec vraisemblance exacte - modèle 2 (en haut) et modèle 10 (en bas)

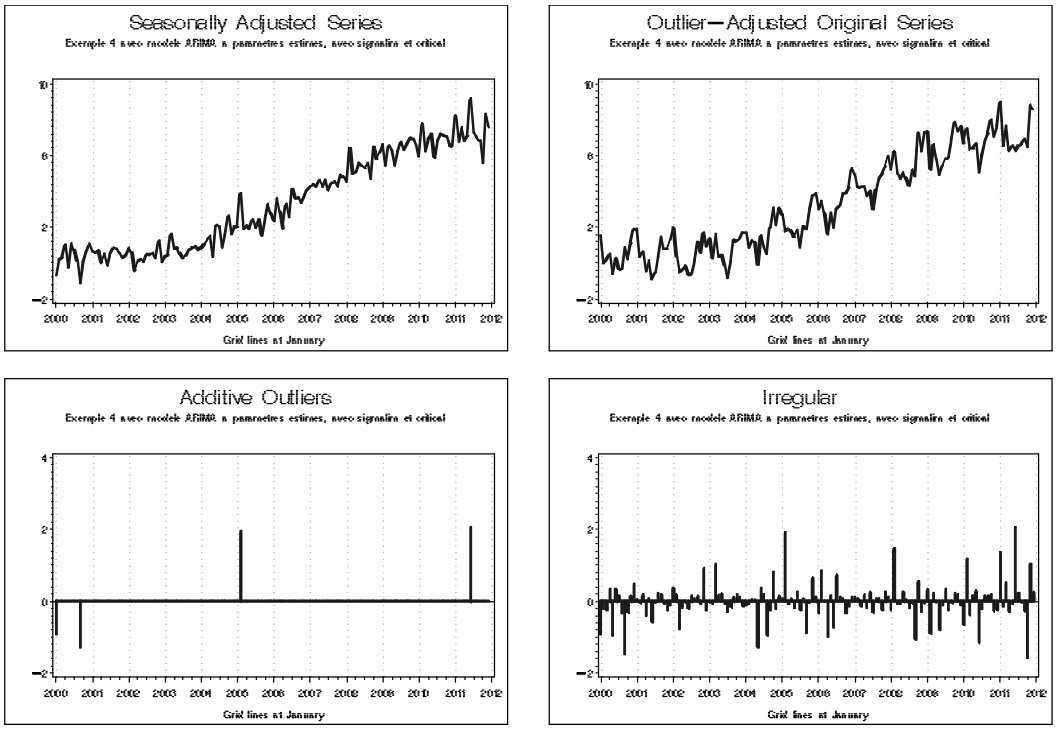


FIG. 1.10: Résultats du modèle 2 avec vraisemblance conditionnelle

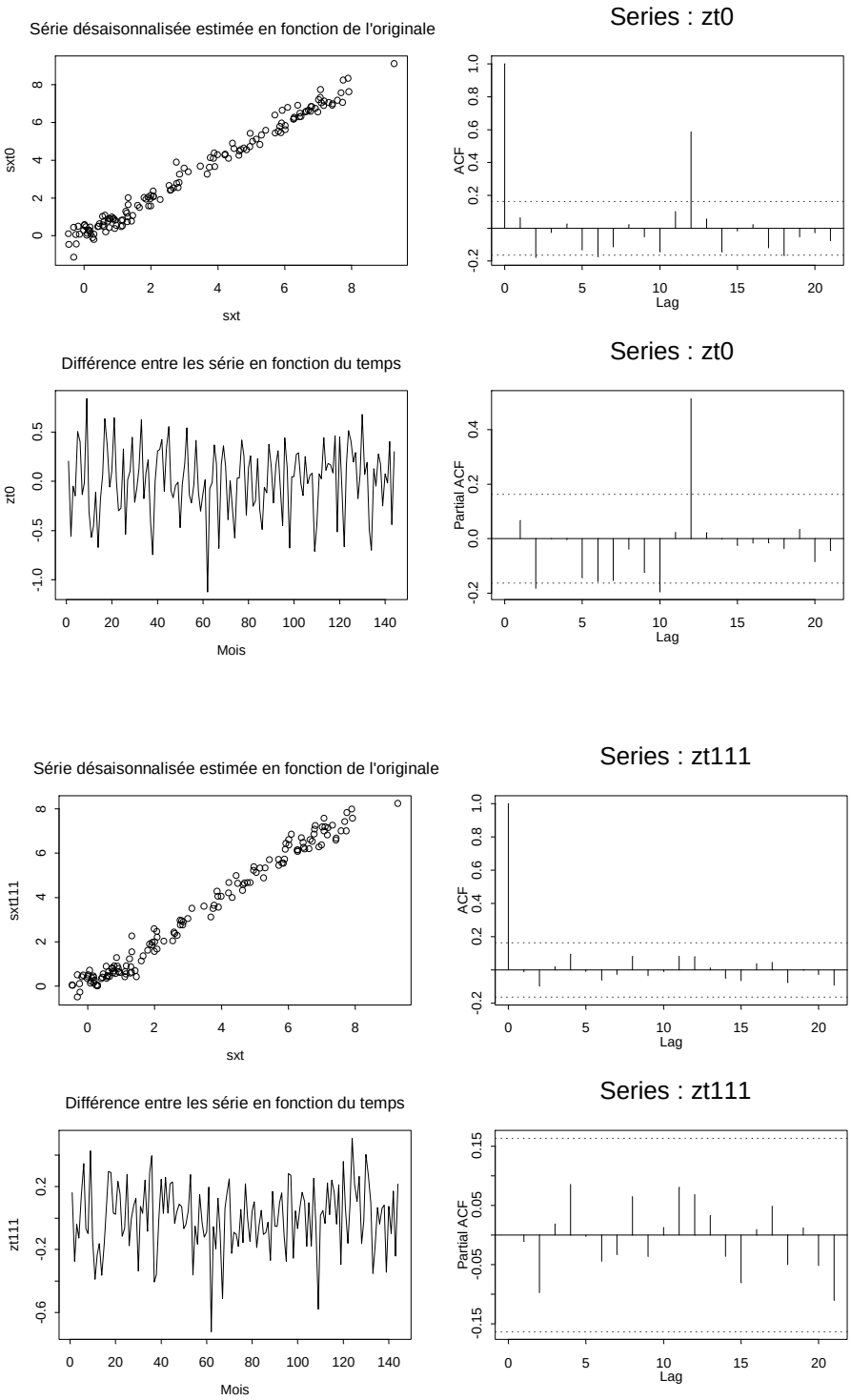


FIG. 1.11: Performances du modèle 2, lorsque la variance de n_t vaut $\frac{1}{4}$ (en haut) et $\frac{1}{16}$ (en bas)

1.4.1 Etude de la statistique M7 (table D 8.A)

Saisonnalité stable

Ce chapitre est une réflexion au sujet de la statistique M_7 qui est celle, parmi les 11 statistiques contribuant au calcul du diagnostic global Q, ayant le poids le plus grand. Cette statistique est définie ci-après.

Dans une ANOVA à un facteur "saison" effectuée sur la composante saisonnière, considérons le test de l'hypothèse H_0 : " pas d'effet saisonnier ". Soit F_S la statistique F associée à ce test.
Dans le cas de données mensuelles, les degrés de liberté de F_S sont 11 et $12(N-1)$, où $N =$ nb d'années de la série.

Recommandation du U.S. Census Bureau : " ne pas désaisonnaliser, si $F_S < 7.0$ "

Cette recommandation implique qu'il faut que la saisonnalité soit très marquée pour qu'on entreprenne une désaisonnalisation, puisque le risque nominal correspondant à la valeur critique de 7 est très faible :

N	3	4	5	6	7	8
Risque nominal	3.6463e-005	3.646e-006	6.746e-007	1.84e-007	6.6e-008	2.8e-008
N	9	10	11	12	13	14	15
Risque nominal	1.4e-008	7.7e-009	4.7e-009	3.0e-009	2.0e-009	1.4e-009	1.1e-009

On voit que la valeur de 7 correspond à une quasi certitude. (Ceci est raisonnable, car la composante saisonnière a été construite). La table D.8.A du fichier de sortie de X-12-Arima indique si F_S est significative au risque 0.1%. A la table 1.3 ci-dessous, on lit c_S, la valeur critique correspondante. Cette valeur est toujours bien inférieure à 7.

Saisonnalité évolutive

Une condition supplémentaire est encore mise à la désaisonnalisation.
Dans une ANOVA à 2 facteurs additifs "saison" et "année", effectuée sur la composante saisonnière, on considère le test de l'hypothèse H_0 : " pas d'effet année " (dénommé *moving seasonality test*). Soit F_M la statistique F associée à ce test. Dans le cas de données mensuelles, les degrés de liberté de F_M sont $N-1$ et $11(N-1)$; remplacer 11 par 3 pour des données trimestrielles.
On définit encore la quantité M_7, donnée par

$$M_7 = \sqrt{\frac{7 + 3F_M}{2F_S}}$$

Recommandation du U.S. Census Bureau :
" ne pas désaisonnaliser, si F_M est significatif au risque 5% et si $M_7 > 1$ "

Etudions cette recommandation dans le cas limite du premier test, c'est-à-dire supposons que F_S = 7. Alors la condition $M_7 > 1$ est équivalente à $F_M > 7/3 = 2.333333$.
A la table 1.3 , on trouve, au risque nominal de 5% pour F_M,

N	3	4	5	6	7	8	9	10	11	12	13	14
Valeur crit. de F_M	3.44	2.89	2.58	2.38	2.24	2.13	2.05	1.98	1.92	1.87	1.83	1.79
risque si $F_S = 7$	5.0	5.0	5.0	5.0	4.2	3.3	2.5	2.0	1.6	1.2	1.0	0.8
risque si $F_S = 10$	2.6	1.1	0.5	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
c_S	4.51	3.87	3.58	3.42	3.32	3.24	3.19	3.15	3.12	3.09	3.07	3.05
α_M	48	13	1	0	0	0	0	0	0	0	0	0

TAB. 1.3: Aspects de la statistique M_7

- les valeurs critiques de F_M qui sont données par le quantile 0.95 de la loi F à $N-1$ et $11(N-1)$ degrés de liberté.
- le risque réel lié à la recommandation (en %), si $F_S = 7$. Le risque est celui de rejeter à tort H_0 : " il n'y a pas de saisonnalité évolutive " (si H_0 est acceptée, on désaisonnalise). Ainsi, on s'abstiendra de désaisonnaliser plus souvent, si le nombre d'années N est inférieur à 7.
- Le risque réel lié à la recommandation (en %), si $F_S = 10$.
- c_S est la valeur critique de F_S pour un risque de 0.1% (elle est toujours inférieure à 7).
- α_M est, lorsque $F_S = c_S$, le risque effectif (en %) lié au test sur F_M correspondant à la règle de décision "rejet de H_0, si $M_7 > 1$". On voit donc que si la série est d'au moins 6 ans, le risque est nul, c'est-à-dire que M_7 est toujours >1.

On tient d'autant plus compte de la propriété de saisonnalité évolutive que le nombre d'années est petit.
On voit facilement que M_7 est une fonction décroissante de F_S. Le critère de saisonnalité évolutive devient de moins en moins opérationnel, lorsque F_S croît. Illustrons encore ce point avec le cas où $F_S = 10$. Si $F_S = 10$, la condition $M_7 > 1$ est équivalente à $F_M > 13/3 = 4.333333$. Cette limite est supérieure à toutes les valeurs critiques au risque 5% (voir table 1.3). Aussi le risque nominal, associé à l'hypothèse nulle " il n'y a pas de saisonnalité évolutive ", diminue.
Le critère de saisonnalité évolutive n'a d'influence que si la série est très courte.

La figure 1.12 indique, en fonction de la longueur de la série, la valeur critique de F_S, lorsque F_M est égale à différents quantiles (de 0 à 0.995) de la loi F à $N-1$ et $11(N-1)$ degrés de liberté. Pour un quantile donné, la zone d'acceptation (correspondant à $M_7 \leq 1$) se situe au-dessus de la courbe correspondante. Ainsi, plus la saisonnalité évolutive est nette, plus il faudra que la saisonnalité stable soit importante, pour que la composante saisonnière soit acceptée.

On constate donc qu'il y a plusieurs façons d'utiliser M_7, F_S et F_M. On trouvera dans [8, p. 118-119] un arbre de décision utilisant les 3 quantités (dans ce tableau, M_7 est noté T). En pratique dans X-12-Arima, la statistique M_7 est utilisée sans référence explicite à F_S et F_M.

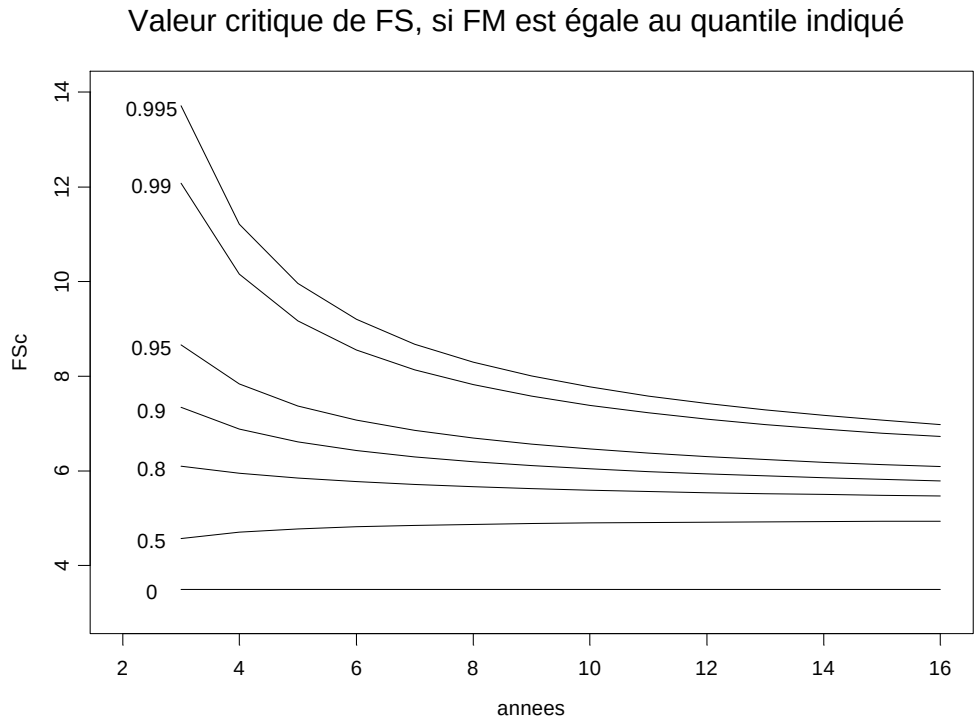


FIG. 1.12: Valeur critique de F_S donnée par la condition $M_7 = 1$

Chapitre 2

Désaisonnalisation de la statistique de l'emploi

2.1 Adaptations aux recensements et aux changements d'échantillon

La statistique de l'emploi (STATEM, allemand BESTA : Beschäftigungsstatistik) et sa problématique sont décrites de façon complète dans les rapports [11] et [16]. Cette série trimestrielle présente une saisonnalité très marquée pour certaines branches d'activité. C'est pourquoi les méthodes de désaisonnalisation prennent, pour l'analyse de l'emploi, un intérêt tout particulier.

La BESTA subit un certain nombre d'adaptations servant à la faire coïncider avec les résultats du recensement des entreprises (RE, allemand BZ : Betriebszählung). Ces aménagements sont de 3 types (voir la note de F. Saucy [15] pour une liste exhaustive, [11], [12] et [13] pour les justifications) :

- Correction d'un saut lors d'un changement d'échantillon
- Interpolation rétrospective entre deux BZ
- Extrapolation.

2.1.1 Correction d'un saut lors d'un changement d'échantillon

Posons :
$B_t =$ valeur de la BESTA au trimestre t.

$$r_t = \frac{B_t - B_{t-1}}{B_{t-1}} \tag{2.1}$$

r_t est donc l'accroissement relatif de la série entre deux trimestres consécutifs. On a :

$$B_t = (1 + r_t)\, B_{t-1} \tag{2.2}$$

Au 1er trimestre 1997, noté $1Q97$, un changement partiel de l'échantillon a eu lieu, dont la conséquence fut un accroissement relatif jugé anormal. Ainsi (dans le cas d'une série saisonnière - voir [13]), r_{1Q97} a été remplacé par la moyenne $\overline{r}_{1Q97}$ des accroissements relatifs au 1er trimestre des deux années adjacentes : $\overline{r}_{1Q97} = \frac{1}{2}(r_{1Q96} + r_{1Q98})$. Posons

$$\begin{aligned} B_t &= \text{valeur de la BESTA non corrigée, } t \geq 1Q97 \\ B_t^c &= \text{valeur de la BESTA corrigée du saut en } t = 1Q97 \end{aligned}$$

$$B^c_{1Q97} = (1 + \overline{r}_{1Q97})B_{4Q96} = \frac{1 + \overline{r}_{1Q97}}{1 + r_{1Q97}} B_{1Q97} = c_{1Q97} B_{1Q97} \tag{2.3}$$

Le facteur de correction, c_{1Q97}, est le rapport du taux de croissance corrigé $(1 + \overline{r}_{1Q97})$ et du taux de croissance observé $(1 + r_{1Q97})$.

Pour les trimestres suivants, on suppose que l'accroissement relatif ne dépend plus du changement d'échantillon, c'est-à-dire :

$$\frac{B^c_t - B^c_{t-1}}{B^c_{t-1}} = \frac{B_t - B_{t-1}}{B_{t-1}} \qquad t > 1Q97$$

On a donc :

$$B^c_t = c_{1Q97} B_t \qquad\qquad t \geq 1Q97 \tag{2.4}$$

2.1.2 Interpolation rétrospective entre deux BZ

La statistique de l'emploi est déjà extrapolée pour représenter l'emploi total. Cependant, les résultats du recensement des entreprises sont considérés comme plus fiables. Aussi a-t-il été choisi de corriger la BESTA rétrospectivement, de telle sorte que la nouvelle série coïncide avec les valeurs disponibles du BZ. Les recensements ont eu lieu aux trimestres $3Q91$, $3Q95$ et $3Q98$.
Considérons d'abord l'adaptation pendant la période $3Q91 \leq t \leq 3Q95$. Posons :

$$a_{95} = \frac{BZ_{95}}{BZ_{91}} \qquad b_{95} = \frac{B_{3Q95}}{B_{3Q91}}$$

qui sont, respectivement, les taux de croissance globaux du BZ et de la BESTA sur la période de 16 trimestres séparant le 3e trimestre 91 du 3e trimestre 95. Appelons $\widehat{BZ}_t$ la série adaptée. On

peut écrire[1] :

$$\begin{aligned} b_{95} &= \prod_{t=4Q91}^{3Q95} (1+r_t) \\ a_{95} &= \prod_{t=4Q91}^{3Q95} (1+s_t) \end{aligned}$$

où r_t est donné par (2.1) et s_t est le taux d'accroissement relatif (non observé) de la série trimestrielle que l'on aurait, si le recensement des entreprises était effectué chaque trimestre. Remarquons que $\sqrt[16]{b_{95}}$ est la moyemnne géométrique des taux d'accroissement relatifs trimestriels de la BESTA sur la période allant du 3e trimestre 91 au 3e trimestre 95.

Hypothèse de travail *Le taux de croissance trimestriel qu'on observerait, si le recensement était effectué chaque trimestre, est proportionnel au taux de croissance observé de la BESTA. Le facteur de proportionnalité est donné par le rapport des moyennes géométriques des taux de croissance des deux séries sur la période séparant deux recensements.*

Pour la période 4Q91 à 3Q95, on a donc :

$$1+\widehat{s}_t = \sqrt[16]{\frac{a_{95}}{b_{95}}}\,(1+r_t) \tag{2.5}$$

Par définition, on a $B_t = B_{3Q91} \prod_{\tau=4Q91}^{t} (1+r_\tau)$. Définissons, par analogie,

$$\begin{aligned} \widehat{BZ}_t &= BZ_{91} \prod_{\tau=4Q91}^{t} (1+\widehat{s}_\tau) \\ &= BZ_{91} \prod_{\tau=4Q91}^{t} \left\{ \sqrt[16]{\frac{a_{95}}{b_{95}}}\,(1+r_\tau) \right\} \\ &= BZ_{91} \left(\sqrt[16]{\frac{a_{95}}{b_{95}}} \right)^{t-3Q91} \prod_{\tau=4Q91}^{t} (1+r_\tau) \end{aligned}$$

donc :

$$\widehat{BZ}_t = \frac{BZ_{91}}{B_{3Q91}} \left(\sqrt[16]{\frac{a_{95}}{b_{95}}} \right)^{t-3Q91} B_t \qquad 3Q91 \leq t \leq 3Q95 \tag{2.6}$$

On a bien $\widehat{BZ}_{3Q91} = BZ_{91}$ et $\widehat{BZ}_{3Q95} = BZ_{95}$.

[1] $\prod$ représente le produit des termes entre parenthèses.

Autres représentations Posons

$$z_{91} = \frac{BZ_{91}}{B_{3Q91}} \qquad z_{95} = \frac{BZ_{95}}{B_{3Q95}}$$

on peut alors écrire $\widehat{BZ}_t$ sous la forme suivante, qui montre les effets décroissant de la correction en 91 et croissant de celle en 95 :

$$\widehat{BZ}_t = \left(\sqrt[16]{z_{91}}\right)^{3Q95-t} \left(\sqrt[16]{z_{95}}\right)^{t-3Q91} B_t \qquad 3Q91 \leq t \leq 3Q95 \tag{2.7}$$

La représentation la plus pratique pour les calculs, et que nous utiliserons dans la suite, est donnée par (2.8) :

$$\widehat{BZ}_t = z_{91} \left(\sqrt[16]{z_{95}/z_{91}}\right)^{t-3Q91} B_t \qquad 3Q91 \leq t \leq 3Q95 \tag{2.8}$$

Puisque $z_{95}/z_{91} = a_{95}/b_{95}$, on voit que (2.6), (2.7) et (2.8) sont des formulations équivalentes.

Le même raisonnement s'applique pour la deuxième adaptation sur la période de 12 mois séparant les recensements de 1995 et et 1998. Posons donc

$$z_{98} = \frac{BZ_{98}}{B_{3Q98}} \tag{2.9}$$

Appelons $\widehat{BZ}_t$ la série adaptée. Sans tenir compte du saut au trimestre $1Q97$, on aurait

$$\widehat{BZ}_t = z_{95} \left(\sqrt[12]{z_{98}/z_{95}}\right)^{t-3Q95} B_t \qquad 3Q95 \leq t \leq 3Q98$$

(On obtient bien $\widehat{BZ}_{3Q95} = BZ_{95}$ et $\widehat{BZ}_{3Q98} = BZ_{98}$.)

Cependant, $B_t^c = c_{1Q97} B_t$, $t \geq 1Q97$. Il faut en tenir compte :

$$z_{98}^c = \frac{BZ_{98}}{B_{3Q98}^c} = \frac{BZ_{98}}{c_{1Q97} B_{3Q98}} = \frac{z_{98}}{c_{1Q97}}$$

Les références ([11], [12] et [13]) ne précisent pas comment la correction du saut est prise en compte au moment de l'adaptation au BZ98. En comparant les valeurs non corrigées avec les valeurs reconstituées des branches saisonnières (Données de EWL [2]), on peut retrouver empiriquement la transformation adoptée. En adaptant les formules ci-dessus, on peut exprimer cette transformation mathématiquement.

$$\widehat{BZ}_t = \begin{cases} z_{95} \left(\sqrt[12]{z_{98}^c/z_{95}}\right)^{t-3Q95} B_t & 3Q95 \leq t < 1Q97 \\ z_{95} \left(\sqrt[12]{z_{98}^c/z_{95}}\right)^{t-3Q95} c_{1Q97} B_t & 1Q97 \leq t \leq 3Q98 \end{cases} \tag{2.10}$$

On voit que la double adaptation - au changement d'échantillon du 1Q97 et aux BZ 95 et 98 :

- laisse intégralement le saut ponctuellement au 1Q97,
- modifie les taux de croissance uniformément sur toute la période : $\widehat{BZ}_t / \widehat{BZ}_{t-1} = \sqrt[12]{z_{98}^c/z_{95}} \cdot B_t/B_{t-1} \quad (t \neq 1Q97)$.

[2] Voir les données de la branche 45 en Annexe 1.

2.1.3 Adaptation au nouvel échantillon du deuxième trimestre 2000

Un renouvellement de l'échantillon d'entreprises a eu lieu au deuxième trimestre 2000. A cette date, nous disposons de deux valeurs BESTA, basées sur l'ancien ($B^{(a)}_{2Q00}$) et le nouvel échantillon ($B^{(n)}_{2Q00}$). Pour adapter la BESTA à ce changement, une correction analogue aux précédentes est introduite : elle répartit la valeur relative de l'emploi du nouvel échantillon par rapport à l'ancien sur les 7 trimestres séparant le recensement de 98 de ce changement d'échantillon. Posons :

$$z_{98} = \frac{BZ_{98}}{B_{3Q98}} \qquad b_{00} = \frac{B^{(n)}_{2Q00}}{B^{(a)}_{2Q00}}$$

$$\widehat{BZ}_t = z_{98} \left(\sqrt[7]{b_{00}/z_{98}} \right)^{t-3Q98} B_t \qquad 3Q98 \leq t \leq 2Q00 \tag{2.11}$$

Remarquons que si l'on écrit $b^c_{00} = b_{00}/c_{1Q97}$, on a $b^c_{00}/z^c_{98} = b_{00}/z_{98}$ et $z^c_{98} c_{1Q97} B_t = z_{98} B_t$, si bien que $\widehat{BZ}_t$ peut s'écrire de manière équivalente à l'aide des valeurs corrigées du saut au 1Q97 :

$$\widehat{BZ}_t = z^c_{98} \left(\sqrt[7]{b^c_{00}/z^c_{98}} \right)^{t-3Q98} c_{1Q97} B_t \qquad 3Q98 \leq t \leq 2Q00$$

On n'a donc plus à s'inquiéter de ce saut après l'adaptation au BZ98.

2.1.4 Modèle global d'adaptation

La table 2.1 récapitule les notations.
Le modèle peut s'écrire de façon beaucoup plus simple, si l'on effectue une transformation logarithmique. Dans cette optique, nous allons simplifier les notations. Par exemple, $\ln\left(\sqrt[16]{\frac{z_{95}}{z_{91}}} \right) = (\ln(z_{95}) - \ln(z_{91}))/16$. Posons donc

$$\begin{array}{llll} x_t = \ln\left(\widehat{BZ}_t\right) & u_{97} = \ln(c_{1Q97}) & & \\ c_{91} = \ln(z_{91}) & c_{95} = \ln(z_{95}) & c_{98} = \ln(z_{98}) & c'_{00} = \ln(b_{00}) \\ d_{95} = (c_{95} - c_{91})/16 & d_{98} = (c_{98} - c_{95})/12 & d'_{00} = (c'_{00} - c_{98})/7 & \end{array}$$

Alors la formule (2.12) est une description complète du modèle d'adaptation de la BESTA au BZ pour une branche saisonnière :

$$x_t = \ln(B_t) + \begin{cases} c_{91} + (t - 3Q91)\ d_{95} & 3Q91 \leq t \leq 3Q95 \\ c_{95} + (t - 3Q95)\left(d_{98} - \frac{1}{12}u_{97}\right) & 3Q95 \leq t < 1Q97 \\ c_{95} + u_{97} + (t - 3Q95)\left(d_{98} - \frac{1}{12}u_{97}\right) & 1Q97 \leq t \leq 3Q98 \\ c_{98} + (t - 3Q98)\ d'_{00} & 3Q98 \leq t \leq 2Q00 \\ 0 & t \geq 2Q00 \end{cases} \tag{2.12}$$

B_t	valeur de la BESTA non corrigée au trimestre t
$\widehat{BZ}_t$	valeur de la BESTA adaptée au BZ et aux changements d'échantillon
$r_t = \frac{B_t - B_{t-1}}{B_{t-1}}$	accroissement relatif trimestriel de B_t
$1 + r_t = \frac{B_t}{B_{t-1}}$	taux d'accroissement trimestriel de B_t
$\widehat{s}_t = \frac{\widehat{BZ}_t - \widehat{BZ}_{t-1}}{\widehat{BZ}_{t-1}}$	accroissement relatif trimestriel de $\widehat{BZ}_t$
$1 + \widehat{s}_t = \frac{\widehat{BZ}_t}{\widehat{BZ}_{t-1}}$	taux d'accroissement trimestriel de $\widehat{BZ}_t$
$B_{3Q91,} B_{3Q95}, B_{3Q98}$	valeur de B_t au 3e trimestre 1991,1995,1998
$BZ_{91}, BZ_{95}, BZ_{98}$	valeur du BZ en 1991,1995, 1998
$a_{95} = \frac{BZ_{95}}{BZ_{91}}, (a_{98})$	taux de croissance total du BZ entre les recensements 91 et 95, (95 et 98)
$b_{95} = \frac{B_{3Q95}}{B_{3Q91}}, (b_{98})$	taux de croissance total de la BESTA entre $3Q91$ et $3Q95$, ($3Q95$ et $3Q98$)
$z_{91} = \frac{BZ_{91}}{B_{3Q91}}, (z_{95}, z_{98})$	facteurs de passage BESTA -> BZ
$\sqrt[16]{\frac{a_{95}}{b_{95}}} = \sqrt[16]{\frac{z_{95}}{z_{91}}}$	facteur de proportionnalité des taux de croissance trimestriels de $\widehat{BZ}_t$ et B_t pendant les 16 premiers trimestres
$c_{1Q97} = \frac{1+(r_{1Q96}+r_{1Q98})/2}{1+r_{1Q97}}$	correction du saut de la BESTA au $1Q97$
$z^c_{98} = \frac{z_{98}}{c_{1Q97}}$	facteur de passage BESTA corrigée du saut -> BZ
$\sqrt[12]{(z^c_{98}/z_{95}}$	facteur de proportionnalité des taux de croissance trimestriels de $\widehat{BZ}_t$ et B_t pour $t \neq 1Q97, \quad 3Q95 \leq t \leq 3Q98$
$b_{00} = \frac{B^{(n)}_{2Q00}}{B^{(a)}_{2Q00}}$	facteur de passage de l'ancien au nouvel échantillon au $2Q00$
$\sqrt[7]{\frac{b_{00}}{z_{98}}} = \sqrt[7]{\frac{B^{(n)}_{2Q00}}{B^{(a)}_{2Q00}} \frac{B_{3Q98}}{BZ_{98}}}$	facteur de proportionnalité des taux de croissance trimestriels adaptés au nouvel éch. pendant la période $4Q98 \leq t \leq 2Q00$

TAB. 2.1: Récapitulation des notations

n	91	95	98		2000
BZ_n	382732	339454	293893	$B^{(n)}_{2Q00}$	300090
$B_{3Q\,n}$	387783	329236	322281	$B^{(a)}_{2Q00}$	322858
c_n	-0.0131	0.0306	-0.0922	c'_{00}	-0.0731
d_n		0.0027	-0.0102	d'_{00}	0.0027
$c_{1Q97} =$	0.9780	$u_{97} =$	-0.0222		

TAB. 2.2: Statistiques pour la branche 45

On peut aussi exprimer la transformation à l'aide des différences : $\ln(B_t) - \ln(B_{t-1}) = \ln(B_t/B_{t-1}) = \ln(1+r_t)$ et $x_t - x_{t-1} = \ln(1+\widehat{s}_t)$. On a

$$\ln(1+\widehat{s}_t) = \ln(1+r_t) + \begin{cases} d_{95} & 4Q91 \leq t \leq 3Q95 \\ d_{98} - \frac{1}{12}u_{97} & 4Q95 \leq t \leq 4Q96 \\ d_{98} + \frac{11}{12}u_{97} & t = 1Q97 \\ d_{98} - \frac{1}{12}u_{97} & 2Q97 \leq t \leq 3Q98 \\ d'_{00} & 3Q98 \leq t \leq 2Q00 \\ 0 & t \geq 2Q00 \end{cases} \tag{2.13}$$

Comme $\widehat{s}_t \simeq \ln(1+\widehat{s}_t)$ et de même $r_t \simeq \ln(1+r_t)$, on peut interpréter les formules (2.13) comme des corrections apportées aux taux d'accroissement relatifs de la BESTA. Ces corrections reviennent à ajouter une constante, différente selon la période, au taux d'accroissement relatif de la statistique de l'emploi, pour estimer le taux correspondant du recensement des entreprises.

Exemple : branche 45 (construction) Pour la branche 45, les quantités introduites sont données à la table 2.2.
Rappelons que a_{95} et a_{98} représentent les taux de croissance globaux du recensement des entreprises, entre 91 et 95 et entre 95 et 98 respectivement ; b_{95} et b_{98} sont les taux correspondants de la statistique de l'emploi au 3e trimestre. On voit que les taux de croissance globaux du BZ ne varient pas dans le même sens que ceux de la BESTA : $a_{98} = 0.87 < a_{95} = 0.89$, mais $b_{98} = 0.98 > b_{95} = 0.85$. La correction au 1Q97 atténue le phénomème, mais sans changer le sens des inégalités : avec la correction du saut, on a $b^c_{98} = b_{98}c_{1Q97} = 0.96$.
Pour la branche 45, on a donc :

$$x_t = \ln(B_t) \begin{cases} -0.0131 + 0.0027\,(t - 3Q91) & 3Q91 \leq t \leq 3Q95 \\ +0.0306 - 0.0084\,(t - 3Q95) & 3Q95 \leq t < 1Q97 \\ +0.0084 - 0.0084\,(t - 3Q95) & 1Q97 \leq t \leq 3Q98 \\ -0.0922 + 0.0027\,(t - 3Q98) & 3Q98 \leq t \leq 2Q00 \\ +0 & t \geq 2Q00 \end{cases} \tag{2.14}$$

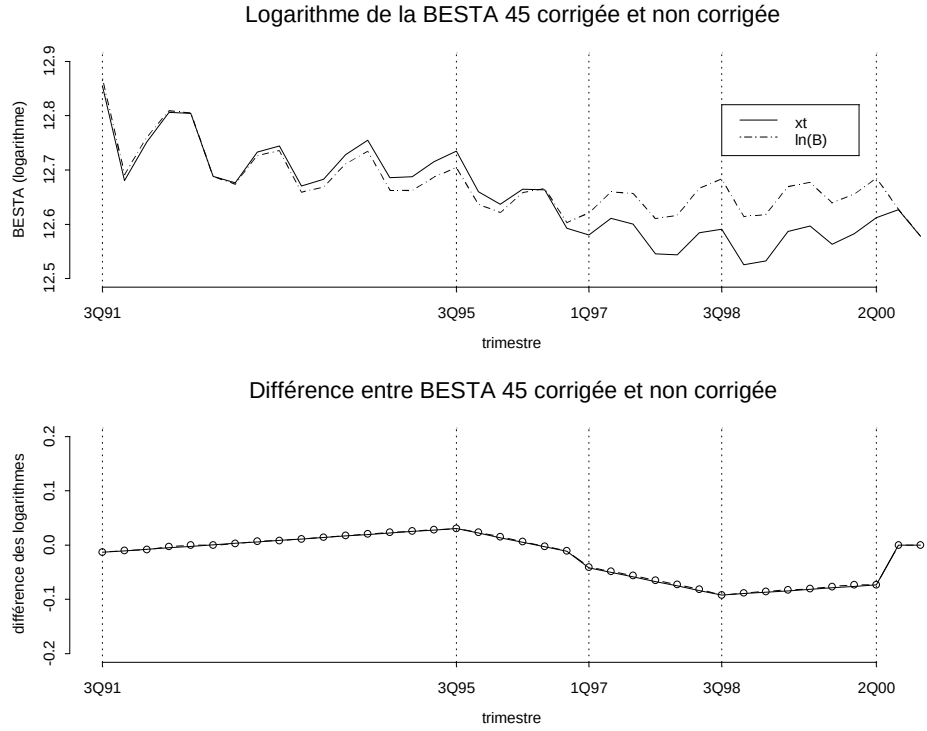


FIG. 2.1: Logarithme de la BESTA corrigée - x_t - et non corrigée - $\ln(B_t)$ - (en haut). Différences observées : o–o et selon formules : — (en bas).

On trouve une représentation graphique de (2.14) à la fig. 2.1. Dans la partie du haut, on voit x_t et $\ln(B_t)$ et dans celle du bas, la différence, donnée par l'expression après l'accolade de (2.14) (ligne continue), et la différence observée (cercles). La fonction mathématique que nous proposons ne recouvre pas exactement les différences $x_t - \ln(B_t)$.
La représentation donnée en (2.13) devient, pour la branche 45 :

$$\ln(1+\widehat{s}_t) = \ln(1+r_t) \begin{cases} +0.0027 & 4Q91 \leq t \leq 3Q95 \\ -0.0084 & 4Q95 \leq t \leq 4Q96 \\ -0.0302 & t = 1Q97 \\ -0.0084 & 2Q97 < t \leq 3Q98 \\ +0.0027 & 3Q98 < t \leq 2Q00 \\ +0 & t > 2Q00 \end{cases} \tag{2.15}$$

On constate que la correction sur les taux trimestriels dans (2.15) est, sauf au 1er trimestre 1997, de l'ordre de quelques millièmes. L'accumulation de ces faibles corrections a cependant un effet non négligeable sur la série elle-même, comme on le voit à la fig. 2.1 (en haut).
Les paramètres correspondant à (2.14) pour les principales branches saisonnières sont donnés à l'Annexe 3.

2.2 Désaisonnalisation de la branche 45

2.2.1 Identification de l'ordre du modèle ARIMA et pré-ajustement

Nous désirons analyser la série adaptée, plutôt que la série brute. Tout d'abord, il nous faut identifier un modèle correct et le plus simple possible. Cela consiste en :

- rechercher l'ordre de différentiation du modèle ARIMA (`identify spec`)
- tester si le modèle global d'adaptation a un effet perturbateur sur la procédure de désaisonnalisation (`regression spec`)
- rechercher les valeurs aberrantes (`outlier spec`).

Identification de l'ordre Rechercher l'ordre de différentiation du modèle ARIMA signifie regarder si le fait de remplacer la série d'origine $\{x_t\}$ par les différences d'ordre d $\left\{\nabla^d x_t\right\}$ rend les données plus compatibles avec un modèle stationnaire. De même, pour la partie saisonnière, rechercher l'ordre D du modèle revient à tester s'il y a avantage à remplacer $\{x_t\}$ par $\left\{\nabla_s^D x_t\right\}$ où $s = 4$ pour des données trimestrielles. On peut combiner les deux recherches, ce qui revient à déterminer d et D, pour évaluer la stationnarité de $\left\{\nabla^d \nabla_s^D x_t\right\}$. X-12-ARIMA possède une procédure d'identification (`identify spec`) recherchant l'ordre de différentiation du modèle.

Dans un premier temps, nous allons travailler sur la série des logarithmes de la BESTA adaptée ($x_t = \ln \widehat{BZ}_t$), car - comme nous l'avons vu précédemment - la structure de la série temporelle des logarithmes est plus simple. Une fois ses propriétés trouvées, cette série est retransformée et le modèle est adapté aux données d'origine.
Nous regardons toutes les combinaisons de $d = 0, 1$ et de $D = 0, 1$. Le choix se base sur l'examen des autocorrélations et des autocorrélations partielles des résiduelles de chaque modèle. La procédure d'appel et quelques détails des résultats d'identification pour la branche 45 se trouvent en Annexe 2. On en conclut qu'il faut prendre $d = D = 1$ pour stationnariser la série. Pour la branche 45, le modèle ARIMA(0,1,0)(0,1,0)$_4$ est même suffisant pour rendre ces résiduelles semblables à un bruit blanc. Nous n'aurons donc pas besoin d'introduire des paramètre autorégressifs et à moyenne mobile.

Pré-ajustement

- Pour tester si le modèle global d'adaptation a un effet perturbateur sur la procédure de désaisonnalisation, nous prenons $\{x_t\}_{3Q91 \leq t \leq 3Q00}$ comme série input du programme X12-ARIMA. Nous introduisons la fonction donnée à droite de l'accolade de (2.14) dans la procédure de régression (`regression spec`) comme une variable de régression définie par l'utilisateur. Les résultats (en Annexe 3) montrent que cette fonction n'est pas significative, si bien que le programme ne l'introduit pas dans la procédure de désaisonnalisation.
- La procédure de détection (`outlier spec`) met en évidence une seule valeur aberrante ponctuelle (AO="additive outlier") au 4e trimestre 1991. Le paramètre de régression correspondant est donné au tableau 2.3. Pour corriger cette valeur, le programme rajoute donc 0.07

	Estimate	Standard Error	t-value
AO1991.4	-0.0700	0.01130	-6.19

TAB. 2.3: Valeurs aberrantes détectées automatiquement à partir du modèle additif sur x_t

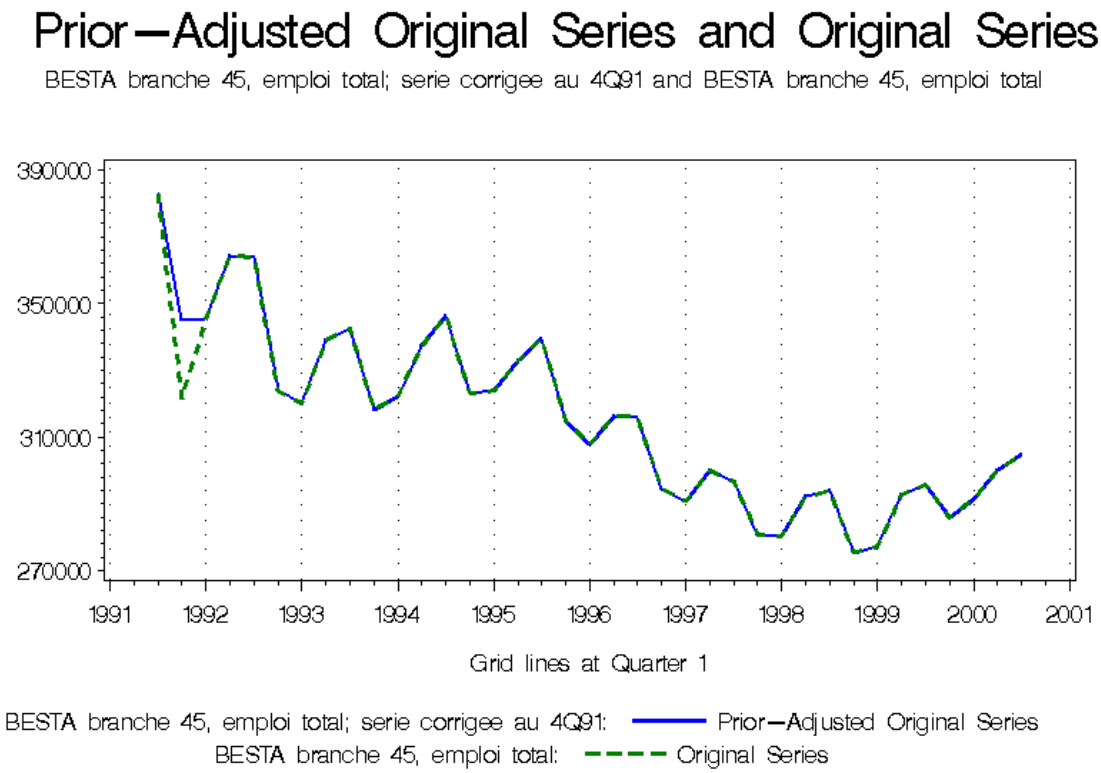


FIG. 2.2: Comparaison des séries originale et corrigée au 4e trimestre 1991

à x_{4Q91}, ce qui revient à multiplier B_{4Q91} par $\exp(0.07) = 1.0725$. La correction est montrée à la fig. 2.2.

La procédure d'appel et un résumé des résultats du pré-ajustement se trouvent en Annexe 3.

Interprétation *On peut donc admettre que la structure du modèle RegARIMA n'est pas plus simple avec la série brute qu'avec la série adaptée aux BZ. Nous traiterons donc dorénavant la série adaptée. Par ailleurs, la valeur x_{4Q91} est détectée comme trop basse (-0.07), relativement aux fluctuations observées ailleurs. Nous corrigerons la série en conséquence.*

2.2.2 Décomposition de la série

Comme nous désirons avoir les résultats dans l'échelle originale (plutôt que logarithmique), nous allons analyser la série $\left\{\widehat{BZ}_t\right\}$ corrigée au 4Q91.
L'analyse préliminaire montre que $\{x_t\}$ admet une décomposition additive en tendance + saisonnalité + composante irrégulière, le modèle analogue pour $\left\{\widehat{BZ}_t\right\}$ est multiplicatif, c'est-à-dire

$$\widehat{BZ}_t = T_t \cdot S_t \cdot I_t \tag{2.16}$$
$$SA_t = T_t \cdot I_t \tag{2.17}$$

Nous faisons la même hypothèse sur la forme du modèle ARIMA que pour la série $\{x_t\}$: $\left\{\widehat{BZ}_t\right\}$ suit un modèle ARIMA(0,1,0)(0,1,0)$_4$.
Le fichier de commandes faisant la désaisonnalisation est donné au tableau 2.4.

Qualité du modèle ARIMA Le tableau 2.5 permet d'évaluer l'importance de l'amélioration de l'ajustement due à la correction au 4Q91. On voit que la variance des résiduelles décroît de $0.62 \cdot 10^8$ à $0.26 \cdot 10^8$. Les critères d'ajustement globaux basés sur la vraisemblance (AIC, AICC etc) montrent aussi l'impact de cette valeur aberrante - qu'il serait naturellement important de pouvoir interpréter. Le lecteur averti pourra s'étonner que le nombre de paramètres estimés soit 1 dans les deux cas. En fait, X12-ARIMA ne permet la détection de valeurs aberrantes qu'avec le modèle additif. (C'est pourquoi nous sommes passés par le logarithme.) Comme nous voulons plutôt ajuster un modèle multiplicatif (2.16), nous avons exécuté X12-ARIMA deux fois sur la série $\left\{\widehat{BZ}_t\right\}$, avec et sans la correction. Le nombre de paramètres estimés est donc de 2, pour le modèle avec correction au 4Q91 (la correction de la valeur aberrante et la variance). Cela ne change pas nos conclusions.

Concernant la qualité du modèle regARIMA de la série corrigée, nous vérifions encore les propriétés distributionnelles des résiduelles.

- Il n'y a pas d'autocorrélations ou d'autocorrélations partielles significatives.
- Extrêmes : $\min$ =-7918 $\max$ =15636 médiane = 1233 (en nombre d'emplois)
- Ecart-type robuste = 5200
- Kurtosis (non évaluable avec moins de 50 observations)

Interprétation *La différence du nombre d'emplois entre deux trimestres consécutifs ne se distingue de la différence correspondante un an avant que par une fluctuation aléatoire imprévisible. L'écart-type de cette fluctuation est 5200 emplois.*

La fig. 2.3 montre le résultat de l'appel à X11 (`X11 spec`) sur la série prolongée selon le modèle ARIMA(0,1,0)(0,1,0)$_4$. La tendance $\{T_t\}$ se trouve dans le premier panneau, en-dessous, on a les facteurs saisonniers $\{S_t\}$ et la composante irrégulière $\{I_t\}$. A droite en haut, on voit la série désaisonnalisée $\{SA_t\}$.

```
series{
title = “BESTA branche 45, emploi total; serie corrigee au 4Q91“
file = “h :\daten\besta\x12aB\B45Bnouv.prn“
decimals = 5
name = “B45Bnouv“
start = 1991.3
period=4
}
#
arima{
model=(0 1 0)(0 1 0)4
}
#
x11{
mode=mult
}
#
check{
print=(all)
}
#
history{
estimates=(FCST seasonal)
save=(FCSThistory SFrevisions)
savelog=(asa ach atr atc asf asp)
start=1997.1}
```

TAB. 2.4: Fichier de commandes de X12-Arima

Modèle Parameter	Corrigé Value (fixed)	Non corrigé Value (fixed)
Variance	$(5096)^2$	$(7931)^2$
Likelihood Statistics		
Effective number of observations (nefobs)	32	32
Number of parameters estimated (np)	1	1
Log likelihood (L)	-318.5675	-332.7193
AIC	639.1351	667.4386
AICC (F-corrected-AIC)	639.2684	667.5719
Hannan Quinn	639.6209	667.9245
BIC	640.6008	668.9043

TAB. 2.5: Comparaison des modèles avec et sans la correction au 4Q91

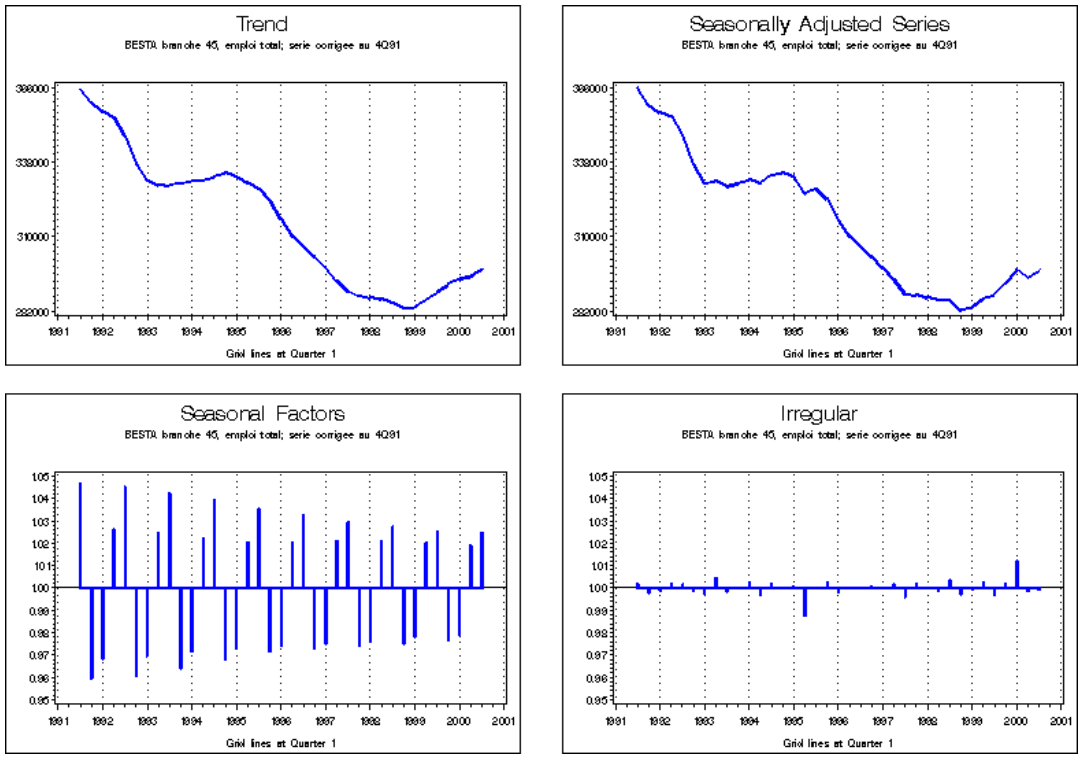


FIG. 2.3: Composantes de la série corrigée

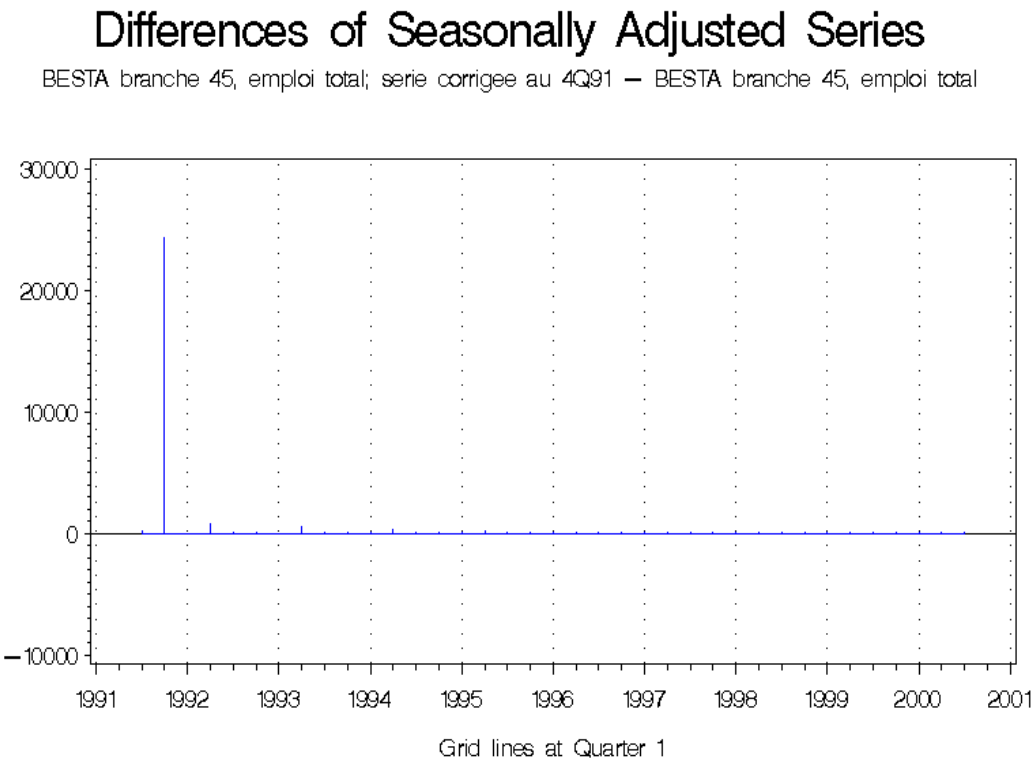


FIG. 2.4: Effet de la correction du 4e trimestre 1991 sur la série désaisonnalisée

Interprétation des composantes La composante irrégulière représente donc le facteur de correction à apporter à la tendance pour obtenir la série désaisonnalisée. Elle a des fluctuations d'amplitude très faibles, inférieures à 1%, ce qui indique une grande stabilité à court terme dans la branche 45. Deux trimestres ressortent : I_{2Q95} et I_{1Q00}. Sans eux, l'amplitude de $\{I_t\}$ ne dépasserait pas 0.5%.

Les facteurs saisonniers représentent le facteur de passage de la série désaisonnalisée à la série originale. L'amplitude de ces facteurs est bien supérieure à celle de $\{I_t\}$, ce qui est évidemment une caractéristique d'un phénomène à la saisonnalité marquée, comme l'emploi dans la construction. On est frappé par la régularité du phénomène. Il est intéressant de constater que l'amplitude de ces facteurs décroît avec le temps.

Quelle est l'influence de la correction du 4Q91 sur la série désaisonnalisée ? La fig. 2.4, montrant la différence entre les séries désaisonnalisées avec et sans la correction du 4Q91, permet de constater que cette correction n'a qu'une influence négligeable sur la désaisonnalisation des autres trimestres.

2.2.3 Contrôle de la qualité de l'ajustement saisonnier

Pour le contrôle de la qualité, nous nous sommes inspirés de la publication de la Banque centrale européenne [1, A2. Example of Monthly Quality Report].
Le tableau 2.6 montre que tout se passe bien :

- Il y a une saisonnalité identifiable
- Il n'y a pas de saisonnalité résiduelle dans la composante irrégulière.
- La contribution relative à la variance de la série originale stationnarisée est pour la composante irrégulière de 0.2%, contre 44% pour la tendance et 48% pour la composante saisonnière.
- Le rapport I/C est le rapport de la moyenne des valeurs absolues du taux d'accroissement relatif $|I_t - I_{t-1}| / I_t$ de la composante irrégulière, sur la moyenne des taux correpondants $|T_t - T_{t-1}| / T_t$ pour la tendance. I/C = 0.29, ce qui prouve qu'en valeur relatives, la variabilité moyenne à court terme due à la tendance est plus de 3 fois celle de la composante irrégulière.
- Le rapport I/S, aussi noté MSR (moving seasonnality ratio) détermine la longueur du filtre saisonnier. Puisqu'il est plus petit que 2.5, on utilisera un filtre court 3x3.
- Toutes les statistiques M1 à M11 sont inférieures à 1.
- Les statistiques globales Q et Q2 sont largement inférieures à 1.

La fig. 2.5 présente l'évolution au cours des années du facteur saisonnier. Elle donne les mêmes informations que la fig. 2.3 (en bas à g.), mais groupées par trimestres.
La fig. 2.6 montre la construction des facteurs saisonniers. A partir de la série brute débarassée de sa tendance ($*$), le programme calcule la série corrigée des valeurs extrêmes (o), qui est enfin lissée par un filtre 3x3 (ligne continue). Le degré de correction des valeurs extrêmes est ajusté par le paramètre SIGMALIM. La longueur du filtre dépend de la régularité de l'estimation provisoire. On reconnaît les courbes présentées à la fig. 2.5.
Enfin, à la fig. 2.7, on voit superposées la tendance, la série désaisonnalisée et la série brute.

X-11 Seasonal Adjustment

Seasonal filter length :	M.S.R.
Seasonal filter length selected by MSR :	3x3
Trend filter length :	default
X-11 Extreme adjustment :	Standard Error
Type of Adjustment :	multiplicative seasonal adjustment

Diagnostics for direct seasonally adjusted series

F-test for Stable Seasonality (D8) :	163.915
F-test for Moving Seasonality (D8) :	6.448

Identifiable Seasonality present
No evidence of residual seasonality in D11.

Relative Contribution to the Variance (F2.F :)

I	C	S	P	TD	Total
0.20	43.49	47.67	0.00	0.00	91.36

I/C RATIO	I/S RATIO
0.29	1.84

M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11
0.064	0.020	0.000	0.259	0.200	0.863	0.283	0.610	0.584	0.554	0.520

Q	Q (without M2)
0.25	0.29

Visually distinct seasonal spectral peaks were not found.
Visually distinct trading day spectral peaks were not found.

TAB. 2.6: Résumé de la désaisonnalisation

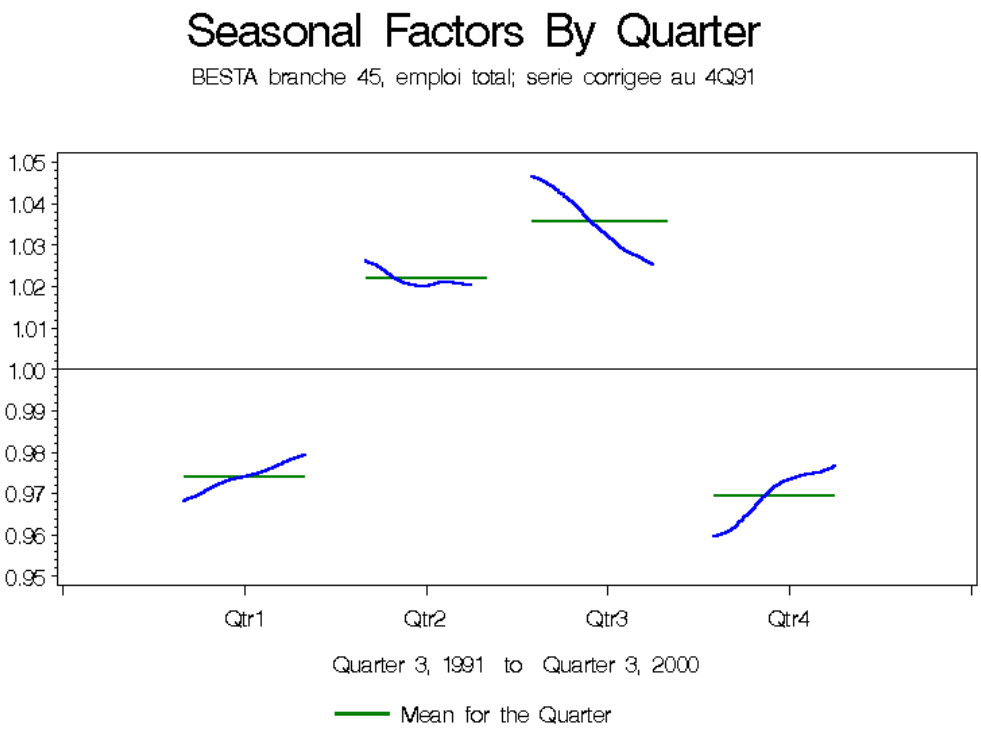


FIG. 2.5: Evolution par trimestre de la composante saisonnière

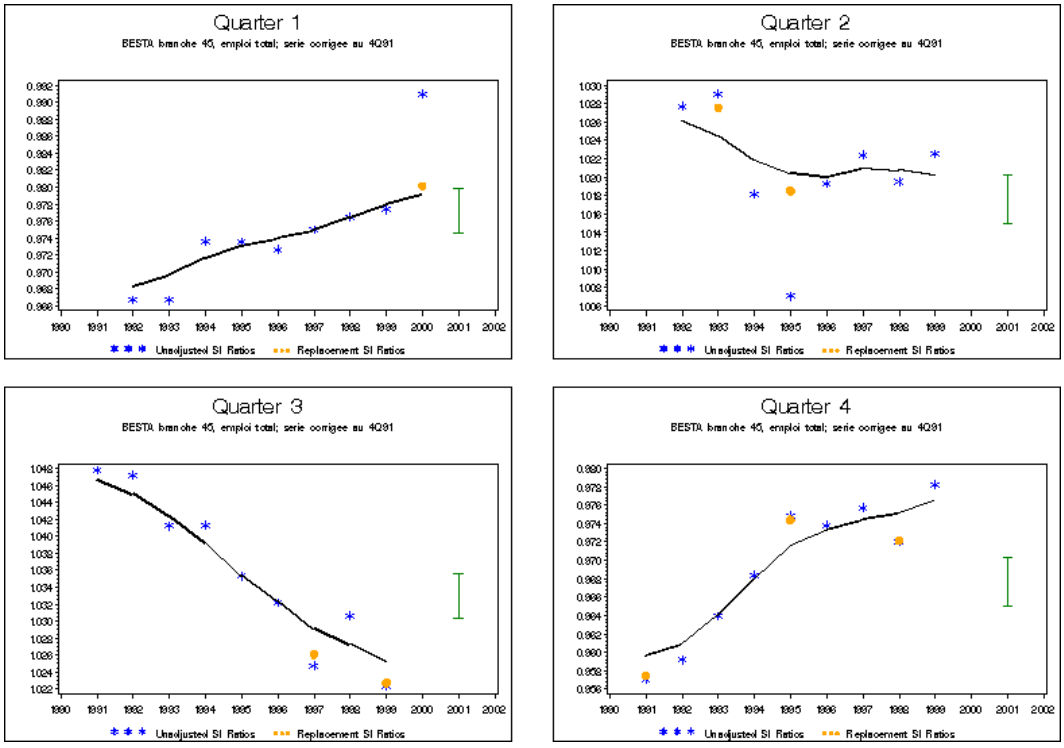


FIG. 2.6: Composante saisonnière+irrégulière (SI) : $*$; SI corrigée des valeurs extrêmes : o ; comp. saisonnière (S) : -.

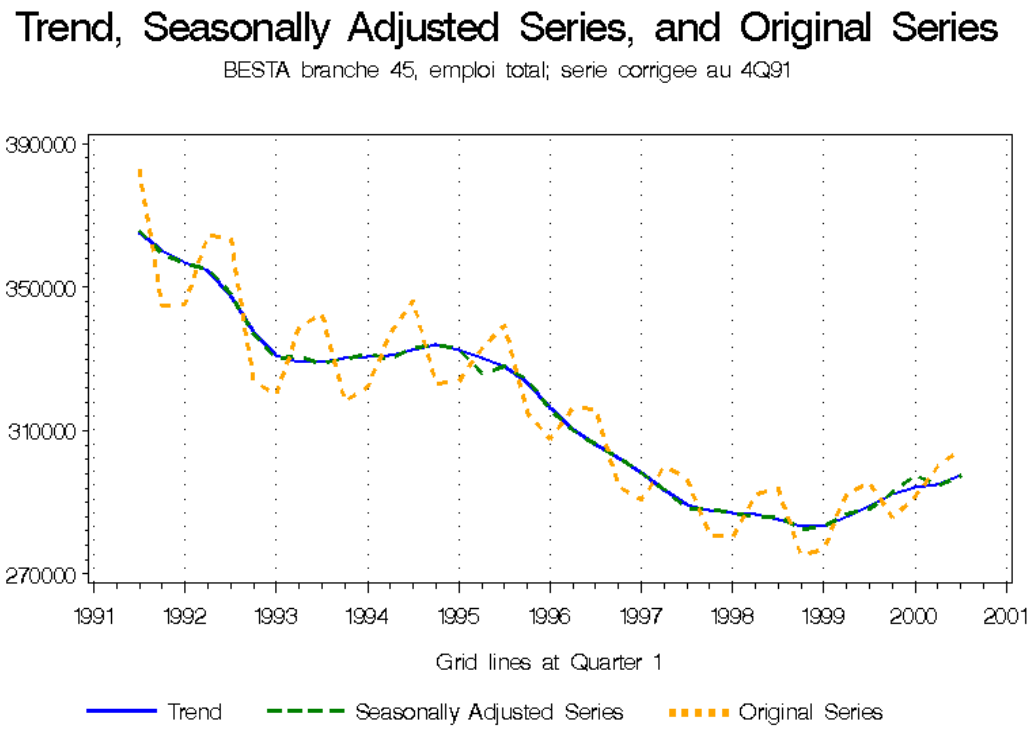


FIG. 2.7: Tendance, série désaisonnalisée et série brute.

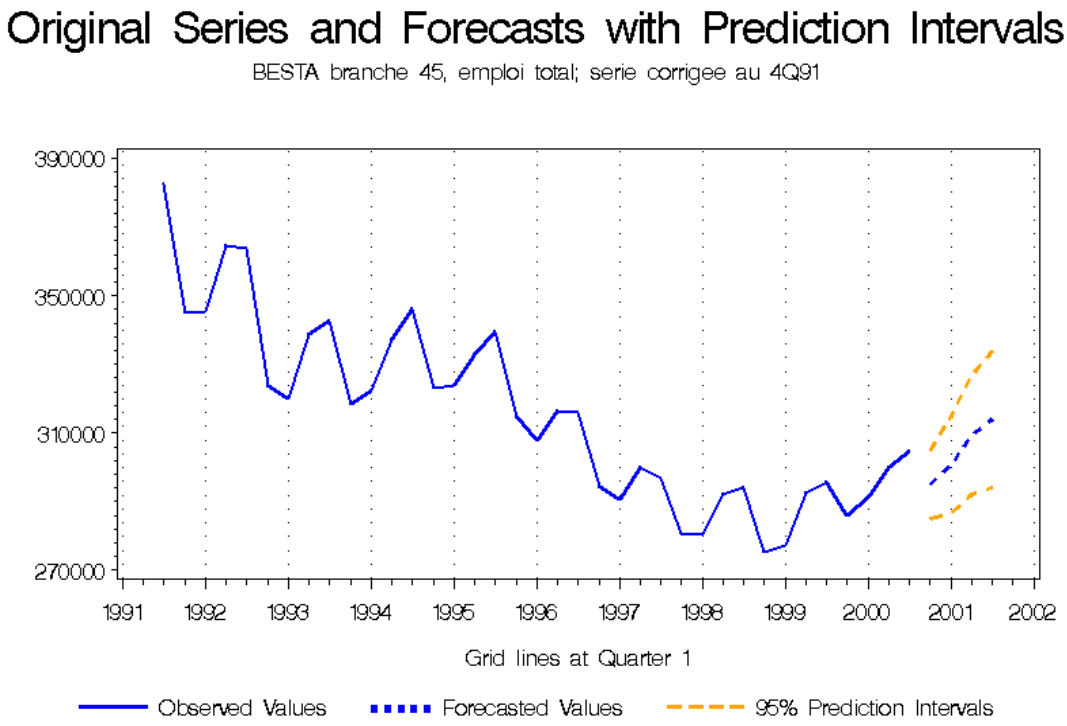


FIG. 2.8: Prévisions sur un an données par le modèle ARIMA

2.2.4 Evaluation de l'impact des révisions

La fig. 2.8 montre les valeurs anticipées (après le 3Q00, dernière valeur entrée dans l'analyse) de la série brute, basées sur le modèle ARIMA(0,1,0)$(0,1,0)_4$, avec l'intervalle de confiance point par point des prévisions. Pour juger de la qualité de ces anticipations, X12-ARIMA propose une analyse historique des prévisions.

Supposons que nous n'utilisions que partiellement notre connaissance de la série, et que nous voulions anticiper (a) au trimestre suivant et (b) au même trimestre de l'année suivante. Faisons l'exercice sur 15 trimestres, du 1Q97 jusqu'au 3Q00.

La table 2.7 montre les prévisions de la série brute faites au trimestre précédent et au même trimestre de l'année précédente, ainsi que l'erreur (valeur finale - prévision). On constate que l'erreur relative est faible, mais que le modèle a tendance à sous-estimer la valeur finale. Cela provient du fait - visible sur la fig. 2.8 - que la forte décroissance amorcée en 1995 s'atténue régulièrement, puis s'inverse sur la période des prévisions. L'ordre de grandeur des erreurs est en accord avec l'écart-type de 5200 mentionné plus haut. Dans le cas de notre modèle très simple, la prévision à 1 trimestre est facile à calculer. Notons la valeur anticipée à 1 trimestre de $\widehat{BZ}_t$ par $\widehat{BZ}_{t.1}$, et par

t	Prévision 1 trim. avant $\widehat{BZ}_{t.1}$	Erreur de la prévision $e_{t.1}$	Prévision 4 trim. avant $\widehat{BZ}_{t.4}$	Erreur de la prévision $e_{t.4}$
1997.2	299340	520	-	-
1997.3	299550	-2930	-	-
1997.4	274950	5830	-	-
1998.1	277140	3150	273720	6570
1998.2	289440	2600	283390	8650
1998.3	288800	5090	277230	16670
1998.4	278050	-2760	267220	8080
1999.1	274790	2280	269860	7220
1999.2	288830	3730	284220	8340
1999.3	294410	1110	291160	4360
1999.4	276920	8830	269810	15950
2000.1	287530	3880	273870	17550
2000.2	306900	-6810	293080	7010
2000.3	303050	1700	297140	7610

TAB. 2.7: Prévisions de la série brute (arrondies à la dizaine)

$e_{t.1}$ l'erreur de prévision :

$$\begin{aligned} \widehat{BZ}_t &= \widehat{BZ}_{t-1} + \left(\widehat{BZ}_{t-4} - \widehat{BZ}_{t-5}\right) + \epsilon_t \qquad \text{(modèle ARIMA}(0,1,0)(0,1,0)_4\text{)} \\ \widehat{BZ}_{t.1} &= \widehat{BZ}_{t-1} + \left(\widehat{BZ}_{t-4} - \widehat{BZ}_{t-5}\right) \\ e_{t.1} &= \widehat{BZ}_t - \widehat{BZ}_{t.1} = \epsilon_t \end{aligned}$$

On voit que dans ce cas, la résiduelle du modèle ARIMA(0,1,0)(0,1,0)$_4$ n'est autre que l'erreur de prévision.

Exemple

$\widehat{BZ}_{2Q97} = 299\,340 + 520 = 299\,860$
$\widehat{BZ}_{3Q97} = 299\,550 - 2930 = 296\ 620$
$\widehat{BZ}_{2Q98} = 289\,440 + 2600 = 292\ 040$
$\widehat{BZ}_{3Q98.1} = 292\,040 + (296\ 620 - 299\ 860) = 288\,800$

On trouve de même la valeur anticipée à 4 trimestres, $\widehat{BZ}_{t.4}$, en remplaçant $\widehat{BZ}_{t-1}$ (et ensuite $\widehat{BZ}_{t-2}$ et $\widehat{BZ}_{t-3}$) dans la formule ci-dessus par leur valeur anticipée. L'erreur de prévision à 4 trimestres est donnée par $\epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} + \epsilon_{t-3}$. Sa variance est donc 4 fois celle du bruit blanc $\{\epsilon_t\}$.

	Prévision au trimestre suivant		Prévision à 4 trimestres	
Date de la prévision	SC de l'erreur	EQM de l'erreur	SC de l'erreur	EQM de l'erreur
1997.2	269662.1	269662.1	-	-
1997.3	8848000.2	4424000.1	-	-
1997.4	42836550.4	14278850.1	-	-
1998.1	52733300.1	13183325.0	43116164.4	43116164.4
1998.2	59494548.1	11898909.6	117890924.0	58945462.0
1998.3	85396031.6	14232671.9	395628814.3	131876271.4
1998.4	92993937.9	13284848.3	460900186.4	115225046.6
1999.1	98204520.2	12275565.0	512968389.0	102593677.8
1999.2	112081337.3	12453481.9	582536499.5	97089416.6
1999.3	113302892.7	11330289.3	601516811.6	85930973.1
1999.4	191309589.7	17391780.9	855766533.6	106970816.7
2000.1	206392560.0	17199380.0	1163636369.8	129292930.0
2000.2	252745917.4	19441993.6	1212814751.9	121281475.2
2000.3	255633368.0	18259526.3	1270677245.3	115516113.2

TAB. 2.8: (R8) Evolution sur 14 trimestres.
SC = somme des carrés cumulée ; EQM = erreur quadratique moyenne=SC/nb. trimestres

Au tableau 2.8 (R8 dans la numérotation de X12-ARIMA) on trouve l'évolution de la somme des carrés des erreurs et celle de l'erreur quadratique moyenne des prévisions du 2Q97 au 3Q00 des données brutes ajustées pour les valeurs aberrantes (il n'y en a pas sur cette période).

Effet sur la composante saisonnière Le tableau 2.9 (R6 dans la numérotation de X12-ARIMA) compare le taux de révision (en %) à appliquer à la composante saisonnière pour chaque trimestre lors de l'anticipation au trimestre suivant (anglais : concurrent, 2e col.), ou au même trimestre de l'année suivante (anglais : projected, 3e col.). Les paramètres du modèle regARIMA sont réestimés chaque fois. Dans ce cas, le seul paramètre est la variance. On constate que ces révisions sont minimes : elles ne dépassent en valeur absolue 0.5% qu'au 3Q99.
Le tableau 2.10 (R6.S dans la numérotation de X12-ARIMA) donne les moyennes par période des valeurs absolues des révisions de la composante saisonnière. On voit (3e partie du tableau) que si les taux de révision extrêmes sont du même ordre à 1 trimestre ou à 1 an, les valeurs centrales sont plus élevées pour les prévisions à 1 an.

Remarque Le modèle permet également une estimation optimale de $\widehat{BZ}_t$ au sens de la variance de prévision, qui est différente de la valeur anticipée, car elle utilise également les valeurs postérieures à t. Pour notre modèle, cet estimateur optimal (c'est-à-dire minimisant de la variance de

Date	Conc - Final	Proj - Final
1997		
1st	0.13	0.42
2nd	0.22	0.34
3rd	-0.26	-0.57
4th	-0.22	-0.28
1998		
1st	-0.01	0.26
2nd	-0.08	-0.01
3rd	-0.19	-0.07
4th	0.26	-0.25
1999		
1st	0.12	0.26
2nd	-0.34	-0.03
3rd	-0.71	-0.62
4th	-0.05	0.45
2000		
1st	0.16	0.41
2nd	0.01	-0.38

TAB. 2.9: (R6) Révisions (en %) à 1 trimestre (col. 2) et à 1 an (col. 3) de la composante saisonnière

l'erreur parmi tous les estimateurs linéaires) est donné par :

$$\begin{aligned} \widehat{BZ}_{t.opt} &= \frac{1}{2}\left(\widehat{BZ}_{t-1} + \widehat{BZ}_{t+1} + \left(\widehat{BZ}_{t-4} - \widehat{BZ}_{t-5}\right) + \left(\widehat{BZ}_{t+4} - \widehat{BZ}_{t+5}\right)\right) \quad (2.18) \\ e_{t.opt} &= \widehat{BZ}_t - \widehat{BZ}_{t.opt} = \frac{1}{2}\left(\epsilon_t + \epsilon_{t+5}\right) \end{aligned}$$

La variance de l'erreur d'estimation est alors, selon le modèle, la moitié de la variance de l'erreur de prévision, car $\{\epsilon_t\}$ forme un bruit blanc.

2.3 Discussion

Les révisions dont nous avons parlé n'ont rien à voir avec l'adaptation de la BESTA aux BZ. La série brute dont nous partons est la série BESTA $\left\{\widehat{BZ}_t\right\}$ telle qu'elle est connue au 3Q00, c'est à dire :

- avec adaptations aux 3 recensements des entreprises jusqu'au 3Q98, date du dernier BZ,
- avec adaptation au nouvel échantillon du 2Q00 et rétropolation au BZ98 sur la période 3Q98 à 2Q00.

Date	Conc - Final	Proj - Final
Quarters :		
1st	0.11	0.34
2nd	0.16	0.19
3rd	0.39	0.42
4th	0.18	0.33
Years :		
1997	0.21	0.40
1998	0.14	0.15
1999	0.31	0.34
2000	0.09	0.39
Total :	0.20	0.31
Hinge	Values :	
Min	0.01	0.01
25%	0.08	0.25
Med	0.18	0.31
75%	0.26	0.42
Max	0.71	0.62

TAB. 2.10: (R6.S) Moyennes par période des taux absolus de révision (en %) à 1 trimestre (col 2) et à 1 an (col 3)

Cette série subira encore, lors du prochain recensement 2001 des entreprises, une adaptation rétrospective jusqu'au 3Q98.

Impact d'une adaptation rétrospective Supposons que les données du recensement soient disponibles au 3Q01, et posons

$$z_{01} = \frac{BZ_{01}}{B_{3Q01}}$$

où BZ_{01} est la valeur de l'emploi total dans la branche 45, telle qu'elle sera donnée par le BZ de 2001, et B_{3Q01} est la valeur de l'emploi estimée au 3Q01 selon le nouvel échantillon de 2000.
La transformation doit tenir compte de l'adaptation (2.11), qui introduit déjà une correction liée au BZ98 :
Posons

$$\widehat{B}_t^{(n)} = \left(\sqrt[7]{b_{00}}\right)^{t-3Q98} B_t^{(a)} \quad 3Q98 \leq t < 2Q00$$

$\widehat{B}_t^{(n)}$ modélise le passage progressif de l'ancien échantillon $B_t^{(a)}$ au nouvel échantillon $B_t^{(n)}$.

Cependant, la BESTA actuelle n'est pas $\widehat{B}_t^{(n)}$, mais $\widehat{BZ}_t$, statistique dépendant de z_{98} par (2.11), et nous pouvons écrire

$$\widehat{BZ}_t = z_{98}\left(\sqrt[7]{1/z_{98}}\right)^{t-3Q98}\widehat{B}_t^{(n)} \quad 3Q98 \leq t < 2Q00 \tag{2.19}$$

Notons par un exposant $^{(r)}$ les série adaptées rétrospectivement. L'adaptation rétrospective aux recensements de 1998 et 2001 peut s'écrire à l'aide de $\widehat{B}_t^{(n)}$ de manière analogue à (2.8) :

$$\widehat{BZ}_t^{(r)} = z_{98}\left(\sqrt[12]{z_{01}/z_{98}}\right)^{t-3Q98}\widehat{B}_t^{(n)} \quad 3Q98 \leq t < 2Q00 \tag{2.20}$$

donc en exprimant $\widehat{B}_t^{(n)}$ à l'aide de (2.19) et en substituant cette expression dans (2.20), on obtient $\widehat{BZ}_t^{(r)} = \left(\sqrt[12]{z_{01}/z_{98}}\right)^{t-3Q98}\left(\sqrt[7]{z_{98}}\right)^{t-3Q98}\widehat{BZ}_t$. Après simplification, on a enfin :

$$\widehat{BZ}_t^{(r)} = \begin{cases} \left(\sqrt[12]{z_{01}\left(z_{98}\right)^{5/7}}\right)^{t-3Q98}\widehat{BZ}_t & 3Q98 \leq t < 2Q00 \\ z_{98}\left(\sqrt[12]{z_{01}/z_{98}}\right)^{t-3Q98}B_t^{(n)} & 2Q00 \leq t < 3Q01 \end{cases} \tag{2.21}$$

Puisque le modèle de désaisonnalisation est multiplicatif, on peut penser que cette adaptation a posteriori ne modifiera ni les facteurs saisonniers, ni la composante irrégulière, mais se répercutera sur la tendance et la série désaisonnalisée, voir les formules (2.16) et (2.17) :

$$T_t^{(r)} = \begin{cases} \left(\sqrt[12]{z_{01}\left(z_{98}\right)^{5/7}}\right)^{t-3Q98}T_t & 3Q98 \leq t < 2Q00 \\ z_{98}\left(\sqrt[12]{z_{01}/z_{98}}\right)^{t-3Q98}T_t & 2Q00 \leq t < 3Q01 \end{cases}$$

$$SA_t^{(r)} = \begin{cases} \left(\sqrt[12]{z_{01}\left(z_{98}\right)^{5/7}}\right)^{t-3Q98}SA_t & 3Q98 \leq t < 2Q00 \\ z_{98}\left(\sqrt[12]{z_{01}/z_{98}}\right)^{t-3Q98}SA_t & 2Q00 \leq t < 3Q01 \end{cases} \tag{2.22}$$

Analyse de la correction du saut au 1Q97 Notre point de départ pour la désaisonnalisation était la série $\{x_t\}$, où $x_t = \ln\left(\widehat{BZ}_t\right)$. Un modèle ARIMA(0,1,0)(0,1,0)$_4$ a été ajusté à $\{x_t\}$. Ce modèle s'écrit donc :

$$\begin{aligned}
\nabla\nabla_4 x_t &= x_t - x_{t-1} - (x_{t-4} - x_{t-5}) \\
&= \ln\left(\widehat{BZ}_t\right) - \ln\left(\widehat{BZ}_{t-1}\right) - \left(\ln\left(\widehat{BZ}_{t-4}\right) - \ln\left(\widehat{BZ}_{t-5}\right)\right) \\
&= \ln\left(\widehat{BZ}_t/\widehat{BZ}_{t-1}\right) - \ln\left(\widehat{BZ}_{t-4}/\widehat{BZ}_{t-5}\right) \\
&= \ln\left(1+\widehat{s}_t\right) - \ln\left(1+\widehat{s}_{t-4}\right) \\
&= \eta_t
\end{aligned}$$

où $\widehat{s}_t$ est le taux d'accroissement relatif de $\widehat{BZ}_t$, voir (2.13), et $\{\eta_t\}$ est un bruit blanc de variance estimée à $0.26{\cdot}10^{-3}$.
En utilisant l'approximation

$$\ln\left(1+\widehat{s}_t\right) \simeq \widehat{s}_t$$

on obtient un modèle pour la série des taux d'accroissement relatifs :

$$\widehat{s}_t = \widehat{s}_{t-4} + \eta_t$$

Ce modèle décrit simplement le taux d'accroissement relatif en t comme étant égal au taux correspondant 4 trimestres avant, plus une erreur aléatoire imprévisible de variance $0.26{\cdot}10^{-3}$. Selon ce modèle, l'estimation optimale de $\widehat{s}_t$, lorsque la série est connue au moins jusqu'en $t+4$, est donnée par

$$\begin{aligned}
\widehat{s}_{t.opt} &= \frac{1}{2}\left(\widehat{s}_{t-4} + \widehat{s}_{t+4}\right) \\
\widehat{\eta}_{t.opt} &= \widehat{s}_t - \widehat{s}_{t.opt} = \frac{1}{2}\left(\eta_t + \eta_{t+4}\right)
\end{aligned}$$

et la variance de cette prévision est $0.13{\cdot}10^{-3}$. On reconnaît la formule de la correction du saut au 1Q97 utilisée, voir (2.3). La nouveauté est que le modèle ARIMA fournit l'estimation de la variance de l'erreur associée à une correction de ce type.
En utilisant (2.15) avec l'approximation : $\ln\left(1+\widehat{s}_t\right) \simeq \widehat{s}_t$, on voit que, pour $t =$1Q97 :

$$\begin{aligned}
\widehat{\eta}_{t.opt} &= \widehat{s}_t - \frac{1}{2}\left(\widehat{s}_{t-4} + \widehat{s}_{t+4}\right) \\
&= r_t - \frac{1}{2}\left(r_{t-4} + r_{t+4}\right) - 0.0302 + 0.0084
\end{aligned}$$

On a donc

$$\Delta_t \doteqdot r_t - \frac{1}{2}\left(r_{t-4} + r_{t+4}\right) = \widehat{\eta}_{t.opt} + 0.0218$$

où le membre de gauche représente la correction effectuée initialement sur la série non adaptée. Cette correction change-t-elle significativement le taux d'accroissement relatif au 1Q97 ?
Sous l'hypothèse que $\widehat{\eta}_{t.opt}$ suit une loi normale $\mathbf{N}\left(0, 0.13 \cdot 10^{-3}\right)$, le test de $\mathbf{H}_0 : \Delta_t = 0$, contre $\mathbf{H}_1 : \Delta_t \neq 0$ au risque $\alpha = 5\%$, donne une valeur critique de 0.0223. Cette valeur est légèrement supérieure à 0.0218, nous pouvons donc admettre que la différence n'est pas significative.

Annexes

ANNEXE 1 : Données pour la branche 45

Ces données sont extraites du fichier de la section EWL "Besta Emploi total.xls".

B45xt.txt	Contenu de B45diff.txt	B45Bnouv.prn
12.855	-0.013	382732.00
12.681	-0.010	344832.81
12.752	-0.008	345229.84
12.806	-0.003	364499.63
12.805	0.000	363863.89
12.688	0.000	323820.96
12.676	0.003	320042.92
12.733	0.007	338806.48
12.744	0.008	342604.51
12.670	0.011	318198.00
12.683	0.014	322156.24
12.729	0.017	337245.37
12.755	0.020	346213.94
12.686	0.023	323167.84
12.688	0.026	323791.54
12.715	0.028	332835.15
12.735	0.031	339454.00
12.660	0.023	314986.04
12.637	0.015	307691.45
12.665	0.006	316328.65
12.664	-0.002	316022.59
12.593	-0.010	294352.36
12.580	-0.041	290703.51
12.611	-0.049	299860.00
12.600	-0.056	296625.06
12.545	-0.065	280784.80
12.544	-0.073	280281.86
12.585	-0.082	292038.59
12.591	-0.092	293893.00
12.526	-0.088	275296.31
12.532	-0.086	277076.04
12.586	-0.083	292557.93
12.596	-0.080	295517.58
12.563	-0.077	285753.03
12.583	-0.073	291416.44
12.612	-0.073	300090.00
12.627	0.000	304748.90

- B45Bnouv.prn contient la série adaptée ;
- B45xt.txt contient le logarithme naturel de la série adaptée ;
- B45diff.txt contient la différence entre le logarithme naturel de la série adaptée et de la série brute

ANNEXE 2 : Etape d'identification du modèle

Série analysée: x_t

Contents of spc file h:\daten\besta\x12a\B45xt0.spc

```
series{     title = "BESTA branche 45, emploi total"
            file = "h:\daten\besta\x12a\B45xt.txt"
            decimals = 5
            name = "B45xt"
            start = 1991.3
            period=4}
identify{   diff=(0,1)
            sdiff=(0,1)
            print=all}
```

MODEL IDENTIFICATION

Autocorrelations of the Residuals					Partial Autocorrelations of the Residuals				
Differencing: none									
Lag	1	2	3	4	Lag	1	2	3	4
ACF	0.72	0.61	0.66	0.71	PACF	0.72	0.19	0.37	0.29
SE	0.16	0.23	0.27	0.31	SE	0.16	0.16	0.16	0.16
Q	20.65	35.81	54.37	76.20					
DF	1	2	3	4					
P	0.000	0.000	0.000	0.000					
Differencing: Seasonal Order=1									
Lag	1	2	3	4	Lag	1	2	3	4
ACF	0.73	0.58	0.35	0.16	PACF	0.73	0.11	-0.23	-0.13
SE	0.17	0.25	0.29	0.30	SE	0.17	0.17	0.17	0.17
Q	19.13	31.66	36.29	37.29					
DF	1	2	3	4					
P	0.000	0.000	0.000	0.000					
Differencing: Nonseasonal Order=1									
Lag	1	2	3	4	Lag	1	2	3	4
ACF	-0.08	-0.59	-0.11	0.69	PACF	-0.08	-0.60	-0.38	0.44
SE	0.17	0.17	0.22	0.22	SE	0.17	0.17	0.17	0.17
Q	0.28	14.27	14.77	35.36					
DF	1	2	3	4					
P	0.597	0.001	0.002	0.000					
Differencing: Nonseasonal Order=1, Seasonal Order=1									
Lag	1	2	3	4	Lag	1	2	3	4
ACF	-0.17	-0.04	-0.07	0.02	PACF	-0.17	-0.07	-0.09	-0.01
SE	0.18	0.18	0.18	0.18	SE	0.18	0.18	0.18	0.18
Q	1.06	1.10	1.28	1.30					
DF	1	2	3	4					
P	0.304	0.576	0.733	0.862					

A titre d'illustration, nous donnons les autocorrélations et les autocorrélations partielles des modèles jusqu'à l'ordre 4. On voit qu'elles ne sont plus significatives, lorsqu'on prend les différences d'ordre 1, non saisonnière et saisonnière.

ANNEXE 3 : Etape de pré-ajustement

Série analysée: x_t

Contents of spc file h:\daten\besta\x12a\B45xt6.spc

series{	title = "BESTA branche 45, emploi total"
	file = "h:\daten\besta\x12a\B45xt.txt"
	decimals = 5
	name = "B45xt"
	start = 1991.3
	period=4}
outlier{	}
regression{	user=(difference)
	start=1991.3
	file = "h:\daten\besta\x12a\B45diff.txt"
	aictest=(user)}
arima{	model=(0 1 0)(0 1 0)4}
x11{	mode=add }
check{	print=(all)}

Likelihood Statistics for model	with user-defined regressors	without user-defined regressors	with outlier detection AO1991.4
Effective number of observations (nefobs)	32	32	32
Number of parameters estimated (np)	2	1	2
Loglikelihood (L)	74.4723	74.3461	86.9525
AIC	-144.9446	-146.6922	-169.9051
AICC (F-corrected-AIC)	-144.5308	-146.5589	-169.4913
Hannan Quinn	-143.9729	-146.2064	-168.9334
BIC	-142.0132	-145.2265	-166.9736

On constate que
- l'ajustement au modèle n'est pas moins bon pour la série adaptée que pour la série brute.
- l'introduction d'une correction au 4Q91 améliore nettement les statistiques.

ANNEXE 4 : Corrections des branches saisonnières

BESTA, emploi total : principales branches saisonnières

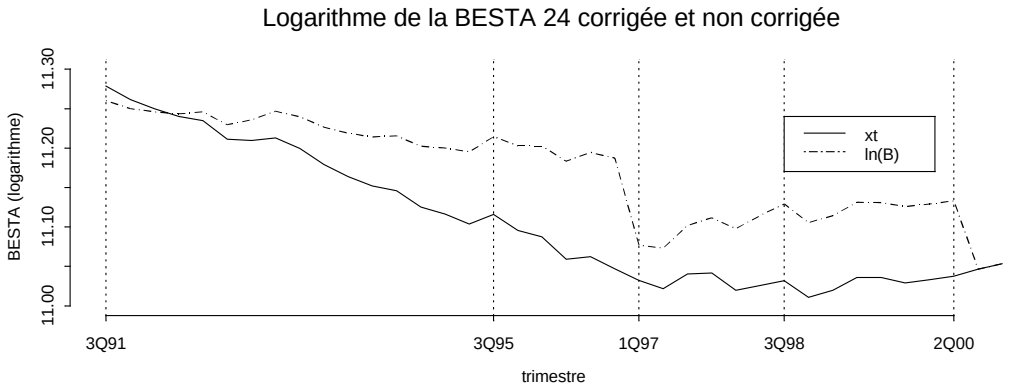
branche	p1	h1	p2	h2	p3	h3	p4	h4
24	-0.0073	0.0186	-0.0084	-0.0986	-0.0084	0.0045	0.0002	-0.0969
28	0.0093	0.0077	-0.0045	0.1561	-0.0045	-0.0114	0.0018	-0.0656
29	-0.0034	-0.0094	-0.0028	-0.0635	-0.0028	0.0349	-0.0045	0.0015
45	0.0027	-0.0131	-0.0084	0.0306	-0.0084	0.0084	0.0027	-0.0922
55	0.0006	0.0002	0.0007	0.0091	0.0007	0.0071	-0.0023	0.0153
72	0.0117	-0.1792	0.0098	0.0084	0.0098	-0.0546	0.0211	0.0634
74	0.0037	0.0722	-0.0036	0.1307	-0.0036	0.0374	0.0042	-0.0063
85	0.0039	-0.0009	0.0010	0.0608	0.0010	0.0324	0.0016	0.0443

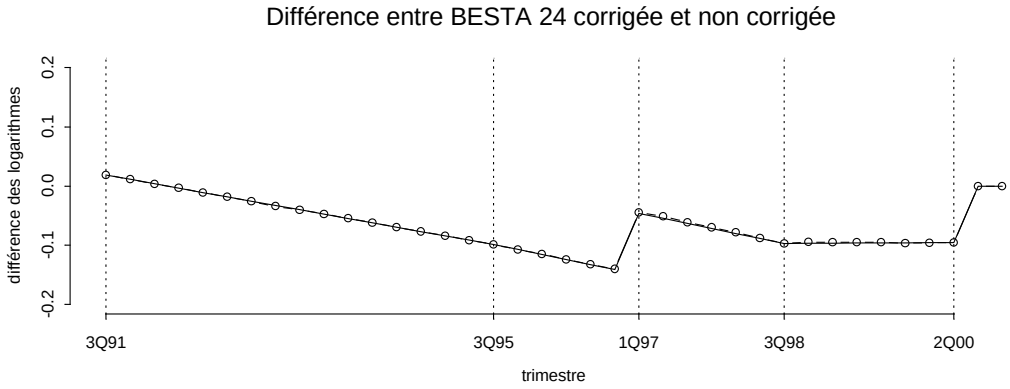
Pente et ordonnée à l'origine des corrections (échelle logarithmique) pour chaque période
(1 = 3Q91-3Q95 ; 2 = 3Q95-1Q97 ; 3 = 1Q97-3Q98 ; 4 = 3Q98-2Q00)

Les corrections appliquées à la BESTA pour les branches saisonnières suivent le schéma décrit au §2.1.4. Au tableau ci-dessus, on trouve les valeurs permettant le passage du logarithme de la BESTA non corrigée ($\ln(B_t)$) à la BESTA adaptée aux recensements:

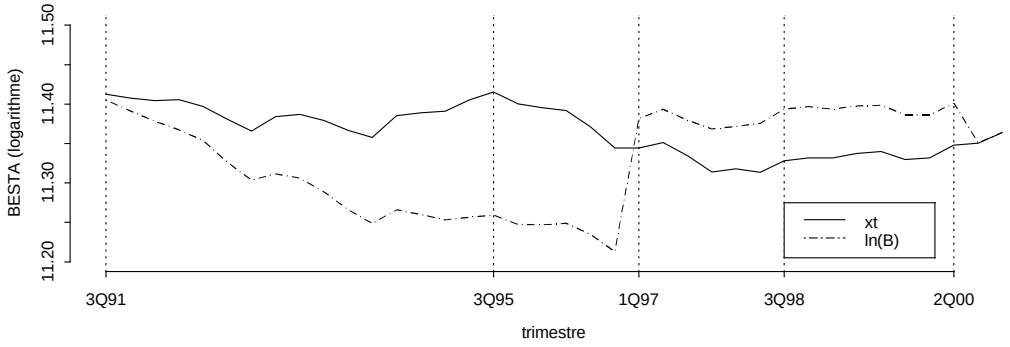
$$x_t = \ln(B_t) + h_i + p_i(t - d_i), i = 1,2,3,4 \text{ avec } d_1 = 3Q91, d_2 = d_3 = 3Q95, d_4 = 3Q98.$$

La ligne de la branche 45 réplique les valeurs introduites dans (2.14).

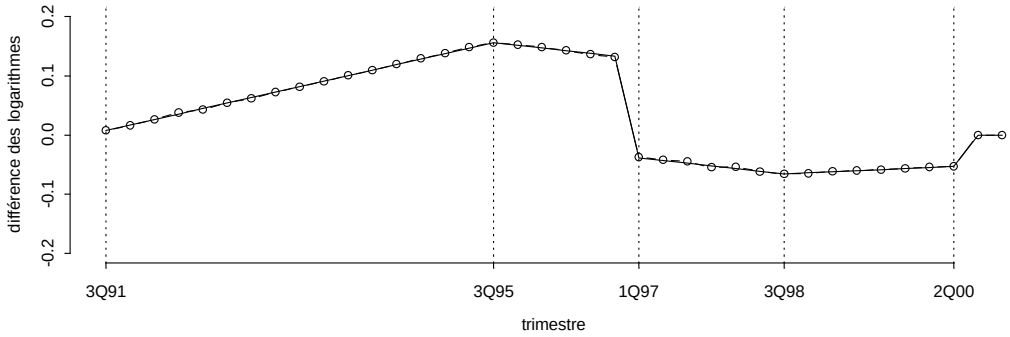




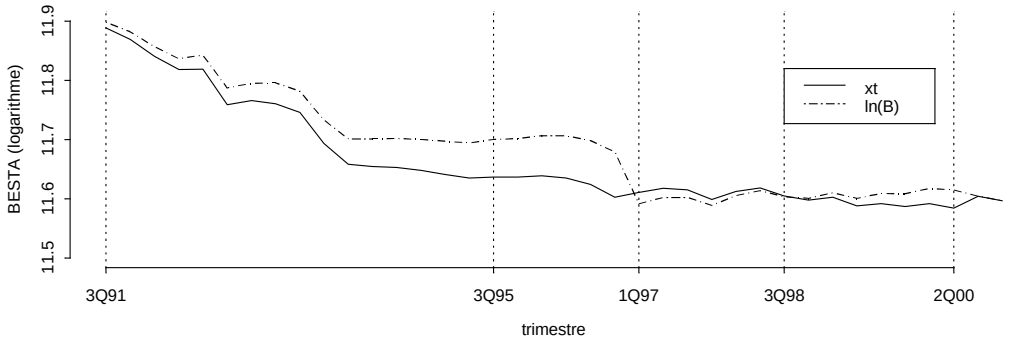
Logarithme de la BESTA 28 corrigée et non corrigée



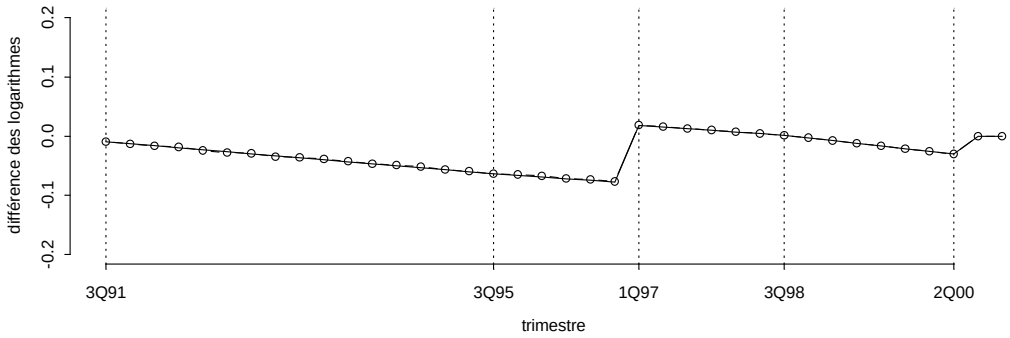
Différence entre BESTA 28 corrigée et non corrigée



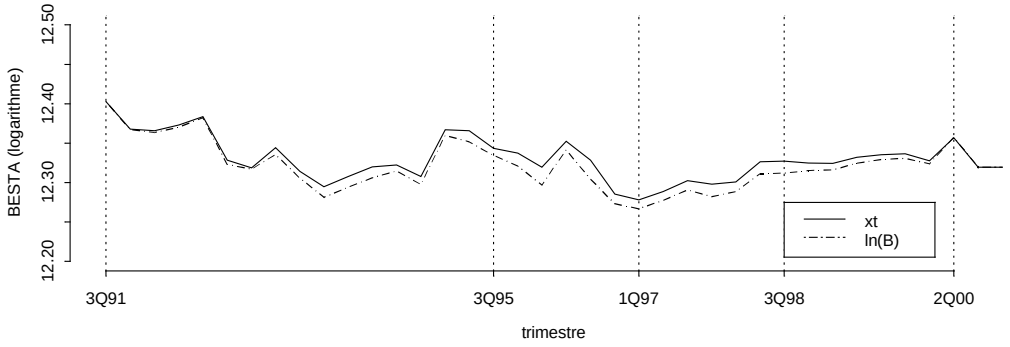
Logarithme de la BESTA 29 corrigée et non corrigée



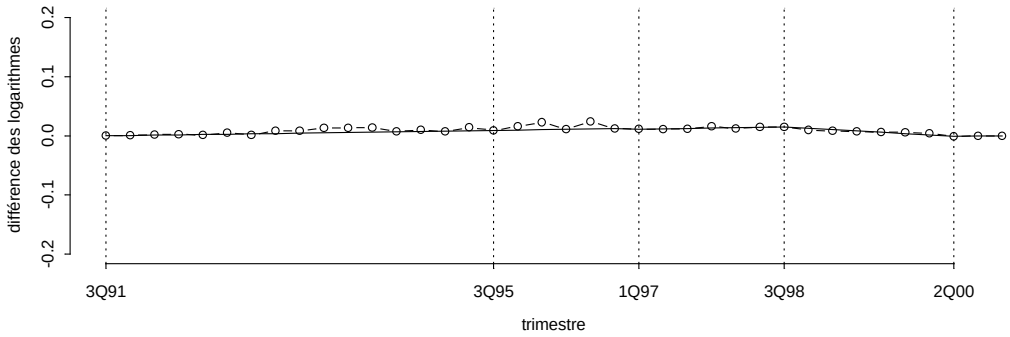
Différence entre BESTA 29 corrigée et non corrigée



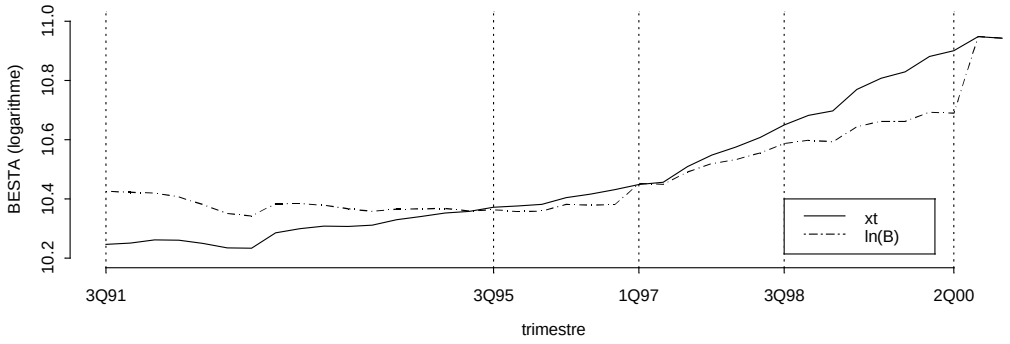
Logarithme de la BESTA 55 corrigée et non corrigée



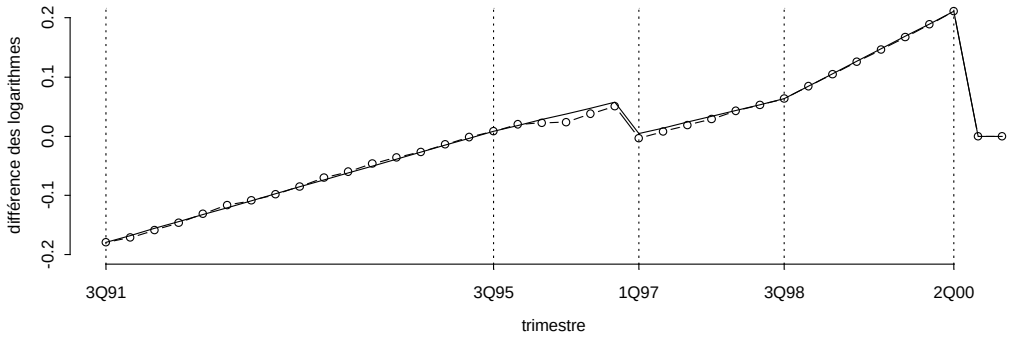
Différence entre BESTA 55 corrigée et non corrigée

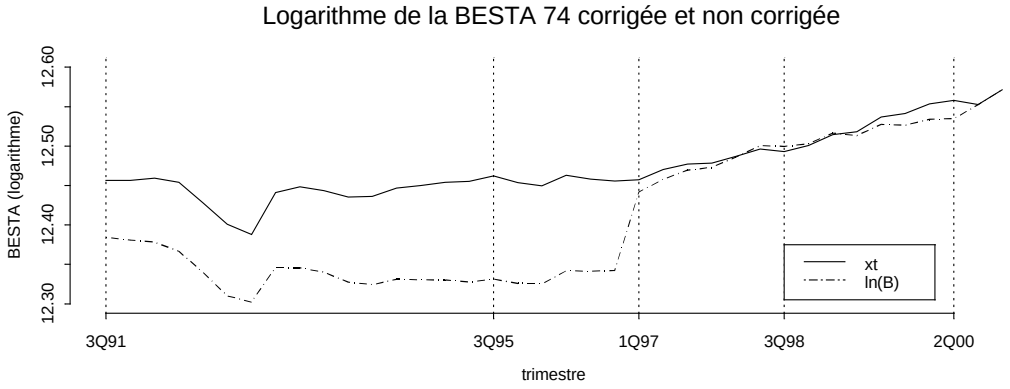


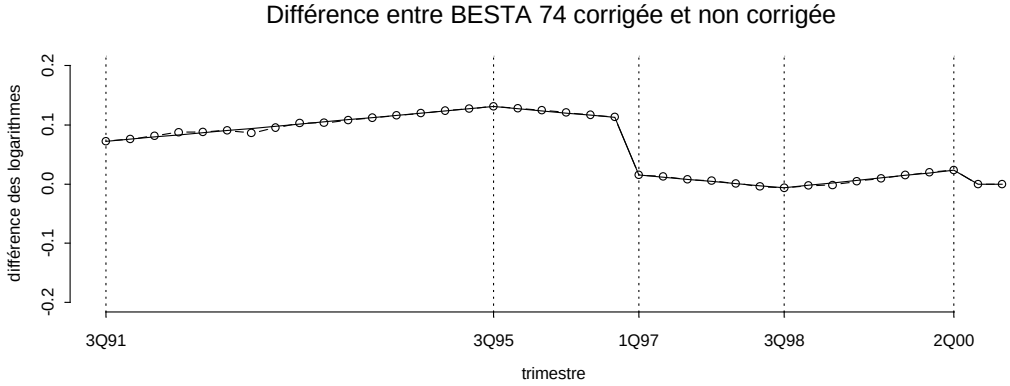
Logarithme de la BESTA 72 corrigée et non corrigée

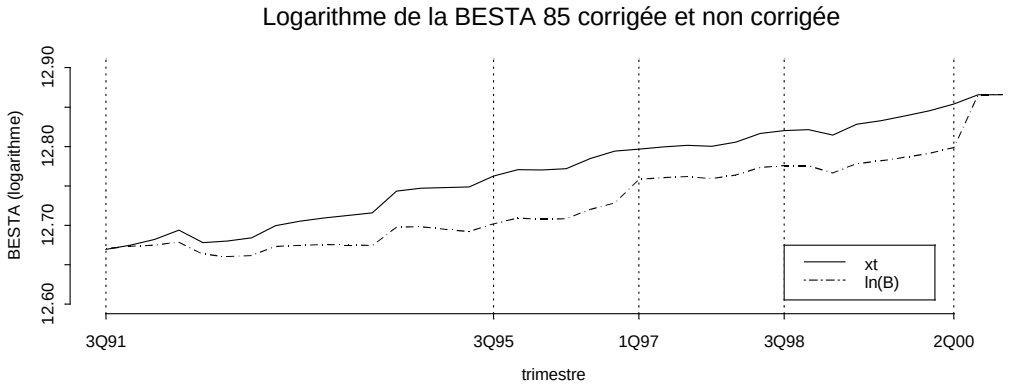


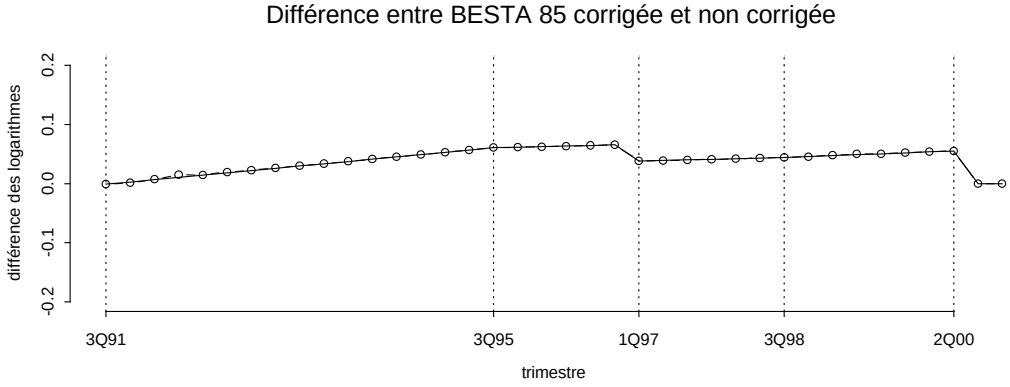
Différence entre BESTA 72 corrigée et non corrigée




Logarithme de la BESTA 74 corrigée et non corrigée
BESTA (logarithme)
12.30
12.40
12.50
12.60
xt
ln(B)
3Q91
3Q95
1Q97
3Q98
2Q00
trimestre


Différence entre BESTA 74 corrigée et non corrigée
différence des logarithmes
-0.2
-0.1
0.0
0.1
0.2
3Q91
3Q95
1Q97
3Q98
2Q00
trimestre


Logarithme de la BESTA 85 corrigée et non corrigée
BESTA (logarithme)
12.60
12.70
12.80
12.90
xt
ln(B)
3Q91
3Q95
1Q97
3Q98
2Q00
trimestre


Différence entre BESTA 85 corrigée et non corrigée
différence des logarithmes
-0.2
-0.1
0.0
0.1
0.2
3Q91
3Q95
1Q97
3Q98
2Q00
trimestre

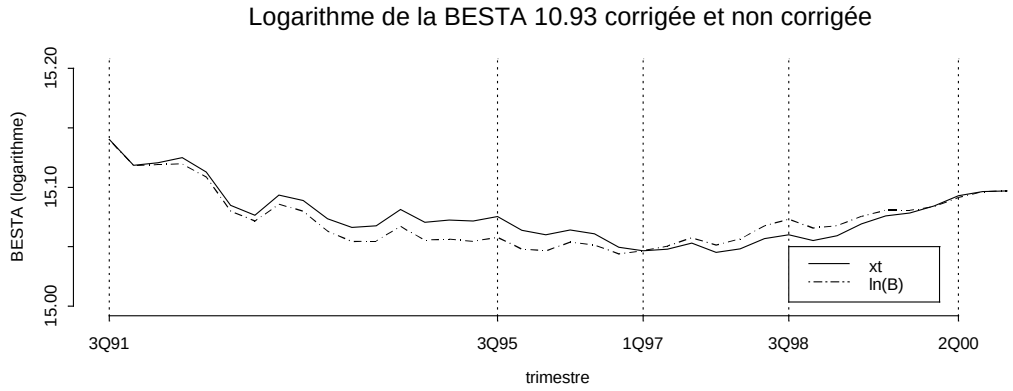
Analyse par secteurs

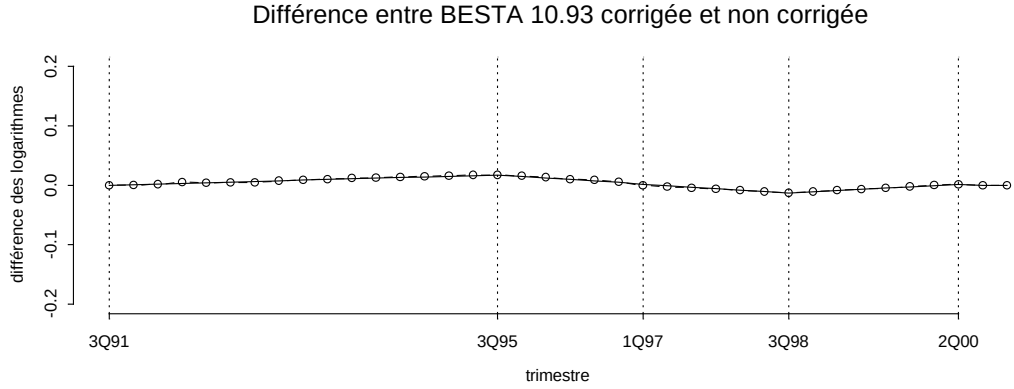
BESTA, emploi total: par secteurs

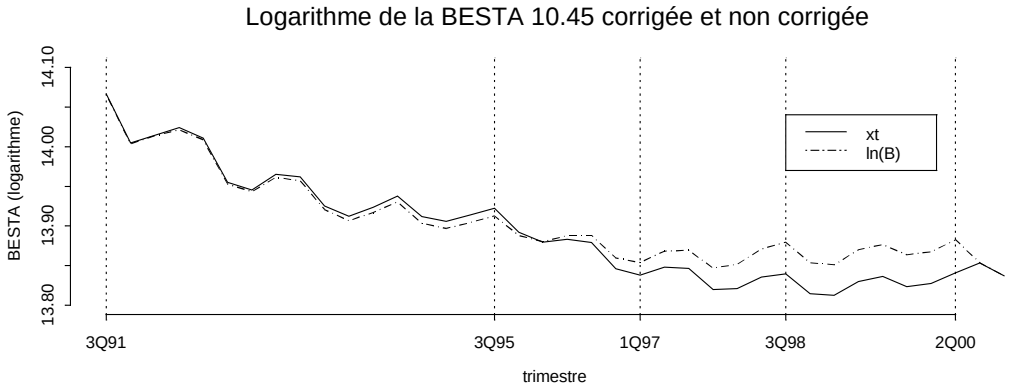
branche	p1	h1	p2	h2	p3	h3	p4	h4
10-93	0.0011	0.0001	-0.0025	0.0175	-0.0025	0.0166	0.0021	-0.0129
10-45	0.0006	0.0003	-0.0045	0.0099	-0.0045	0.0140	-0.0002	-0.0401
50-93	0.0013	-0.0001	-0.0016	0.0210	-0.0016	0.0180	0.0030	-0.0013

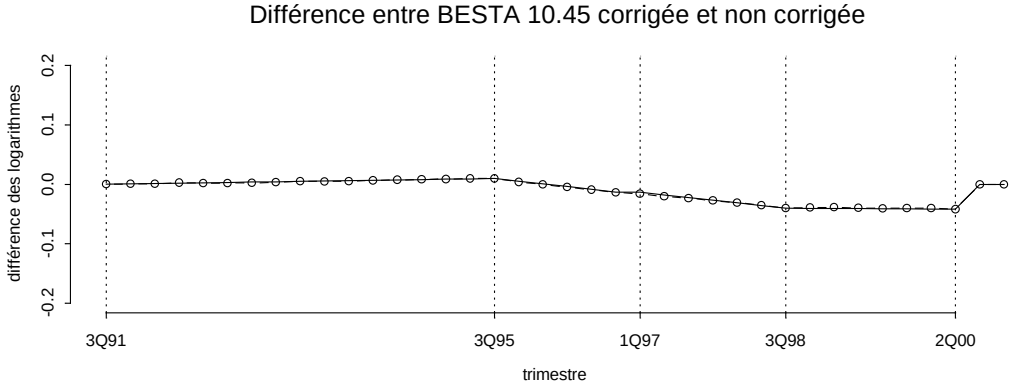
Pente et ordonnée à l'origine des corrections (échelle logarithmique) pour chaque période
(1 = 3Q91-3Q95 ; 2 = 3Q95-1Q97 ; 3 = 1Q97-3Q98 ; 4 = 3Q98-2Q00)

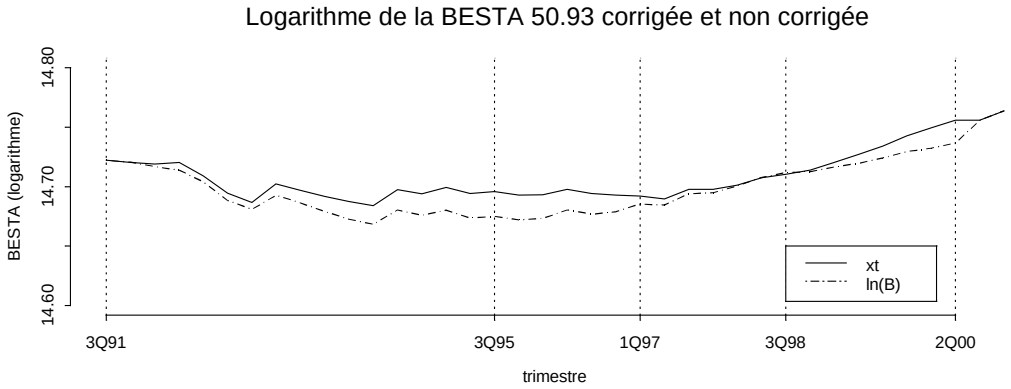
Le tableau ci-dessus montre la transformation calculée selon le même schéma pour le total et les totaux par secteurs. Cette transformation n'est pas exacte, car les totaux par secteurs sont calculés comme somme des branches adaptées concernées. Pourtant, ainsi que les figures ci-dessous le montrent, le résultat est satisfaisant.

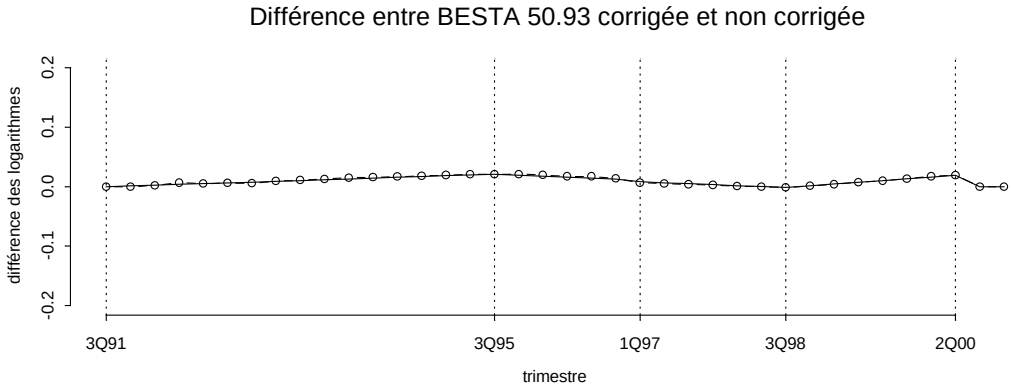





Logarithme de la BESTA 10.45 corrigée et non corrigée
BESTA (logarithme)
14.10
14.00
13.90
13.80
xt
ln(B)
3Q91
3Q95
1Q97
3Q98
2Q00
trimestre


Différence entre BESTA 10.45 corrigée et non corrigée
différence des logarithmes
0.2
0.1
0.0
-0.1
-0.2
3Q91
3Q95
1Q97
3Q98
2Q00
trimestre


Logarithme de la BESTA 50.93 corrigée et non corrigée
BESTA (logarithme)
14.80
14.70
14.60
xt
ln(B)
3Q91
3Q95
1Q97
3Q98
2Q00
trimestre


Différence entre BESTA 50.93 corrigée et non corrigée
différence des logarithmes
0.2
0.1
0.0
-0.1
-0.2
3Q91
3Q95
1Q97
3Q98
2Q00
trimestre

ANNEXE 5 : Programmes S-Plus

Les deux programmes suivants produisent les résultats donnés à l'annexe 4.

- corbesta.ssc définit une fonction S-plus corbesta faisant les dessins et calculant les paramètres définis à la formule (2.12).
- adapt.ssc appelle corbesta pour chaque branche et construit le fichier des paramètres (tableaux de l'annexe 4).

```
##############
# corbesta.ssc
##############
# Calcul de la correction induite par l'adaptation aux BZ et changement
   d'échantillon
# pour les branches saisonnières

corbesta <-function(BESTA, branche){
BZ <- BESTA$B[c(1,17,29)]                # années 91,95,98
B<-BESTA$Bnc[c(1,17,29)]                   # 3e trim. 91,95,98
B2Q00n <- BESTA$B[36]                     # nouvel éch. 2e trim. 00
B2Q00a <- BESTA$Bnc[36]                   # ancien éch. 2e trim. 00
# correction du saut au 1Q97:
c1Q97 <- ((BESTA$Bnc[19]/BESTA$Bnc[18] +
   BESTA$Bnc[27]/BESTA$Bnc[26])/2)/(BESTA$Bnc[23]/BESTA$Bnc[22])


cc <- log(BZ/B)
c00 <- log(B2Q00n/B2Q00a)
d <- c((cc[2]-cc[1])/16, (cc[3]-cc[2])/12)
u97<- log(c1Q97)
d00 <- (c00-cc[3])/7

trim <-c(3,4,rep(1:4,8),1:4)
ann <-
   c(rep(91,2),rep(92,4),rep(93,4),rep(94,4),rep(95,4),rep(96,4),rep(97,4),r
   ep(98,4),rep(99,4),rep("00",4))
trimQ <-paste(trim,"Q",sep="")
trimQ <-paste(trimQ,ann,sep="")

names(BESTA) <- c("Bnc","B")
row.names(BESTA) <-trimQ
BESTA

BESTA$lnBnc <-log(BESTA$Bnc)
BESTA$xt <-log(BESTA$B)

indi <-c(1,17,23,29,36)         # trimestres spéciaux
   (1Q97,3Q95,1Q97,3Q98,2Q00)
indi
mesticks <-trimQ[indi]
mesnoms <-c("xt","ln(B)")
difference <- BESTA$xt-BESTA$lnBnc
range(difference)
BESTA$difference<-difference
```

```
# Figure
par(mfrow=c(2,1))
y1<-min(min(BESTA$xt),min(BESTA$lnBnc))
y1 <- floor(10*y1)/10
y2 <-max(max(BESTA$xt),max(BESTA$lnBnc))
y2 <- ceiling(10*y2)/10

# position de la légende
y3 <- y1 + (y2-y1)/4
if (min(BESTA$lnBnc[33],BESTA$xt[33]) < y3 ) y3 <- y2 - (y2-y1)/5

titre <- paste("Logarithme de la BESTA",branche,"corrigée et non corrigée")
plot(1:38,BESTA$xt,type="l", xlab="trimestre",ylab="BESTA (logarithme)",
   axes=F,
   main= titre,ylim=c(y1,y2))
lines(1:38,BESTA$lnBnc, lty=3)
axis(1, at=indi,label=mesticks) ; axis(2)
abline(v=indi, lty=2) ; legend(29,y3,legend=mesnoms, lty=c(1,3))

# les 3 lignes suivantes donnent le calcul exact
# k2 <- (d[2]-(difference[22]-cc[2])/5)/u97
# k3 <-(d[2]-(difference[29]-difference[23])/6)/u97
# k0 <- (2*difference[23]-difference[29]-cc[2])/u97

# valeurs théoriques
k2 <- 1/12 ; k3 <- 1/12 ; k0 <- 1
# titre <- paste(as.character(round(k2,4)),
                 as.character(round(k3,4)),as.character(round(k0,4)))

# pentes et ordonnées à l'origine des segments
p1 <- d[1]
h1 <- cc[1]
p2 <- d[2] - k2*u97
h2 <- cc[2]
p3 <- d[2] - k3*u97
h3 <- cc[2]+ k0*u97
p4 <- d00
h4 <- cc[3]
tQ1 <- ( 0:(indi[2]-1) )*p1 + h1                        # période 3Q91 à 3Q95
tQ2 <- ( 1:(indi[3]-indi[2]-1) )*p2 + h2                # période 3Q95 à 4Q96
tQ3 <- ( (indi[3]-indi[2]):(indi[4]-indi[2]) )*p3 + h3 # période 1Q97 à 3Q98
tQ4 <- ( 1:(indi[5]-indi[4]) )*p4 + h4                  # période 3Q98 à 2Q00
tQ <-c( tQ1 , tQ2 , tQ3, tQ4 , 0, 0)
titre <-paste("Différence entre BESTA",branche,"corrigée et non corrigée")
plot(1:38, difference, xlab="trimestre",ylab="différence des logarithmes",
   axes=F, ylim=c(-0.2,0.2),
   type="b", main= titre)
   axis(1, at=indi,label=mesticks)
axis(2)
abline(v=indi, lty=2)
lines(1:38,tQ)

return(branche,p1,h1,p2,h2,p3,h3,p4,h4)
}
# fin de la fonction corbesta
```

```
##############
# adapt.ssc #
##############
# Fichier input: Besta Emploi total.xls
# (feuilles 1:série adaptée et 2:série reconstituée)

# Btot =besta emploi total série adaptée
# nbesta: data.frame contenant les no de branches selon la feuille 1

names(Btot) <- as.character(nbesta$V1)
dim(Btot)

# Btotnc =besta emploi total série non corrigée
# ordre différent des col. pour la feuille 2

ordnc <- c(3,5:28,31:34,36:40,42:44,46:53,55:58,4,30,35,41,45,54)
names(Btotnc) <-c(names(Btot)[ordnc],"B200","B300","B500")


# Valeurs branche ** (** = no branche)
# B**: fichier à 2 colonnes Bnc et B
# Bnc: Valeurs BESTA recomposées
# B: Valeurs BESTA adaptées

# Appels à corbesta

B24 <- data.frame(Btotnc$B24,Btot$B24)
names(B24) <- c("Bnc","B")
cor24 <- corbesta(B24,"24")

B28 <- data.frame(Btotnc$B28,Btot$B28)
names(B28) <- c("Bnc","B")
cor28 <- corbesta(B28,"28")

B29 <- data.frame(Btotnc$B29,Btot$B29)
names(B29) <- c("Bnc","B")
cor29 <- corbesta(B29,"29")

B45 <- data.frame(Btotnc$B45,Btot$B45)
names(B45) <- c("Bnc","B")
cor45 <- corbesta(B45,"45")

B55 <- data.frame(Btotnc$B55,Btot$B55)
names(B55) <- c("Bnc","B")
cor55 <- corbesta(B55,"55")

B72 <- data.frame(Btotnc$B72,Btot$B72)
names(B72) <- c("Bnc","B")
cor72 <- corbesta(B72,"72")

B74 <- data.frame(Btotnc$B74,Btot$B74)
names(B74) <- c("Bnc","B")
cor74 <- corbesta(B74,"74")
```

```
B85 <- data.frame(Btotnc$B85,Btot$B85)
names(B85) <- c("Bnc","B")
cor85 <- corbesta(B85,"85")

# Fichier des paramètres de la fonction de passage de ln(B) à xt
#

corbr <- rbind(data.frame(cor24),data.frame(cor28),data.frame(cor29) )
corbr <- rbind(corbr,data.frame(cor45),data.frame(cor55) )
corbr <- rbind(corbr,data.frame(cor72),data.frame(cor74),data.frame(cor85)
   )
corbp <- corbr
corbp[,2:9] <- round(corbr[,2:9] ,4)
corbp
#########################################################


# total
B10.93 <- data.frame(Btotnc$B500,Btot$B10.93)
names(B10.93) <- c("Bnc","B")
cor10.93 <- corbesta(B10.93,"10.93")
data.frame(cor10.93)

# secteurs
B10.45 <- data.frame(Btotnc$B200,Btot$B10.45)
names(B10.45) <- c("Bnc","B")
cor10.45 <- corbesta(B10.45,"10.45")
data.frame(cor10.45)

B50.93 <- data.frame(Btotnc$B300,Btot$B50.93)
names(B50.93) <- c("Bnc","B")
cor50.93 <- corbesta(B50.93,"50.93")
data.frame(cor50.93)

cortotp <- cortot
cortot <-
   rbind(data.frame(cor10.93),data.frame(cor10.45),data.frame(cor50.93) )
cortotp[,2:9] <- round(cortot[,2:9] ,4)
cortotp
```

Bibliographie

[1] European Central Bank (2000). *Seasonal Adjustment of Monetary Aggregates and HICP for the Euro Area.*, 93pp

[2] EUROSTAT (1998). *Seasonal Adjustment Methods : A Comparison for Industry Statistics, revised edition*. European Business Trends, Site CIRCA, EUROSTAT, Luxembourg.

[3] Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C., Chen, B.C. (1998). New Capabilities and Methods of the X-12-ARIMA Seaonal-Adjustment Program. *J. of Business & Economic Statistics,* **16,** 2.

[4] Harvey, A. (1993) *Time Series Models (2nd Ed).*The MIT Press, Cambridge, MA.

[5] Kirchner, R. (1999). Auswirkungen des neuen Saisonbereinigungsverfahrens Census X-12-ARIMA auf die aktuelle Wirtschaftsanalyse in Deutschland. *Volkswirtschaftliche Forschungsgruppe der Deutschen Bundesbank, Diskussionspapier 7/99*

[6] Kaiser, R. and Maravall, A. (to appear) *Measuring Business Cycles in Economic Time Series*. Lecture Notes in Statistics, Vol. 154, Springer. ISBN : 0-387-95112-1

[7] Planas, Ch. (1997). *Applied Time Series Analysis : Modelling, Forecasting Unobserved Components Analysis and the Wiener-Kolmogorov Filter.* Site CIRCA, EUROSTAT, Luxembourg.

[8] Ladiray, D. & Quenneville, B. (1999). *Comprendre la méthode X11.* http ://www.census.gov/pub/ts/papers/x11doc.pdf.

[9] Schips, B, Stier, W.(1995) The CENSUS-X-11 Procedure. Theory, Assessment and Alternatives. *Swiss Statistics.* Swiss Federal Statictical Office, Neuchâtel.

[10] U.S. Census Bureau (2000). *X-12-ARIMA Reference Manual (Version 0.2.7).* http ://www.census.gov/srd/www/x12a/x12down_pc.html

[11] La Statistique de l'emploi : Bases Méthodologiques. *OFS* 1997, 31pp

[12] Revision der Beschäftigungstatistik Grobkonzept : 1999/2000. *OFS* 1999, 56pp

[13] Révision de la Statistique de l'emploi, Conception des détails. *OFS* 2000, 53 pp

[14] La Statistique des places vacantes : Bases Méthodologiques. *OFS* 2000, 35 pp

[15] Saucy, F. Statistique de l'emploi, désaisonnalisation des données. Note. *Population et emploi, OFS* 2000, 5p

[16] Salamin, P.-A. Evaluation de la Statistique de l'emploi. *Rapport de Méthodes, OFS* 1997, 63 pp

[17] Salamin, P.-A. Renouvellement de l’échantillon de la Statistique de l’emploi. *Rapport de Méthodes, OFS* (en prép.)

Rapports de méthodes du Service de méthodes statistiques

Hulliger, B., Peters, R. (1994). Enquête sur la structure des salaires en Suisse : stratégie d'échantillonnage pour le secteur privé.

Peters, R., Hulliger, B. (1994). La technique de pondération des données : application à l'enquête suisse sur la santé.

Schwendener, P. (1995). Verbrauchserhebung 1990 - Vertrauensintervalle.

Hulliger, B. (1995). Konjunkturelle Mietpreiserhebung : Stichprobenplan und Schätzverfahren.

Peters, R. (1995). Enquête de structure sur les loyers : établissement d'un plan d'échantillonnage stratifié.

Salamin, P.A. (1995). Estimation des flux pour le module II des comptes globaux du marché de travail.

Hulliger, B. (1996). Gütertransportstatistik 1993 : Schätzverfahren mit Kompensation der Antwortausfälle.

Hulliger, B., Peters, R. (1996). Enquête sur le comportement de la population suisse en matière de transport en 1994 : plan d'échantillonnage et pondération.

Peters, R. (1996). Schéma de pondération des indices PAUL.

Peters, R., Hulliger, B. (1996). Schätzverfahren für die Lohnstruktur-Erhebung 1994 / Procédure d'estimation pour l'enquête de 1994 sur la structure des salaires.

Hulliger, B. (1996). Haushalterhebung Familie 1994 : Stichprobenplan, Stichprobenziehung und Reservestichproben.

Comment, T., Hulliger, B., Ries, A. (1996). Gewichtungsverfahren für die Schweizerische Arbeitskräfteerhebung (1991-1995).

Peters, R. (1996). Pondération des données de l'enquête sur la famille en Suisse.

Peters, R. (1997). Enquête 1996 sur la structure des salaires en Suisse : établissement du plan d'échantillonnage.

Peters, R. (1997). Etablissement du plan d'échantillonnage pour l'enquête 1996 sur la recherche et le développement dans l'économie privée en Suisse.

Salamin, P.A. (1997). Evaluation de la Statistique de l'emploi.

Bender, A., Hulliger, B. (1997). Enquête suisse sur la population active : rapport de pondération pour 1996.

Peters, R. (1998). Extrapolation des données de l'enquête de structure sur les loyers.

Salamin, P.A. (1998). Etablissement d'une clef de passage pondérée entre l'ancienne (NGAE 85) et la nouvelle nomenclature (NOGA 95) générale des activités économiques.

Hulliger, B., Kassab, M. (1998). Evaluation of Estimation Methods for the Survey on Environment Protection Expenditures of Swiss Communes.

de Rossi, F.-X. (1998). Méthodes statistiques pour le compte routier suisse.

Hulliger, B., Eichenberger, P. (2000). Stichprobenregister für Haushalterhebungen : Umstellung auf Telefonnummern ohne Namen und Adressen, Abläufe für Erstellung und Stichprobenziehung. Bestellnummer : 338-0000

Renaud, A. (2001). Statistique suisse des bénéficiaires de l'aide sociale. Plan d'échantillonnage des communes. Numéro de commande : 338-0001

Hüsler, J., Müller, S. (2001). Sclussbericht Betriebszählung 1995 (BZ 95). Mehrfach imputierte Umsatzzahlen. Bestellnummer : 338-0002

Graf, M. (2001). Désaisonnalisation. Aspects méthodologiques et application à la statistique de l'emploi. Numéro de commande : 338-0003

Rapports de méthodes du Service de méthodes statistiques

Une liste complète des titres de cette série de publications figure à la fin de ce rapport.
Les rapports de méthodes peuvent être commandés auprès de :

Office fédéral de la statistique
Service des publications
Espace de l'Europe 10
CH-2010 Neuchâtel
Tél. : (+41 32) 71 36060
Fax : (+41 32) 71 36061
Internet : Order@bfs.admin.ch

Methodenberichte des Dienstes Statistische Methoden

Ein vollständige Liste der Titel dieser Serie befindet sich am Ende dieses Berichts.
Die Methodenberichte können bestellt werden bei :

Bundesamt für Statistik
Publikationsdienst
Espace de l'Europe 10
CH-2010 Neuchâtel
Tel. : (+41 32) 71 36060
Fax : (+41 32) 71 36061
Internet : Order@bfs.admin.ch