# **Economie et Statistiques**Working papers du STATEC

juillet 2009

Auteur: Véronique ELTER STATEC

# La désaisonnalisation ou l'ajustement saisonnier

## Introduction

Des facteurs aussi bien climatiques qu'institutionnels, qui surviennent régulièrement à la même époque chaque année (jour, semaine, mois, trimestre, semestre) font varier les habitudes de consommation et de production. Ces variations perturbent l'interprétation conjoncturelle, qui met l'accent sur la tendance globale à long terme, plutôt qu'à l'état actuel brouillé par des effets saisonniers. La question de la saisonnalité se pose inévitablement lors de toute comparaison de l'évolution du PIB (Produit Intérieur Brut) d'un trimestre à l'autre. La désaisonnalisation d'une série économique consiste à utiliser des techniques mathématiques afin d'éliminer les phénomènes purement saisonniers.

Voici quelques exemples concrets d'effets saisonniers:

- Avant certaines fêtes (comme par exemple Noël, St. Valentin, Pâques), la consommation (achats de cadeaux, nourriture, boissons) augmente considérablement.
- Les déplacements touristiques connaissent une large croissance lors des congés scolaires, ce qui se répercute entre autres sur les secteurs de la restauration et de l'hôtellerie.
- La production agricole présente des niveaux de crête liés aux mois de récolte.
- Dans le domaine de la construction les variations saisonnières sont dues au climat et aux congés collectifs.

Le Statec n'engage en aucun cas sa responsabilité concernant la pertinence, l'exactitude, l'exhaustivité ou la qualité du contenu de la présente publication. Il ne pourra donc être tenu responsable des dommages causés par l'utilisation de ces informations, y compris de celles qui pourraient s'avérer incomplètes ou incorrectes. Les articles publiés dans la série « Économie et statistiques» n'engagent que leurs auteurs.

juin 2009

2

- La fin de l'année fiscale peut également apporter des variations saisonnières dans les séries.
- Le chômage augmente chaque année au cours des mois d'hiver. En hiver, à cause des températures basses, la construction, l'agriculture, le tourisme (et donc aussi le transport, l'hôtellerie et la restauration) connaissent une importance inférieure qu'en été, ce qui réduit le nombre de travailleurs nécessaires et entraîne donc un chômage saisonnier.
- Les prix des vêtements diminuent chaque année en janvier et en juillet, à cause des soldes.

En gros, tous les effets saisonniers peuvent être regroupés selon trois catégories:

- la saisonnalité <u>climatique</u>, attribuable aux variations saisonnières du climat (par exemple agriculture, consommation d'électricité pour le chauffage ou la climatisation)
- la saisonnalité <u>institutionnelle</u>, attribuable aux conventions sociales ou aux règles administratives (par exemple calendrier scolaire, année fiscale)
- la saisonnalité <u>induite</u>, attribuable à la saisonnalité d'autres secteurs (par exemple industrie alimentaire, fabrication de jouets)

L'évolution de la saisonnalité dépend du progrès technique (par exemple développement de matériaux et de techniques adaptés à la construction en hiver ou de céréales résistantes au froid), de changements institutionnels et de la modification de la structure des séries (par exemple l'emploi dans l'industrie; cette activité économique moins dépendante des saisons, a connu une croissance considérable, au détriment de l'emploi dans l'agriculture et la pêche, qui eux sont plus affectés par des effets saisonniers).

Dans le cadre de l'analyse économique conjoncturelle, il convient de savoir dans quelle phase du cycle économique se situer (crise, récession, reprise, expansion), au-delà de tout effet saisonnier. Voilà pourquoi, lors de la désaisonnalisation, la série brute est décomposée en plusieurs composantes: la tendance à long terme, le cycle, la composante saisonnière et la composante irrégulière. En outre, la série brute peut être ajustée pour l'effet de Pâques (qui est une fête de date variable, soit au premier, soit au deuxième trimestre) et pour le nombre de jours ouvrables (un dimanche de plus ou de moins dans un trimestre peut affecter la production).

En fait, désaisonnaliser signifie tout simplement extraire la composante saisonnière de la série brute. Par définition, la somme des effets saisonniers est nulle sur l'intervalle de temps considéré, donc les facteurs saisonniers ne jouent pas sur la tendance. Sans ajustement saisonnier des séries présentant de fortes variations saisonnières, il est délicat de comparer chaque trimestre avec le trimestre précédent de la même année, mais uniquement avec les mêmes trimestres des années précédentes.

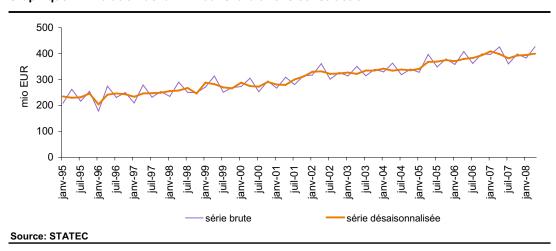
L'ajustement saisonnier facilite donc les comparaisons, à travers le temps (possibilité de comparer le premier trimestre, qui possède le mois court de février au quatrième trimestre, qui contient le mois de décembre à consommation augmentée (Noël, Nouvel An)) et à travers l'espace (possibilité de comparer les différentes économies mondiales, indépendamment du climat et des festivités locales).

Les outils informatiques utilisés pour la désaisonnalisation, peuvent être classés en deux grands groupes. Au Canada et aux Etats-Unis prédomine le programme X 12-ARIMA, développé par David F. Findley et Brian C. Monsell du Bureau of Census des Etats-Unis. X 12-ARIMA utilise une méthode non paramétrique d'origine empirique, basée sur la méthode des moyennes mobiles ne nécessitant pas le recours à un modèle. La modélisation (partie ARIMA de cet outil) n'est utilisée que pour l'étude des extrémités de la série. X 12-ARIMA est gratuit et peut être téléchargé sur le site <a href="http://www.census.gov/srd/www/x12a/">http://www.census.gov/srd/www/x12a/</a> du Bureau of Census des Etats-Unis.

Les pays européens utilisent en majorité le programme TRAMO-SEATS, développé par Víctor Gómez du «Ministerio de Economía de España» et Augustín Maravall de la «Banco de España». TRAMO-SEATS repose sur une méthode paramétrique, basée sur la modélisation économétrique de la série chronologique. Les modèles qui peuvent être utilisés sont fournis par la théorie des processus stochastiques (modèles ARIMA: «Auto-Regressive Integrated Moving Average»). TRAMO-SEATS est gratuit et peut être téléchargé sur le site http://www.bde.es/servicio/software/econome.htm de la «Banco de España». Le Statec emploie également le programme TRAMO-SEATS, à l'aide de l'interface Demetra, mise au point par EUROSTAT. Demetra est gratuit et peut être téléchargé sur le site http://circa.europa.eu/irc/dsis/eurosam/info/data/demetra.htm d'EUROSTAT.

Le graphique suivant illustre la VAB (Valeur Ajoutée Brute) de la branche construction de janvier 1995 à avril 2008. La série brute présente une hausse considérable au premier trimestre et une baisse importante au troisième trimestre de chaque année. Cependant une baisse durant l'été ne signifie pas un ralentissement conjoncturel de l'économie, mais correspond à une période calme (congé collectif dans le secteur du bâtiment). Au fil des années, une évolution positive de l'économie dans ce secteur peut être constatée. Sur des périodes courtes, ceci est nettement plus visible à l'aide de la série désaisonnalisée ou de la tendance à long terme qu'avec la série brute.

Graphique 1: Evolution de la VAB dans la branche construction



Dans la suite de ce document, on parlera de «désaisonnalisation/désaisonnalisé» au lieu de «correction/corrigé des variations saisonnières», afin de simplifier la lecture.

## Sommaire

Le but du présent document est de fournir une introduction, ainsi qu'un aperçu des différentes méthodes utilisées sur l'ajustement des variations saisonnières des séries temporelles.

La première partie explique d'une manière facilement compréhensible la signification du terme ajustement des variations saisonnières et illustre celui-ci à l'aide de quelques exemples.

La deuxième partie développe les différents outils mathématiques nécessaires afin de faciliter la compréhension du mécanisme sous-jacent à la désaisonnalisation.

La troisième partie argumente les choix méthodologiques que le Statec a effectués, en vue de l'ajustement des variations saisonnières de ses séries temporelles concernant tous les indicateurs mensuels et trimestriels. La dernière partie donne un exemple concret de la désaisonnalisation à l'aide de l'interface Demetra.

Ce document se lit plutôt comme un petit manuel didactique et illustré par ces exemples.

Mots clés: désaisonnalisation, moyennes mobiles, modèles ARIMA.

# 1. Méthodologie de la désaisonnalisation

Etant donné que la plupart des séries relevant du domaine économique constituent des séries temporelles, il est d'abord procédé à la définition de celles-ci.

# 1.1 Séries temporelles et composantes

Définition: Série temporelle

La suite régulièrement ordonnée d'observations  $(X_t)_{t \in \tau}$  d'une variable X à différentes dates t est appelée <u>série temporelle</u>.

Selon la nature de l'ensemble  $\, au$  , les problèmes seront différents:

- $\tau$  peut être continu (par exemple électroencéphalogramme),
- $\tau$  peut être discret (cas usuel en économie).

Exemples de séries temporelles:

```
PIB trimestriel ( \tau = \{1,2,3,4\}), IPC (indice des prix à la consommation) mensuel ( \tau = \{1,2,...,12\}), taux de chômage mensuel ( \tau = \{1,2,...,12\}).
```

La série brute  $X_t$  se décompose en plusieurs éléments de base:

- la composante tendance-cycle à long terme  $(TC)_t$  qui regroupe la tendance à long terme et le cycle, vu qu'une distinction entre ces deux composantes s'avère souvent très difficile pour les séries courtes,
- la composante saisonnière  $S_t$  et
- la composante irrégulière  $I_{t}$ .

La tendance à long terme est l'évolution, à la hausse ou à la baisse, de la série observée sur plusieurs décennies. Cette composante montre les variations de fond, observées sur une longue période de temps. C'est une composante lisse, qui ne traduit pas d'évolution brusque.

Le cycle est le mouvement lisse et périodique de la série observé souvent autour de la composante tendance à long terme. Il se caractérise par une alternance de périodes d'expansion et de récession dont la longueur varie entre 3 et 10 ans et parfois même plus (cycle d'affaires).

La composante saisonnière représente les fluctuations (mensuelles ou trimestrielles) à l'intérieur d'une même année qui se répètent plus ou moins régulièrement d'une année à l'autre.

La composante irrégulière représente les fluctuations résiduelles et erratiques qui ne peuvent être attribuées aux autres composantes (par exemple grèves, catastrophes naturelles (foudre, ouragan, grêle, inondations, tremblements de terre), déclenchements de guerres, crises économiques ou politiques subites, erreurs de mesure inaperçues, fins de mesures fiscales, de programmes d'aide).

5

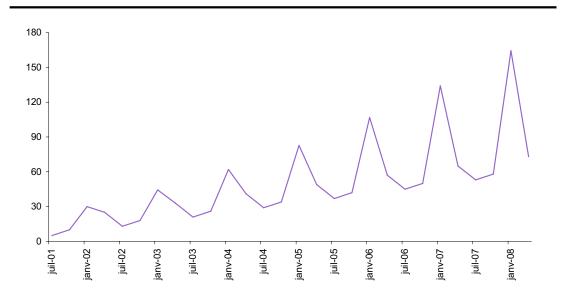
Ainsi la série corrigée des variations saisonnières, notée CVS, est le regroupement de la tendancecycle à long terme et de la composante irrégulière. Elle peut également être définie par la série brute, à laquelle a été enlevée la composante saisonnière et les effets de calendrier, ce qui revient au même.

Afin d'éliminer les variations saisonnières d'une série temporelle, elle est décomposée, soit en un modèle multiplicatif, additif ou pseudo-additif.

Le modèle multiplicatif est défini par:  $X_t \coloneqq (TC)_t \cdot S_t \cdot I_t$  .

Le modèle multiplicatif est utilisé si les composantes sont dépendantes entre elles, c'est-à-dire si les oscillations saisonnières croissent et décroissent avec le niveau des séries.

Graphique 1.1.1: Série nécessitant un modèle multiplicatif de décomposition



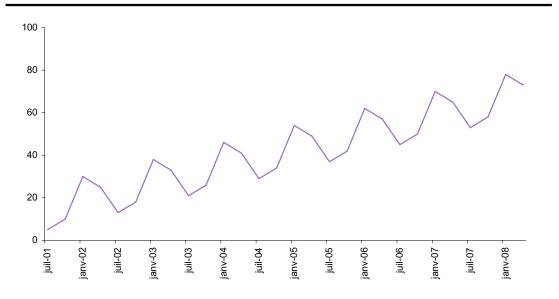
Source: STATEC

Contrairement, le modèle additif est défini par:

$$X_t := \big(TC\big)_t + S_t + I_t \,.$$

Le modèle additif suppose que les composantes sont indépendantes entre elles, c'est-à-dire que les oscillations saisonnières sont indépendantes du niveau des séries.

Graphique 1.1.2: Série nécessitant un modèle additif de décomposition



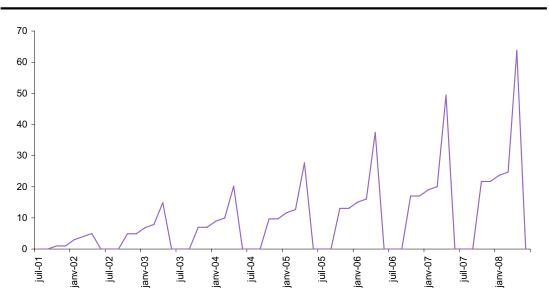
Source: STATEC

Le modèle pseudo-additif est défini par:

$$X_t := (TC)_t \cdot (S_t + I_t - 1).$$

Le modèle pseudo-additif admet que les composantes saisonnières et irrégulières sont indépendantes entre elles, mais qu'elles sont dépendantes du niveau de la tendance-cycle à long terme.

Graphique 1.1.3: Série nécessitant un modèle pseudo-additif de décomposition



**Source: STATEC** 

juin 2009

Le plus souvent, les modèles additif et multiplicatif seront utilisés. Pour des valeurs positives des composantes, le modèle multiplicatif se réduit à un modèle additif en prenant le logarithme des deux côtés de l'équation:

$$\log(X_t) = \log[(TC)_t] + \log(S_t) + \log(I_t).$$

Dans chacun des trois cas, une équation à trois inconnues  $(TC)_t$ ,  $S_t$  et  $I_t$  doit être résolue (en négligeant les effets de calendrier). Donc il n'existe pas de solution unique et différentes méthodes de désaisonnalisation procureront différents résultats.

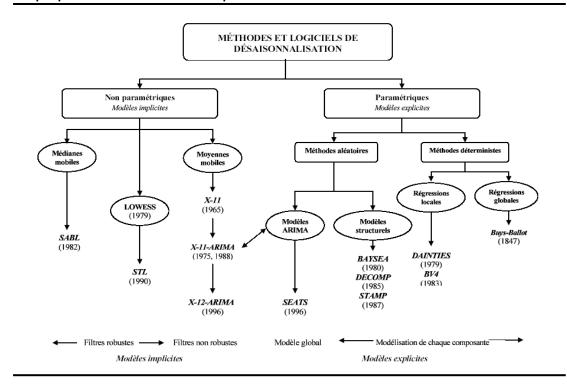
#### 1.2 Aperçu

La nature des hypothèses sur la série et les composantes implique des méthodes et donc des logiciels différents. De manière simplifiée, les différentes méthodes sont regroupées en deux catégories:

- les méthodes paramétriques (modélisation de chaque composante par des équations dépendant de paramètres) et
- les méthodes non paramétriques (modélisation par des hypothèses).

Une bonne visualisation des différentes méthodes disponibles actuellement est donnée par le tableau suivant, issu du document [7] de la bibliographie:

Graphique 1.2.1: Méthodes utilisées pour la désaisonnalisation



Dans la suite, plusieurs de ces méthodes seront étudiées.

Vu l'importance de la notion d'un bruit blanc, dans la construction de ces méthodes, un rappel de la définition de celui-ci est donné:

#### Définition: Bruit blanc

Un processus stochastique discret  $(\mathcal{E}_t)_{t\in\mathcal{T}}$ , pour lequel les  $\mathcal{E}_t$  sont des variables aléatoires indépendantes de distribution identique (i.i.d.: «independant identically distributed) est appelé <u>bruit blanc</u> s'il satisfait aux conditions suivantes:

$$E[\varepsilon_t] = 0$$
 et

$$Cov(\varepsilon_t, \varepsilon_s) = \begin{cases} \sigma^2, t = s \\ 0, t \neq s \end{cases} \quad \forall t, s \in \tau.$$

#### 1.2.1 Méthode de régression

La tendance-cycle est supposée linéaire  $(TC)_t \coloneqq \alpha t + \beta$ , la composante saisonnière (pour une saisonnalité trimestrielle) est supposée constante  $S_t \coloneqq \sum_{i=1}^4 \gamma_i 1_i (t \operatorname{mod} 4)$ , où  $1_t(\cdot)$  est la fonction caractéristique, et la composante irrégulière est donnée par un bruit blanc  $I_t \coloneqq \varepsilon_t$ . Ces hypothèses donnent la décomposition suivante de la série brute:

$$X_t = \alpha t + \beta + \sum_{i=1}^4 \gamma_i 1_i (t \bmod 4) + \varepsilon_t \text{. Pour l'estimation des paramètres } \alpha \text{,} \quad \beta \quad \text{et} \quad \gamma_i,$$
 
$$i = 1, \dots, 4 \text{, la méthode des moindres carrés («OLS: ordinary least squares method») est utilisée.}$$

#### 1.2.2 Méthode des moyennes mobiles

#### Définition: Moyenne mobile

La <u>moyenne mobile</u> d'ordre p+f+1 et de coefficients  $\left\{\theta_i\right\}_{i=-p,\dots,+f}$ , où p et f sont des nombres entiers, d'une série temporelle  $\left(X_t\right)_{t\in\tau}$  est notée  $M\!\left[X_t\right]$  et est définie par:

$$M\big[X_{t}\big] \coloneqq \sum_{i=-p}^{+f} \theta_{i} X_{t+i} = \theta_{-p} X_{t-p} + \theta_{-p+1} X_{t-p+1} + \ldots + \theta_{f-1} X_{t+f-1} + \theta_{f} X_{t+f} \,.$$

La valeur de la série temporelle  $(X_t)_{t \in r}$  à l'instant t est donc remplacée par la moyenne pondérée de p valeurs dans le passé, de la valeur actuelle et de f valeurs dans le futur.

Exemple: La moyenne mobile d'ordre 4 (p=1 et f=2) est donnée par:

$$M[X_t] = \theta_{-1}X_{t-1} + \theta_0X_t + \theta_1X_{t+1} + \theta_2X_{t+2}$$
.

Cas particuliers:

Une moyenne mobile est dite centrée, si p = f .

Une moyenne mobile centrée est dite symétrique, si  $\theta_i=\theta_{_{-i}} \ \ \forall i=-p,\ldots,+f$  .

Il s'agit de trouver une moyenne mobile, c'est-à-dire des valeurs pour p, f et  $\theta_i$   $\forall i \in [-p, ..., +f]$ , qui conserve la tendance-cycle, mais élimine la composante saisonnière et la composante irrégulière. Cette moyenne mobile doit donc répondre au critère suivant:

$$M[X_t] = M[(TC)_t + S_t + \varepsilon_t] = M[(TC)_t] + M[S_t] + M[\varepsilon_t] \approx (TC)_t$$

Dans le contexte d'une décomposition de Fourier, il s'agit donc d'éliminer certaines fréquences.

Décomposition de Fourier et spectre

La décomposition de Fourier consiste en une décomposition d'une fonction en une somme de fonctions sinusoïdales. Il y a donc plusieurs fréquences dans la composition de la fonction. Désaisonnaliser signifie éliminer les fréquences saisonnières de la série temporelle. Pour les séries trimestrielles, la période est de 4, donc la fréquence est de

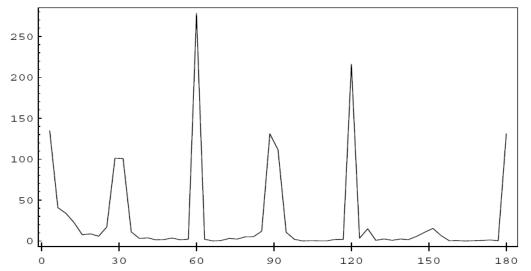
$$\frac{2\pi}{p\acute{e}riode}rad = \frac{2\pi}{4}rad = \frac{\pi}{2}rad = 90^{\circ}.$$

Pour les séries mensuelles, la période est de 12, donc la fréquence est de

$$\frac{2\pi}{p\acute{e}riode}rad = \frac{2\pi}{12}rad = \frac{\pi}{6}rad = 30^{\circ}.$$

Le spectre d'une série est le graphique, qui associe à chaque fréquence son importance dans la série. Les pics sur le spectre donnent donc les fréquences importantes.

Figure 1.2.2.1: Spectre d'une série mensuelle (pics spectraux pour les valeurs multiples entiers de 30°)



Source: Ladiray D. et Quenneville B., mars 2001. Seasonal Adjustment with the X-11 Method, Lecture Notes in Statistics 158, Springer, pages 1-256

De façon générale, la moyenne mobile  $\,M\,$  transforme toute série de la forme

$$X_{t} = R\sin(\omega t + \varphi),$$

où R est l'amplitude,  $\omega$  la fréquence et  $\varphi$  la phase de la série, en une série de la forme

$$M[X_t] = M[R\sin(\omega t + \varphi)] = RG(\omega)\sin[\omega t + \varphi + \Gamma(\omega)],$$

où  $G(\omega)$  est la fonction de gain et  $\Gamma(\omega)$  la fonction de déphasage de la moyenne mobile. Souvent  $\frac{\Gamma(\omega)}{\omega}$  est représenté, ce qui permet de mesurer le déphasage en nombre de périodes.

Afin d'illustrer ces concepts, la moyenne asymétrique d'ordre 3 est considérée:

$$M[X_t] = \frac{1}{3}(X_{t-2} + X_{t-1} + X_t) = \frac{1}{3}R[\sin(\omega t - 2\omega + \varphi) + \sin(\omega t - \omega + \varphi) + \sin(\omega t + \varphi)]$$

$$=\frac{1}{3}R[1+2\cos(\omega)]\sin(\omega t+\varphi-\omega).$$

Dans ce cas, la fonction de gain est donnée par  $G(\omega) = \frac{1 + 2\cos(\omega)}{3}$  et la fonction de déphasage

par 
$$\Gamma(\omega) = -\omega$$
 , donc  $\frac{\Gamma(\omega)}{\omega} = -1$  .

La moyenne asymétrique d'ordre 3, déphase donc la série d'une période, ce qui est gênant pour la détermination des points de retournement.

Il est facile de démontrer que les moyennes mobiles symétriques n'entraînent pas de déphasage. C'est la raison pour laquelle elles sont utilisées de façon privilégiée.

A présent, le problème initial sera de nouveau considéré.

Construction de moyennes mobiles

Il s'agit de construire la moyenne mobile, énoncée précédemment, qui conserve la composante tendance-cycle, mais élimine la tendance saisonnière et le bruit au maximum, c'est-à-dire qui satisfait:

$$M[X_t] = M[(TC)_t + S_t + \varepsilon_t] = M[(TC)_t] + M[S_t] + M[\varepsilon_t] \approx (TC)_t$$

D'abord, il faut trouver des contraintes pour les paramètres  $heta_i$  de la moyenne mobile.

Conservation des tendances:

Soit  $X_t := \sum_{j=0}^d \gamma_j t^j$  un polynôme de degré d en t . Afin de conserver cette tendance polynomiale,

l'équation suivante doit être satisfaite:

$$M[X_t] = X_t \Leftrightarrow \sum_{i=-p}^{+f} \theta_i X_{t+i} = X_t \Leftrightarrow \sum_{i=-p}^{+f} \theta_i \left[ \sum_{j=0}^{d} \gamma_j (t+i)^j \right] = \sum_{j=0}^{d} \gamma_j t^j$$

$$\iff \sum_{i=-p}^{+f} \theta_i \left[ \sum_{j=0}^d \gamma_j t^j + \sum_{j=0}^d \gamma_j \sum_{k=1}^j \binom{j}{k} t^{j-k} i^k \right] = \sum_{j=0}^d \gamma_j t^j$$

$$\Leftrightarrow \sum_{i=-p}^{+f} \theta_i = 1 \text{ et } \sum_{i=-p}^{+f} \theta_i i^k = 0 \ \, \forall k=1,\dots,d \, .$$

Sous forme matricielle ceci s'écrit:

$$C\theta = b$$
,

$$\operatorname{avec} C \coloneqq \begin{pmatrix} 1 & 1 & \dots & 1 & 1 \\ -p & -p+1 & \dots & f-1 & f \\ (-p)^2 & (-p+1)^2 & \dots & (f-1)^2 & f^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (-p)^{d-1} & (-p+1)^{d-1} & \dots & (f-1)^{d-1} & f^{d-1} \\ (-p)^d & (-p+1)^d & \dots & (f-1)^d & f^d \end{pmatrix}, \ \theta = \begin{pmatrix} \theta_{-p} \\ \theta_{-p+1} \\ \theta_{-p+2} \\ \vdots \\ \theta_{f-1} \\ \theta_f \end{pmatrix} \text{ et } b = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}.$$

En outre, toute moyenne mobile qui conserve les constantes (respectivement un polynôme de degré 2d ) conserve les droites (respectivement un polynôme de degré 2d+1).

#### Elimination de la saisonnalité

Si la saisonnalité est mensuelle (donc de période 12), il faut éliminer les fréquences multiples de  $\frac{\pi}{6} rad = 30^\circ$  .

Si la saisonnalité est trimestrielle (donc de période 4), il faut éliminer les fréquences multiples de  $\frac{\pi}{2}rad=90^\circ$  .

Pour ce faire, la fonction de gain est utilisée, car celle-ci doit être nulle pour les fréquences éliminées.

#### Réduction du bruit

La composante irrégulière ne peut être éliminée, vu son aspect imprévisible et aléatoire, mais elle peut cependant être atténuée.

Elle est modélisée par un bruit blanc  $(\varepsilon_t)_{t\in\tau}$ , avec  $E[\varepsilon_t] = 0$  et  $Var(\varepsilon_t) = \sigma^2$ .

 $\text{La moyenne mobile } M\big[\varepsilon_{\scriptscriptstyle t}\big] = \sum_{\scriptscriptstyle i=-p}^{\scriptscriptstyle +f} \theta_{\scriptscriptstyle i} \varepsilon_{\scriptscriptstyle t+i} \quad \text{transforme ce bruit blanc en un autre bruit blanc } \Big(\varepsilon_{\scriptscriptstyle t}^*\Big)_{\!\scriptscriptstyle t\in\tau}\,,$ 

$$\begin{split} E\left[\boldsymbol{\varepsilon}_{t}^{*}\right] &= E\left[M\left[\boldsymbol{\varepsilon}_{t}\right]\right] = E\left[\sum_{i=-p}^{i=+f} \theta_{i} \boldsymbol{\varepsilon}_{t+i}\right] = \sum_{i=-p}^{+f} \theta_{i} E\left[\boldsymbol{\varepsilon}_{t+i}\right] = 0 \end{split} \qquad \text{et} \\ Var\left(\boldsymbol{\varepsilon}_{t}^{*}\right) &= Var\left(M\left[\boldsymbol{\varepsilon}_{t}\right]\right) = Var\left(\sum_{t=-p}^{+f} \theta_{i} \boldsymbol{\varepsilon}_{t+i}\right) = \sum_{i=-p}^{+f} \theta_{i}^{2} Var\left(\boldsymbol{\varepsilon}_{t+i}\right) = \sigma^{2} \sum_{i=-p}^{+f} \theta_{i}^{2} \; . \end{split}$$

Réduire le bruit revient donc à minimiser le terme  $\sum_{i=-n}^{+f} \theta_i^2$  .

Pour construire la moyenne mobile cherchée, il faut réunir toutes les conditions trouvées ci-dessus:

- La moyenne mobile doit être symétrique, donc par exemple d'ordre 2p+1 , allant de -p à +p .
- Elle doit satisfaire à la condition  $\sum_{i=-p}^{+p} \theta_i = 1$ , c'est-à-dire elle doit conserver les droites.
- Elle doit minimiser le terme  $\sum_{i=-p}^{+f} \theta_i^2$  .

En résolvant simultanément ces trois hypothèses, la condition unique suivante peut être obtenue:

$$\theta_i = \frac{1}{2p+1} \quad \forall i = -p, \dots, +p$$
.

#### Moyennes mobiles composées

Le logiciel X 12-ARIMA utilise des moyennes mobiles composées pour estimer les coefficients saisonniers. Ces moyennes sont notées 3x3, 3x5, 3x9, 3x15, où ixj signifie qu'il faut d'abord construire une moyenne mobile d'ordre i, puis une moyenne mobile d'ordre j. Afin d'illustrer ceci, la moyenne mobile composée 3x3 est considérée. D'abord la moyenne mobile d'ordre 3 est calculée en prenant par exemple -p=-1 et +f=+1:

$$M[X_t] = \frac{1}{3}X_{t-1} + \frac{1}{3}X_t + \frac{1}{3}X_{t+1}.$$

Ensuite, la moyenne mobile d'ordre 3 est de nouveau calculée pour chaque terme en prenant de nouveau -p=-1 et +f=+1:

$$M_{3\times3}\big[\boldsymbol{X}_{t}\,\big] = \frac{1}{3}\bigg(\frac{1}{3}\boldsymbol{X}_{t-2} + \frac{1}{3}\boldsymbol{X}_{t-1} + \frac{1}{3}\boldsymbol{X}_{t}\bigg) + \frac{1}{3}\bigg(\frac{1}{3}\boldsymbol{X}_{t-1} + \frac{1}{3}\boldsymbol{X}_{t} + \frac{1}{3}\boldsymbol{X}_{t+1}\bigg) + \frac{1}{3}\bigg(\frac{1}{3}\boldsymbol{X}_{t} + \frac{1}{3}\boldsymbol{X}_{t+1} + \frac{1}{3}\boldsymbol{X}_{t+2}\bigg)$$

$$= \frac{1}{9}X_{t-2} + \frac{2}{9}X_{t-1} + \frac{3}{9}X_{t} + \frac{2}{9}X_{t+1} + \frac{1}{9}X_{t+2}.$$

La moyenne mobile composée 3x3 possède donc les coefficients  $\frac{1}{9}\{1,2,3,2,1\}$ . En fait, la série est ainsi lissée deux fois.

Une bonne application de la mise en œuvre des moyennes mobiles composées est donnée par le principe itératif de X 11 (version antérieure de X 12-ARIMA).

#### Algorithme de base de X 11:

Pour obtenir une première estimation de la série corrigée des variations saisonnières  $X \_CVS_t^{(1)}$ , X 11 parcourt les étapes suivantes:

1. Estimation de la composante tendance-cycle par moyenne mobile composée 2x12:

$$(TC)_t^{(1)} = M_{2 \times 12} [X_t].$$

2. Estimation de la composante saisonnière-irrégulière (dans le cas d'un modèle additif  $X_t = (TC)_t + S_t + I_t$ ):

$$(S_t + I_t)^{(1)} = X_t - (TC)_t^{(1)}$$
.

3. Estimation de la composante saisonnière par moyenne mobile composée 3x3 sur chaque mois:

$$S_{t}^{(1)}=M_{3\times 3}\Big[\!\big(S_{t}+I_{t}\big)^{\!(1)}\,\Big]\!,$$
 afin d'éliminer la composante irrégulière et

 $S\_norm_t^{(1)} = S_t^{(1)} - M_{2 \times 12} [S_t^{(1)}]$ , afin que la somme des coefficients sur toute période de 12 mois soit approximativement nulle (normalisation).

4. Estimation de la série corrigée des variations saisonnières:

$$X \_CVS_t^{(1)} = [(TC)_t + I_t]^{(1)} = X_t - S \_norm_t^{(1)}.$$

Cette première estimation de la série corrigée des variations saisonnières  $X \_CVS_t^{(1)}$  doit, par construction contenir une saisonnalité inférieure.

Pour obtenir une deuxième estimation de la série corrigée des variations saisonnières  $X \_CVS_t^{(2)}$ , X 11 parcourt les mêmes étapes, mais en prenant d'autres moyennes mobiles:

5. Estimation de la composante tendance-cycle par moyenne mobile de Henderson sur 13 termes à coefficients  $\frac{1}{16796}$  {-325,-468,0,1100,2475,3600,4032,3600,2475,1100,0,-468,-325}:

$$(TC)_{t}^{(2)} = H_{13} [X \_CVS_{t}^{(1)}].$$

6. Estimation de la composante saisonnière-irrégulière (dans le cas d'un modèle additif):

$$(S_t + I_t)^{(2)} = X_t - (TC)_t^{(2)}$$
.

7. Estimation de la composante saisonnière par moyenne mobile composée 3x5 sur chaque mois:

$$S_t^{(2)} = M_{3 \times 5} \Big[ \! \big( S_t + I_t \big)^{(2)} \, \Big]$$
 , afin d'éliminer la composante irrégulière et

 $S\_norm_t^{(2)} = S_t^{(2)} - M_{2 \times 12}[S_t^{(2)}]$ , afin que la somme des coefficients sur toute période de 12 mois soit approximativement nulle (normalisation).

8. Estimation de la série corrigée des variations saisonnières:

$$X \_CVS_t^{(2)} = [(TC)_t + I_t]^{(2)} = X_t - S \_norm_t^{(2)}.$$

Le programme X 12-ARIMA utilise X 11 pour la désaisonnalisation et une modélisation ARIMA pour stabiliser les estimations de fin de série en évitant le recours à des moyennes mobiles asymétriques.

1.2.3 Méthodes ayant recours à des modèles économétriques:

Etant donné que la méthode de prévision de Box-Jenkins, développée dans la suite, ne peut être appliquée qu'à des séries stationnaires, un rappel de la définition de celles-ci est donné:

#### Stationnarité

Une série chronologique est stationnaire si elle ne comporte pas de tendance, ni de saisonnalité, c.-à-.d si elle est homogène par rapport au temps.

#### Définition: Stationnarité

Une série temporelle  $(X_t)_{t=\tau}$  est dite <u>faiblement stationnaire</u> ou tout court <u>stationnaire</u>, si sa moyenne, sa variance et sa covariance sont indépendantes du temps et si sa variance est finie, c'est-à-dire si

- $E[X_t] =: \mu = cst. \ \forall t \in \tau$  processus générateur sans tendance,
- $Var(X_t) =: \sigma^2 = cst. < \infty \quad \forall t \in \tau$  variance indépendante de t et bornée et
- $Cov(X_s, X_t) = E[X_s X_t] \mu^2 =: \gamma_l$ , avec l := t s fonction auto-covariance ne dépend que de l'écart l entre s et t et non pas de s ni de t.

Vu la nécessité de séries stationnaires pour pouvoir appliquer la méthode de prévision de Box-Jenkins, il convient de s'interroger sur les méthodes, à l'aide desquelles une série peut être rendue stationnaire.

Soit  $X_t := \alpha t + \beta + \varepsilon_t$  une série, où  $\varepsilon_t$  est un bruit blanc, avec  $E[\varepsilon_t] = 0$  et  $Var(\varepsilon_t) = \sigma^2$ .

Il est évident que cette série n'est pas stationnaire, car  $E[X_t] = \alpha t + \beta \neq cst$ .

Soit d un nombre entier et  $abla^d$  l'opérateur différence d ème, abla étant l'opérateur différence saisonnière. Donc

$$\nabla^{d} X_{t} = X_{t} - B^{d} X_{t} = X_{t} - X_{t-d}.$$

Toutefois la série  $Y_t \coloneqq \nabla X_t = X_t - X_{t-1} = \alpha + \varepsilon_t - \varepsilon_{t-1}$  est stationnaire, car

$$E[Y_t] = \alpha = cst. \ \forall t \in \tau \text{ et}$$

$$Var(Y_t) = 2\sigma^2 = cst. < \infty \ \forall t \in \tau$$
 et

$$Cov(Y_s, Y_t) = E[Y_s Y_t] - \alpha^2 = \begin{cases} -\sigma^2 & t - s = -1 \\ \sigma^2 & t - s = 1 \\ 0 & \forall t - s \notin \{-1, 1\} \end{cases}$$

Par application de l'opérateur différence première, une série stationnaire a donc pu être construite. Un essai pour rendre une série stationnaire peut être effectué, en appliquant un opérateur différence  $\nabla^d$ . Souvent, en présence de phénomènes saisonniers, une série devient stationnaire par différenciation sur la saison.

#### Méthode de prévision de Box-Jenkins

Afin de modéliser une série temporelle donnée, il peut être fait recours à un modèle auto-projectif  $X_t = f(X_{t-1}, X_{t-2}, \dots, \mathcal{E}_t)$ , pour lequel la série ne dépend que de son passé et d'un terme d'erreur. Il s'agit donc de trouver la fonction f. Si la série  $(X_t)_{t=\tau}$  est stationnaire, il existe un modèle ARMA («<u>A</u>uto-<u>Regressive Moving A</u>verage») qui l'approche au mieux.

<u>Définition: Modèle</u> AR(p)

Un  $\underline{\text{modèle auto-régressif}}$  (« $\underline{\text{A}}$ uto- $\underline{\text{R}}$ egressive  $\underline{\text{M}}$ odel») d'ordre p est noté AR(p) et est défini par:

$$X_{\scriptscriptstyle t} := \varphi_{\scriptscriptstyle 1} X_{\scriptscriptstyle t-1} + \varphi_{\scriptscriptstyle 2} X_{\scriptscriptstyle t-2} + \ldots + \varphi_{\scriptscriptstyle p} X_{\scriptscriptstyle t-p} + \varepsilon_{\scriptscriptstyle t},$$

ce qui revient à

$$(1 - \varphi_1 B - \varphi_2 B^2 - \dots \varphi_p B^p) X_t =: \Phi(B) X_t = \varepsilon_t,$$

où  $\left(\mathcal{E}_{t}\right)_{t\in\tau}$  est un bruit blanc.

Exemple: Le modèle AR(1) est donné par:  $X_t = \varphi_1 X_{t-1} + \varepsilon_t$  .

<u>Définition: Modèle</u>  $\mathit{MA}(q)$ 

Un  $\underline{\text{modèle moyenne mobile}}$  (« $\underline{\text{M}}$ oving  $\underline{\text{A}}$ verage  $\underline{\text{M}}$ odel») d'ordre q est noté MA(q) et est défini par:

$$X_t := \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_a \varepsilon_{t-a}$$

ce qui revient à

$$X_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t =: \Theta(B) \varepsilon_t$$

où  $\left(\mathcal{E}_{t}\right)_{t\in\tau}$  est un bruit blanc.

Exemple: Le modèle  $M\!A(1)$  est donné par:  $X_{t} = \mathcal{E}_{t} - \theta_{1}\mathcal{E}_{t-1}$  .

Définition: Modèle ARMA(p,q)

Un <u>modèle moyenne mobile</u> (d'ordre q) <u>auto-régressive</u> (d'ordre p) («<u>A</u>uto-<u>R</u>egressive <u>M</u>oving <u>A</u>verage <u>M</u>odel») est noté ARMA(p,q) et est défini par:

$$\Phi(B)X_t := \Theta(B)\varepsilon_t,$$

où  $\left(\mathcal{E}_{t}\right)_{t\in\tau}$  est un bruit blanc.

Exemple: Le modèle  $\mathit{ARMA}(1,1)$  est donné par:  $X_t - \varphi_1 X_{t-1} = \varepsilon_t - \theta_1 \varepsilon_{t-1}$  .

Définition: Modèle  $\mathit{ARIMA}(p,d,q)$ 

Un <u>modèle moyenne mobile</u> (d'ordre q) <u>intégrée</u> (d'ordre d) <u>auto-régressive</u> (d'ordre p) («<u>A</u>uto-Regressive Integrated Moving Average Model») est noté ARIMA(p,d,q) et est défini par:

$$(1-B)^d \Phi(B) X_t := \Theta(B) \varepsilon_t,$$

où  $\left(\mathcal{E}_{t}\right)_{t\in\tau}$  est un bruit blanc.

Ce type de modèle est utilisé pour les séries non stationnaires, qui présentent une tendance, mais pas de saisonnalité.

Exemple: Le modèle ARIMA(1,1,1) est donné par:  $X_t - (1+\varphi_1)X_{t-1} + \varphi_1X_{t-2} = \varepsilon_t - \theta_1\varepsilon_{t-1}$ .

<u>Définition:Modèle</u>  $SARIMA(p,d,q)(P,D,Q)_s$ 

Un modèle de Box-Jenkins multiplicatif saisonnier, de moyenne mobile (d'ordre q) intégrée (d'ordre

d ) <u>auto-régressive</u> (d'ordre p ) («Box-Jenkins Multiplicative <u>Seasonal Auto-Regressive Integrated Moving Average Model») est noté  $SARIMA(p,d,q)(P,D,Q)_S$  et est défini par:</u>

$$(1-B)^d (1-B^s)^D \Phi(B)\Phi(B^s) X_t := \Theta(B)\Theta(B^s) \varepsilon_t$$
, où  $(\varepsilon_t)_{t\in \tau}$  est un bruit blanc.

Ce type de modèle est utilisé pour les séries non stationnaires, qui présentent une tendance et une saisonnalité.

Un exemple très fréquent de modèle SARIMA est le «Airline Model», qui correspond au modèle  $SARIMA(0,1,1)(0,1,1)_S$ :  $(1-B)(1-B^S)X_t = (1-\theta_1B)(1-\theta_1B^S)\varepsilon_t, \qquad \text{c.-à-d.}$   $X_t - X_{t-1} - X_{t-s} + X_{t-s-1} = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_1\varepsilon_{t-s} + \theta_1\theta_1\varepsilon_{t-s-1}.$ 

#### Méthode de prévision de Box-Jenkins

La méthode de prévision de Box-Jenkins («grid-search procedure») parcourt les étapes suivantes:

- 1. Stationnarisation du processus (choix des ordres de différence d et D).
- 2. Identification a priori du modèle (choix des ordres p, P, q et Q).
- 3. Estimation des coefficients  $\varphi_i$ ,  $\theta_j$ ,  $(\varphi_S)_i$  et  $(\theta_S)_j$ .
- 4. Validation du modèle.
- 5. Choix d'un modèle (si plusieurs modèles validés).
- Prévision.

Ceci est réalisé par application des indications suivantes:

- ad 1. Transformer la série par différenciation.
- ad 2. Par l'intermédiaire des autocorrélations simples et partielles. Il s'agit d'un modèle AR(p), si les autocorrélations simples décroissent exponentiellement et les autocorrélations partielles s'annulent à partir du rang p+1. Il s'agit d'un modèle MA(q), si les autocorrélations simples s'annulent à partir du rang q+1 et les autocorrélations partielles décroissent exponentiellement.
- ad 3. Par l'estimation du maximum de vraisemblance («maximum likelihood estimator») ou la méthode des moindres carrés («least squares method»).
- ad 4. A l'aide de tests sur la nullité des coefficients (par exemple test de Student) et de tests de bruit blanc sur les résidus (par exemple test de Portmanteau connu aussi sous le nom de test de Ljung-Box).
- ad 5. A l'aide de critères (par exemple critère d'Akaike).

Une bonne application de la mise en œuvre des méthodes ayant recours à des modèles économétriques composés est donnée par le programme TRAMO.

TRAMO-SEATS emploie des modèles ARIMA pour la modélisation de la série temporelle. En fait, les programmes TRAMO («Time Series Regression with ARIMA Noise, Missing Values, and Outliers») et SEATS («Signal Extraction in ARIMA Time Series») sont deux programmes indépendants, même s'ils

sont le plus souvent utilisés ensemble. TRAMO est utilisé pour la modélisation ARIMA et SEATS pour la désaisonnalisation. SEATS repose sur une décomposition du modèle obtenu par TRAMO en une somme de composantes et sur une extraction de ces composantes. Ces composantes seront décrites par des processus ARIMA qui seront choisis pour que la somme ait les mêmes autocovariances que le processus obtenu par la méthode TRAMO. Les programmes TRAMO et SEATS ont été développés par Víctor Gómez et Augustín Maravall sur la base d'un programme antérieur de J. Peter Burman (1980). Le programme TSW est une version Windows, qui intègre les deux programmes TRAMO et SEATS.

Dans le programme TRAMO, p et q varient de 0 à 3 et les modèles équilibrés, c'est-à-dire ceux où p=q, sont privilégiés. Les non-linéarités (changements de niveau, points aberrants, changements de niveau transitoire, rampes) ainsi que les variables externes (jours ouvrables, effet de Pâques) sont pris en compte.

Vu que le Statec utilise aussi la combinaison de programmes TRAMO-SEATS, les deux programmes sont décrits ici, alors que seule la partie TRAMO utilise la modélisation ARIMA décrite ci-dessus.

#### Algorithme de base de TRAMO-SEATS:

- D'abord TRAMO détecte, estime et corrige automatiquement les valeurs manquantes, les valeurs aberrantes, les changements de régime et les effets de calendrier.
- TRAMO propose automatiquement un modèle ARIMA (utilisation de la méthode de l'estimation du maximum de vraisemblance («maximum likelihood estimation»)).
- TRAMO transmet alors les séries préalablement corrigées à SEATS qui effectue la décomposition, c'est-à-dire estime la tendance-cycle à long terme, la composante saisonnière et la composante irrégulière.
- 4. En premier lieu, SEATS décompose la fonction de densité spectrale du modèle estimé dans les fonctions de densité spectrale des composantes non observées.
- Ensuite SEATS évalue les paramètres de la tendance-cycle à long terme et de la composante saisonnière.
- 6. En fin de compte, l'effet des valeurs aberrantes est réintroduit dans les composantes.

TRAMO peut également estimer les effets de calendrier et les observations manquantes.

Par utilisation de modèles à intervention Reg-ARIMA, TRAMO peut corriger quatre types de valeurs aberrantes:

- point aberrant («additive outlier»)
- changement de niveau transitoire («transitory change»)
- changement de niveau («level shift»)
- rampe («ramp»)



Figure 1.2.3.1: point aberrant («additive outlier»)

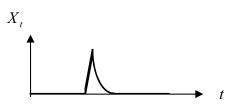


Figure 1.2.3.2: changement de niveau transitoire («transitory change»)

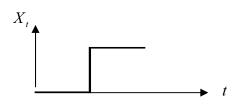


Figure 1.2.3.3: changement de niveau («level shift»)

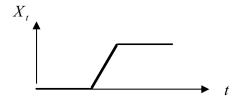


Figure 1.2.3.4: rampe («ramp»)

SEATS met à disposition des outils analytiques, qui permettent de déterminer les erreurs dans les estimateurs des composantes et les erreurs de prévision. Afin de contrôler si le modèle ARIMA choisi est adéquat, les tests de Ljung-Box et de Box-Pierce, implémentés dans SEATS, peuvent être utilisés. SEATS possède également d'importants outils graphiques.

Un dernier point méthodologique important concerne les effets de calendrier, dont traitera le paragraphe suivant:

#### 1.3 Les effets de calendrier

Il existe différents effets de calendrier:

les effets des jours ouvrables

Dans le contexte de la productivité, le terme jours ouvrables («working days») désigne les jours où les entreprises ont travaillé, donc les lundis, mardis, mercredis, jeudis et vendredis, non fériés d'un mois. Vu que le nombre de jours ouvrables varie d'un mois à l'autre, la production en dépend et cet effet doit donc être corrigé.

les effets concernant le type de jour de la semaine

Pour le commerce de détail, les ventes sont plus importantes les samedis que les autres jours de la semaine. Il faudra donc également lisser cet effet.

• les effets des jours fériés et des fêtes mobiles

Un exemple important de ces effets constitue l'effet de Pâques. Pâques est une fête mobile qui a lieu

soit en mars, soit en avril (entre le 22 mars et le 25 avril), donc soit au premier, soit au deuxième trimestre. Ainsi l'activité économique de nombreux secteurs augmente et diminue à différentes dates d'une année à l'autre. Une répartition de cette activité sur les deux mois concernés ou même les deux trimestres concernés est donc nécessaire. D'autres exemples sont le Ramadan, le Nouvel An chinois et Pâques pour les orthodoxes.

Une partie des effets de calendrier est déjà prise en compte dans la composante saisonnière. Ainsi le nombre de jours ouvrables du mois de février est presque chaque année inférieur à celui des autres mois. Certains mois, comme par exemple le mois de mai, comportent chaque année plus de jours fériés que les autres mois. L'Ascension tombe toujours sur un jeudi.

Méthode la plus simple pour corriger les effets des jours ouvrables:

Il est supposé que l'activité économique soit identique chaque jour ouvrable du mois (donc les effets concernant le type de jour de la semaine sont négligés). Comme le mois type a 21 jours ouvrables, l'application de la règle de trois pour un jour ouvrable supplémentaire donne une augmentation de la production de 4,762%. Ceci s'écrit donc:

$$X_t^{cjo} = X_t \frac{n_{jo}}{21} \,,$$

où  $n_{jo}$  constitue le nombre de jours ouvrables effectifs du mois concerné.

Méthode économétrique pour corriger les effets des jours ouvrables, les effets concernant le type de jour de la semaine et les effets des jours fériés:

Il est supposé que l'effet d'un jour ouvrable ou d'un certain type de jour de la semaine soit constant sur la période examinée. Alors le modèle de base suivant peut être construit:

$$X_t = \sum_{i=1}^{7} \alpha_i N_{i,t} + \varepsilon_t ,$$

où  $\alpha_i$  représente l'effet d'un jour de type i (par exemple production moyenne),

 $N_{i,t}$  représente le nombre de lundis (pour i=1), de mardis (pour i=2), de mercredis (pour i=3), de jeudis (pour i=4), de vendredis (pour i=5), de samedis (pour i=6) et de dimanches (pour i=7) du mois t de l'année considérée et

 $\mathcal{E}_{_{t}}$  est un bruit blanc.

Un modèle plus performant est obtenu par distinction des jours fériés et non fériés:

$$\boldsymbol{X}_{t} = \sum_{i=1}^{7} \boldsymbol{\beta}_{i} \boldsymbol{N}_{i,t}^{\textit{nf}} + \sum_{i=1}^{6} \boldsymbol{\gamma}_{i} \boldsymbol{N}_{i,t}^{\textit{f}} + \boldsymbol{\varepsilon}_{t} \,,$$

où  $\beta_i$  représente l'effet d'un jour non férié de type i et  $\gamma_i$  l'effet d'un jour férié de type i ,

 $N_{i,t}^{\it nf}$  représente le nombre de lundis non fériés (pour i=1),..., de dimanches non fériés (pour

i=7 ) du mois t de l'année considérée et  $N_{i,t}^f$  représente le nombre de lundis fériés (pour i=1 ),..., de dimanches fériés (pour i=7 ) du mois t de l'année considérée et

 $\mathcal{E}_t$  est un bruit blanc.

La deuxième  $\Sigma$  s'arrête à la valeur i=6, car les dimanches fériés sont considérés comme des dimanches normaux (vu qu'il n'y a pas ou pratiquement pas de production les dimanches).

Les effets des fêtes mobiles sont traités par des modèles Reg-ARIMA.

Pour des explications plus détaillées concernant la méthodologie, une référence utile est la publication [6].

# 2. Les décisions méthodologiques prises par le Statec

Il importe de savoir que chaque institut de statistique en Europe se voit obligé de fournir les séries brutes ainsi que les séries désaisonnalisées à EUROSTAT. La méthodologie utilisée par le Statec dans ce domaine se base notamment sur les ouvrages suivants: le «Système européen des comptes SEC95», publié par EUROSTAT, le «Handbook on quarterly national accounts», publié également par EUROSTAT, le «Quarterly National Accounts Manuel», publié par le Fond Monétaire International, les documents distribués lors d'un workshop spécial sur la désaisonnalisation, organisé par EUROSTAT, ainsi que les publications reçues de la part de Monsieur Augustín Maravall, développeur du programme TRAMO-SEATS. Après plusieurs réflexions, discussions et tests pratiques sur les données disponibles, le Statec a pris les décisions méthodologiques suivantes:

- Niveau de détail: Le niveau de détail, à partir duquel la désaisonnalisation est appliquée, est le niveau A17 (ou plutôt le niveau A15 qui est un regroupement du niveau A17 utilisé par l'unité C2 du Statec). Vu qu'il s'agit de séries chaînées et que celles-ci sont par définition non additives, les séries chaînées désaisonnalisées au niveau de détail A15 sont alors déchaînées, par simple inversion de la formule du chaînage. Les séries déchaînées, ainsi obtenues, sont ensuite agrégées au niveau A6 et rechaînées. Les résultats calculés par cette méthode sont nettement mieux, car plus précis, que ceux obtenus en désaisonnalisant qu'à partir du niveau A6. La désaisonnalisation entraîne, elle-aussi, une non-additivité des branches désaisonnalisées. Cependant, il convient de noter que l'ampleur des distorsions dues à cette non-additivité est nettement moins grande que celle due au chaînage, de sorte que la non-additivité due à la désaisonnalisation peut être négligée.
- Calcul du PIB: Le calcul du PIB et de la VAB totale se fait de façon directe, c'est-à-dire que le PIB et la VAB totale sont désaisonnalisés directement au lieu de désaisonnaliser leurs éléments et d'agréger par après. Ainsi l'écart de non-additivité dû à la désaisonnalisation (bien que mineur) ne brouille aucunement la tendance globale du PIB.
- Contributions à la croissance: Vu les problèmes de non-additivité dus au chaînage et à la désaisonnalisation, ainsi que l'absence générale de solutions raisonnables de la part d'EUROSTAT ainsi que de toute autre institution, le Statec s'est décidé à ne pas publier de contributions à la croissance, jusqu'à une date ultérieure.

Finalement, il convient encore de noter qu'il n'a pas lieu de négliger l'utilité des séries brutes et ceci pour les raisons suivantes:

- Les séries brutes illustrent les faits économiques actuels qui ont eu lieu, alors que les séries désaisonnalisées et la tendance représentent une élaboration analytique, destinée à révéler les mouvements économiques sous-jacents qui pourraient être masqués par la saisonnalité des données brutes.
- Les méthodes employées pour l'ajustement saisonnier ne donnent pas une solution unique.
- Les séries désaisonnalisées sont susceptibles d'être soumises à des révisions avec l'apparition de nouvelles données, même si les séries originales n'ont pas été révisées.
- Dans le domaine des comptes nationaux trimestriels, il vaut mieux réaliser l'équilibrage sur des données, qui ne sont pas désaisonnalisées.
- Alors que des erreurs dans les données de base sont plus facilement détectées à l'aide des séries désaisonnalisées, il semble plus facile d'identifier la source des erreurs et de corriger ces erreurs en travaillant sur les séries brutes.
- La pratique a démontré qu'en désaisonnalisant les données brutes, au niveau de détail nécessaire pour la compilation des comptes nationaux trimestriels, un résidu de saisonnalité peut demeurer dans les agrégats.

Ainsi, non seulement les données désaisonnalisées, mais aussi les données brutes ont leur raison d'être et fournissent des informations importantes (bien que différentes) sur l'économie.

Comme les utilisateurs sont plutôt intéressés à connaître la tendance globale de l'économie, le Statec s'est résolu à publier plutôt les séries désaisonnalisées que les séries brutes. Cette approche a aussi été adoptée par pratiquement tous les autres pays européens.

Pour conclure, certains tableaux et graphiques, issus de la dernière notification des comptes trimestriels, sont considérés. Ils permettent d'apprécier la différence entre les séries brutes et les séries désaisonnalisées. Les graphiques suivants ne concernent que les séries du PIB, étant donné que les séries de la VAB ont même allure que celles du PIB.

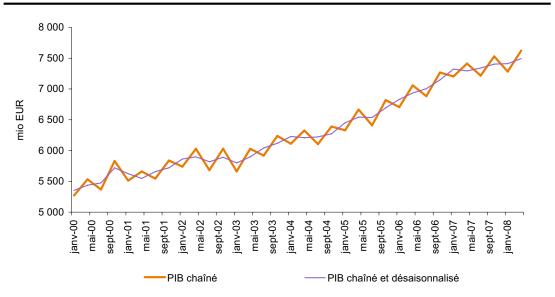
Tableau 2.1: PIB et VAB bruts vs. PIB et VAB désaisonnalisés (valeurs)

		PIB chaîné et		VAB chaînée et
_	PIB chaîné	désaisonnalisé	VAB chaînée	désaisonnalisée
janv-00	5 270	5 353	4 683	4 737
avr-00	5 532	5 437	4 923	4 812
juil-00	5 368	5 474	4 782	4 875
oct-00	5 830	5 718	5 235	5 185
janv-01	5 511	5 624	4 955	5 007
avr-01	5 660	5 548	5 095	4 989
juil-01	5 544	5 658	5 000	5 109
oct-01	5 838	5 720	5 223	5 161
janv-02	5 738	5 862	5 140	5 192
avr-02	6 029	5 898	5 399	5 286
juil-02	5 682	5 818	5 096	5 226
oct-02	6 030	5 892	5 437	5 363
janv-03	5 662	5 798	5 098	5 153
avr-03	6 028	5 898	5 422	5 304
juil-03	5 917	6 042	5 294	5 428
oct-03	6 238	6 117	5 564	5 488
janv-04	6 112	6 229	5 498	5 554
avr-04	6 326	6 209	5 637	5 523
juil-04	6 106	6 223	5 427	5 570
oct-04	6 390	6 272	5 693	5 617
janv-05	6 331	6 451	5 626	5 687
avr-05	6 665	6 541	5 954	5 840
juil-05	6 410	6 537	5 743	5 884
oct-05	6 821	6 694	6 129	6 040
janv-06	6 705	6 830	6 030	6 102
avr-06	7 058	6 934	6 350	6 232
juil-06	6 883	7 005	6 205	6 339
oct-06	7 269	7 149	6 508	6 414
janv-07	7 205	7 323	6 517	6 584
avr-07	7 413	7 293	6 660	6 550
juil-07	7 217	7 340	6 490	6 620
oct-07	7 531	7 405	6 759	6 664
janv-08	7 283	7 411	6 569	6 638
avr-08	7 625	7 496	6 868	6 756

Source: STATEC Aux erreurs d'arrondi près

juin 2009

Graphique 2.1: PIB brut vs. PIB désaisonnalisé



Source: STATEC

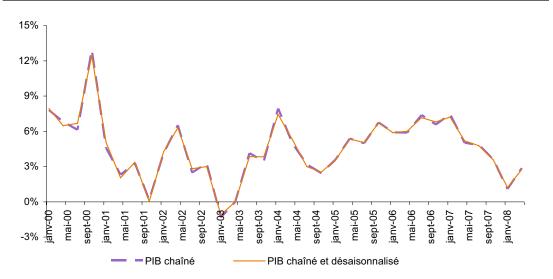
La variation d'un trimestre par rapport au même trimestre de l'année précédente peut être calculée à l'aide des séries brutes ou désaisonnalisées. Les résultats sont toutefois différents:

Tableau 2.2: PIB et VAB bruts vs. PIB et VAB désaisonnalisés (pourcentages de variation)

		PIB chaîné et		VAB chaînée et
	PIB chaîné	désaisonnalisé	VAB chaînée	désaisonnalisée
janv-00	7.8%	7.9%	5.6%	5.6%
avr-00	6.8%	6.5%	5.4%	5.4%
juil-00	6.1%	6.7%	5.6%	5.6%
oct-00	12.9%	12.4%	13.3%	12.8%
janv-01	4.6%	5.0%	5.8%	5.7%
avr-01	2.3%	2.0%	3.5%	3.7%
juil-01	3.3%	3.4%	4.6%	4.8%
oct-01	0.1%	0.0%	-0.2%	-0.5%
janv-02	4.1%	4.2%	3.7%	3.7%
avr-02	6.5%	6.3%	6.0%	6.0%
juil-02	2.5%	2.8%	1.9%	2.3%
oct-02	3.3%	3.0%	4.1%	3.9%
janv-03	-1.3%	-1.1%	-0.8%	-0.7%
avr-03	0.0%	0.0%	0.4%	0.4%
juil-03	4.1%	3.8%	3.9%	3.9%
oct-03	3.4%	3.8%	2.3%	2.3%
janv-04	7.9%	7.4%	7.8%	7.8%
avr-04	4.9%	5.3%	4.0%	4.1%
juil-04	3.2%	3.0%	2.5%	2.6%
oct-04	2.4%	2.5%	2.3%	2.4%
janv-05	3.6%	3.6%	2.3%	2.4%
avr-05	5.4%	5.3%	5.6%	5.7%
juil-05	5.0%	5.0%	5.8%	5.6%
oct-05	6.7%	6.7%	7.7%	7.5%
janv-06	5.9%	5.9%	7.2%	7.3%
avr-06	5.9%	6.0%	6.7%	6.7%
juil-06	7.4%	7.2%	8.0%	7.7%
oct-06	6.6%	6.8%	6.2%	6.2%
janv-07	7.5%	7.2%	8.1%	7.9%
avr-07	5.0%	5.2%	4.9%	5.1%
juil-07	4.8%	4.8%	4.6%	4.4%
oct-07	3.6%	3.6%	3.9%	3.9%
janv-08	1.1%	1.2%	0.8%	0.8%
avr-08	2.9%	2.8%	3.1%	3.2%

Source: STATEC Aux erreurs d'arrondi près

Graphique2.2: Pourcentage de variation du PIB par rapport au même trimestre de l'année précédente



Source: STATEC

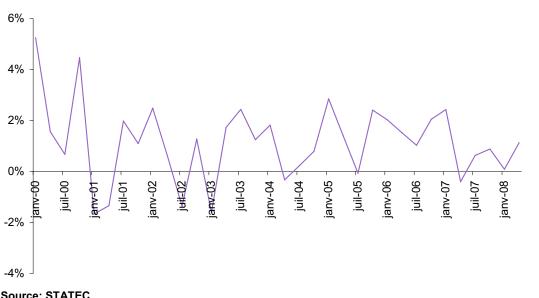
Cependant, les séries désaisonnalisées permettent également d'apprécier la variation d'un trimestre par rapport au trimestre précédent, ce qui est impossible avec les séries brutes:

Tableau 2.3: Pourcentage de variation du PIB et de la VAB par rapport au trimestre précédent

	Pourcentage de variation par rapport au trimestre précédent		
	PIB chaîné et désaisonnalisé	VAB chaînée et désaisonnalisée	
janv-00	5.3%	3.0%	
avr-00	1.6%	1.6%	
juil-00	0.7%	1.3%	
oct-00	4.5%	6.4%	
janv-01	-1.7%	-3.4%	
avr-01	-1.3%	-0.4%	
juil-01	2.0%	2.4%	
oct-01	1.1%	1.0%	
janv-02	2.5%	0.6%	
avr-02	0.6%	1.8%	
juil-02	-1.4%	-1.1%	
oct-02	1.3%	2.6%	
janv-03	-1.6%	-3.9%	
avr-03	1.7%	2.9%	
juil-03	2.4%	2.3%	
oct-03	1.3%	1.1%	
janv-04	1.8%	1.2%	
avr-04	-0.3%	-0.6%	
juil-04	0.2%	0.9%	
oct-04	0.8%	0.8%	
janv-05	2.8%	1.2%	
avr-05	1.4%	2.7%	
juil-05	-0.1%	0.8%	
oct-05	2.4%	2.7%	
janv-06	2.0%	1.0%	
avr-06	1.5%	2.1%	
juil-06	1.0%	1.7%	
oct-06	2.1%	1.2%	
janv-07	2.4%	2.7%	
avr-07	-0.4%	-0.5%	
juil-07	0.6%	-0.3 <i>%</i> 1.1%	
oct-07	0.0%	0.7%	
janv-08	0.9%	-0.4%	
avr-08	1.1%	-0.4% 1.8%	
avi-uo	1.1%	1.8%	

Source: STATEC Aux erreurs d'arrondi près juin 2009

Graphique 2.3: Pourcentage de variation du PIB par rapport au trimestre précédent



**Source: STATEC** 

Ces résultats, obtenus à l'aide des séries désaisonnalisées, sont désormais intégrés dans les publications du Statec, notamment dans le Statnews.

exemple concret de la désaisonnalisation à l'aide de l'interface Demetra.

exemple concret de la désaisonnalisation à l'aide de l'interface Demetra.

# 3. La désaisonnalisation à l'aide de l'interface Demetra

Afin de pouvoir reproduire les séries désaisonnalisées publiées par le Statec, l'utilisation de l'interface Demetra et en particulier le choix des paramètres sont décrits en détail dans ce chapitre.

Les fichiers «input» utilisés par le Statec sont toujours du Type Excel. Afin que Demetra reconnaisse les données des fichiers «input» du type Excel, ceux-ci doivent présenter une des structures suivantes: Dans la première cellule du fichier Excel, c.-à-d. la cellule A1 doit figurer un des mots suivants: «Horizontal», «Vertical», selon la disposition horizontale ou verticale des séries ou tout simplement le mot «Date», qui fonctionne dans les deux cas. Dans le cas de la disposition horizontale des séries, la première ligne, à part la première cellule de cette ligne, doit contenir les dates relatives aux données des séries. Dans le cas de la disposition verticale des séries, la première colonne, à part la première cellule de cette colonne, doit contenir les dates relatives aux données des séries. Demetra reconnaît alors automatiquement s'il s'agit de données mensuelles ou trimestrielles en prenant compte du laps de temps entre deux dates consécutives. A part la première ligne et la première colonne, les cellules ne doivent uniquement contenir les valeurs correspondantes aux séries.

Figure 3.1: Fichier «input» Excel en disposition horizontale

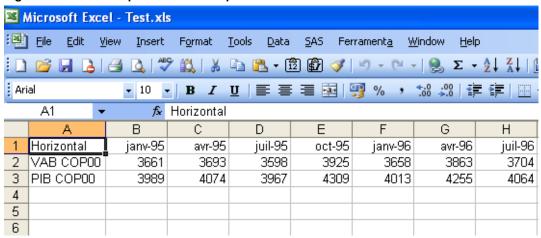
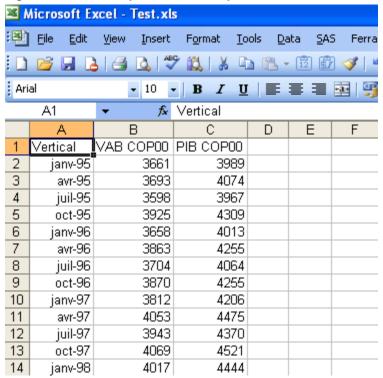


Figure 3.2: Fichier «input» Excel en disposition verticale



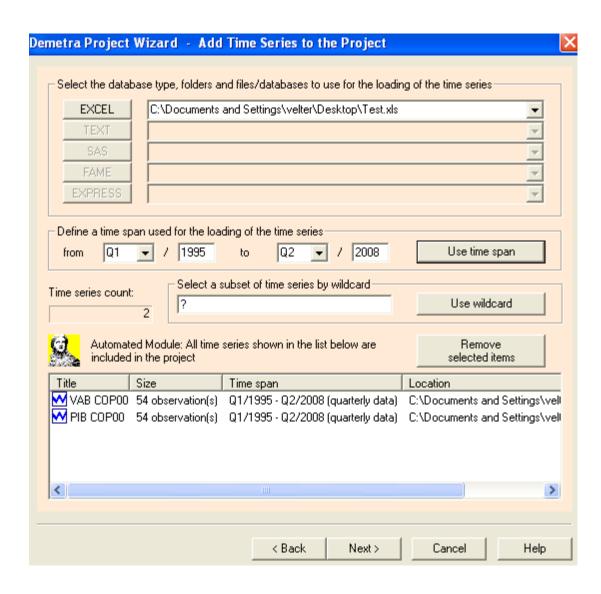
Dès que le fichier «input» présente la structure requise par Demetra, la désaisonnalisation à l'aide de cet interface peut débuter. **Attention!** Avant de lancer Demetra, le fichier «input» du type Excel doit être fermé! En premier lieu, il faut naturellement ouvrir l'interface Demetra. Dans la barre Menu, cliquez sur *Project*, puis sélectionner *New...*.



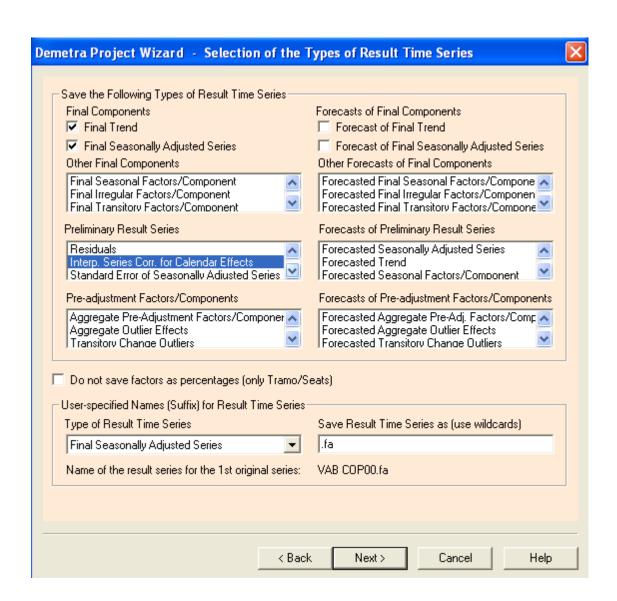
Dans la 1<sup>ère</sup> fenêtre qui apparaît alors, choisissez <sup>//</sup>Automated Module, puis cliquez sur le bouton Next>.



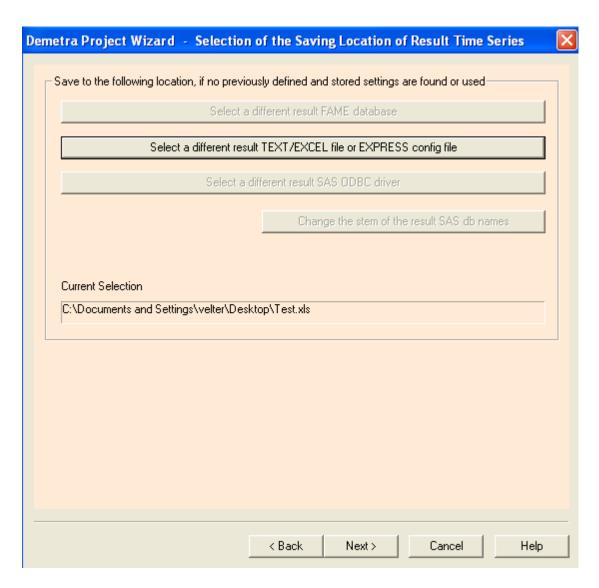
Dans la 2<sup>ième</sup> fenêtre, cliquez sur le bouton Excel puis, dans la fenêtre qui apparaît alors, ouvrez le fichier préparé auparavant. Pour la désaisonnalisation des comptes trimestriels, le Statec choisit, dans la partie *Define a time span used for the loading of the time series*, le début de la publication des comptes trimestriels jusqu'aux dernières données disponibles (p. ex. *Q1/1995* to *Qt/aaaa*, où *t* est le trimestre que l'on veut publier actuellement et *aaaa* l'année correspondant à ce trimestre). En appuyant sue le bouton Use time span, les séries chargées sont adaptés à ce choix. Cliquez sur le bouton Next>.



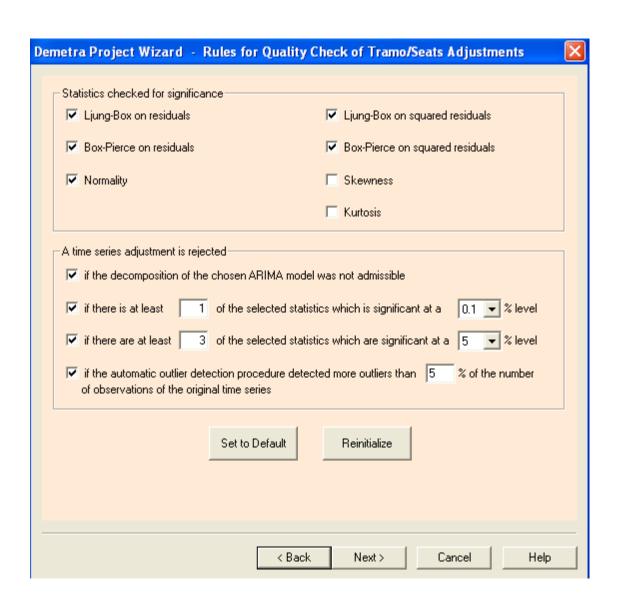
Dans la 3ième fenêtre, dans la partie *Preliminary Result Series*, sélectionnez *Interp. Series Corr. for Calendar Effects*, puis cliquez sur Next>. Cette sélection permet de produire le «working day adjustment», c.-à-d. l'ajustement pour les jours ouvrables seul. La sélection automatique du *Final Trend* et des *Final Seasonally Adjusted Series* permet de calculer la tendance finale ainsi que les séries désaisonnalisées.

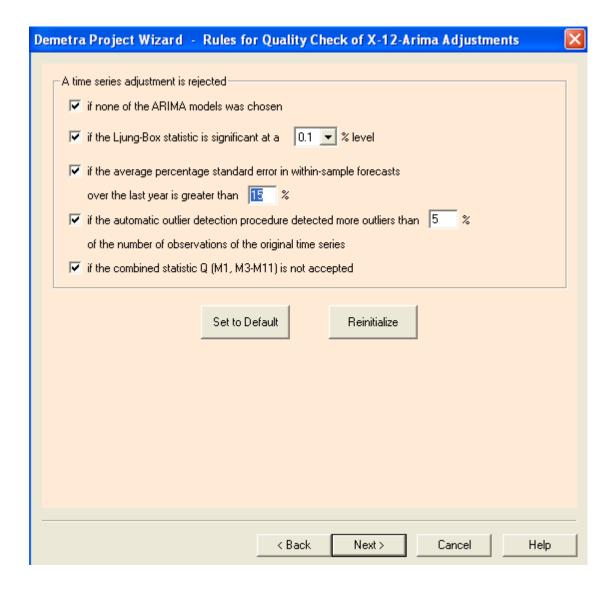


En appuyant sur Next> dans la fenêtre suivante, les séries ajustées des jours ouvrables, la tendance finale et les séries désaisonnalisées sont sauvegardés dans le même fichier Excel que les données source, mais sur d'autres feuilles. Cependant, il est également possible de les sauvegarder dans un autre fichier Excel, dont il faut alors définir l'emplacement. Ensuite, cliquez de nouveau sur Next>.

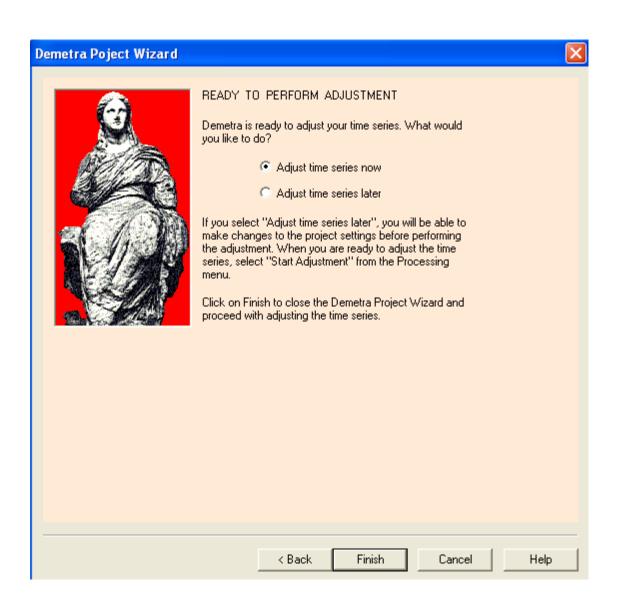


Dans les deux fenêtres suivantes, certains tests de signifiance ainsi que certaines conditions de rejet des séries ajustées peuvent être sélectionnés. Le Statec accepte à cet endroit les choix automatiques de Demetra. Il suffit donc de cliquer sur Next> dans ces deux fenêtres.

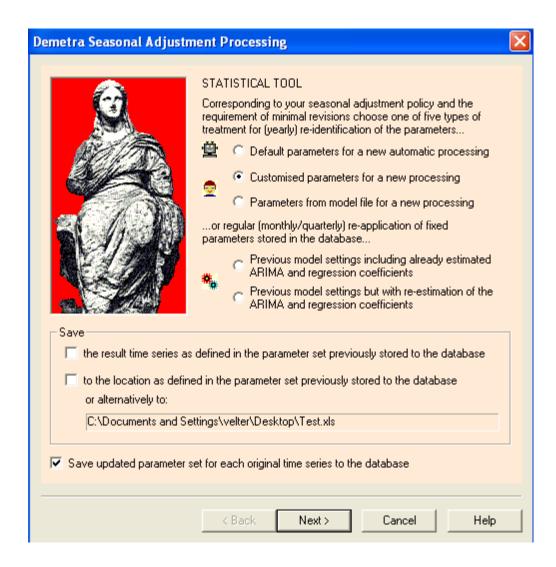




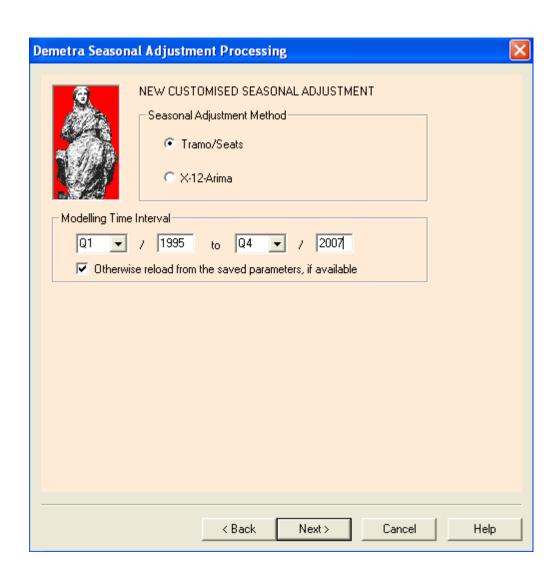
Pour débuter la désaisonnalisation, cliquez également sur le bouton Finish dans la fenêtre suivante. **Attention!** A partir de ce moment, il est impossible de retourner dans les fenêtres précédentes!



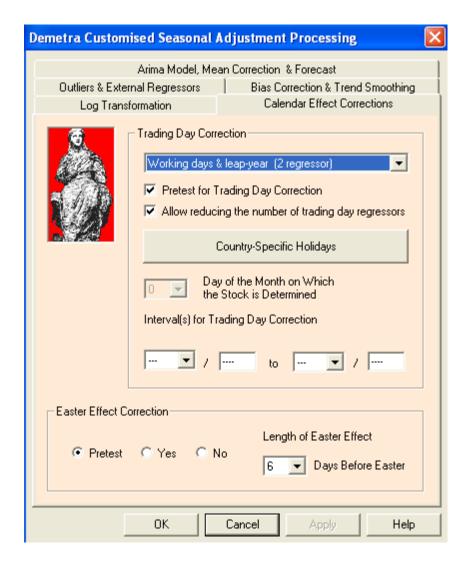
Dans la 8<sup>ième</sup> fenêtre choisissez <sup>M</sup>Customised parameters for a new processing, puis cliquez sur le bouton Next>.



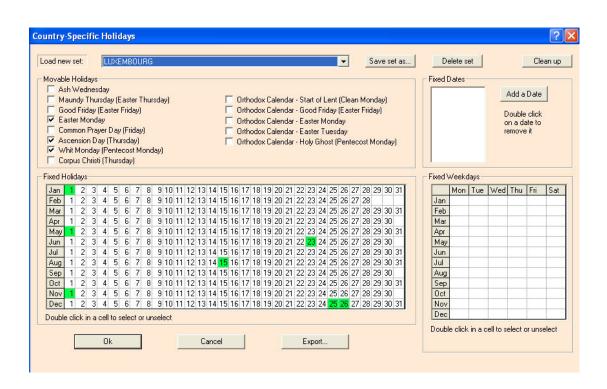
Dans la 9<sup>ième</sup> fenêtre, *TRAMO-SEATS* est sélectionné automatiquement. Dans la partie *Modelling Time Interval*, le Statec a décidé de choisir le début de la publication des comptes trimestriels jusqu'aux dernières données annuelles disponibles (p. ex. *Q1/1995* to *Q4/aaaa*, où *aaaa* est la dernière année pour laquelle des comptes annuels existent), afin de ne pas prendre en compte les données plus volatiles des trimestres, pour lesquels il n'existe pas encore de contrainte annuelle. Ensuite, cliquez sur le bouton Next>.



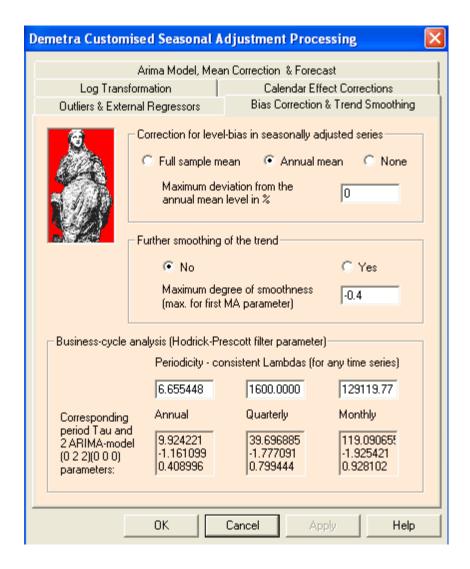
Dans la fenêtre suivante, activez la feuille *Calendar Effect Corrections*. Dans la partie *Trading Day Correction*, choisissez *Working days & leap-year (2 regressor)*, puis cliquez sur le bouton Country-Specific Holidays.



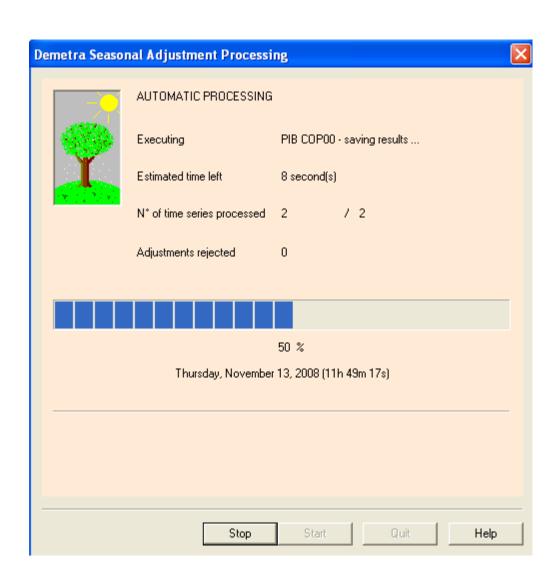
Dans la fenêtre qui s'ouvre alors, dans la partie *Load new set:* choisissez *Luxembourg* et cliquez sur le bouton OK.



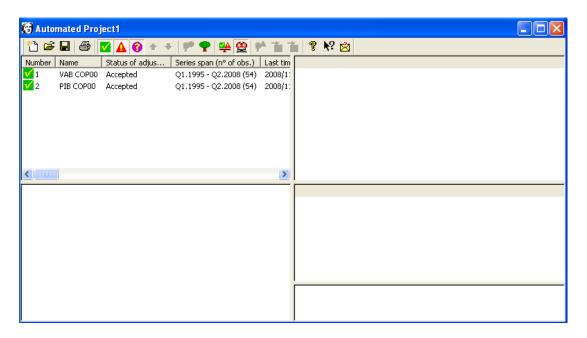
De retour dans la fenêtre précédente, activez la feuille *Bias Correction & Trend Smoothing*. Dans la partie *Correction for level-bias in seasonally adjusted series*, choisissez <sup>#</sup>Annual mean et mettez *Maximum deviation for the annual mean level in %* à 0. Ceci a pour effet que la somme des trimestres des séries désaisonnalisées correspond à la valeur annuelle. Quittez *Demetra Customised Seasonal Adjustment Processing* en cliquant sur le bouton OK.



Pour démarrer les calculs, cliquez sur le bouton Start dans a fenêtre suivante.

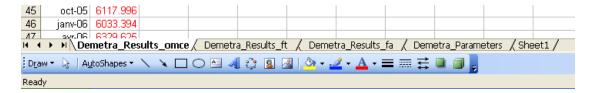


Dans la fenêtre des résultats, si une série est rejetée, détectez les raisons pour le rejet et remédiez-y si possible (p.ex. en supprimant les valeurs erronées dans la série originale, si le nombre de valeurs aberrantes est trop élevé). Si la série est rejetée, parce qu'elle n'est tout simplement pas saisonnière, prenez la série originale en tant que série ajustée des variations saisonnières.



Maintenant le fichier «input» du type Excel, contient de nouvelles feuilles relatives à la désaisonnalisation:

- la feuille Demetra\_Parameters, qui donne pour chaque série les données correspondantes à la désaisonnalisation, tel que p. ex. le modèle ARIMA choisi, le type de décomposition choisi, les résultats des tests de signifiance etc.
- la feuille Demetra\_Results\_fa, qui donne les séries désaisonnalisées (fa étant l'abréviation pour Final Seasonally Adjusted Series),
- la feuille *Demetra\_Results\_ft*, qui donne la tendance finale (*ft* étant l'abréviation pour *Final Trend*),
- le feuille Demetra\_Results\_omce, qui donne les séries ajustées des jours ouvrables (omce étant l'abréviation pour Interpolated Series Corrected for Calendar Effects)



Toutes les séries produites par Demetra sont coloriées en rouge, afin de faciliter la distinction entre les séries originales et les séries ajustées. Les séries de la feuille *Demetra\_Restults\_fa* sont celles utilisées pour la publication des comptes trimestriels. Cependant les séries chaînées désaisonnalisées au niveau de détail A15 sont encore déchaînées et ensuite agrégées au niveau A6, puis rechaînées, ce qui explique la différence entre certains résultats de la feuille *Demetra\_Results\_fa* et les résultats issus de la publication des comptes trimestriels.

En résumé, il a été essayé, dans le présent document, de fournir un aperçu théorique et pratique sur la désaisonnalisation. Pour de plus amples informations concernant la théorie de la désaisonnalisation, la bibliographie qui suit pourra s'avérer utile. En ce qui concerne l'implémentation de cette théorie au Statec et l'utilisation des différents programmes de désaisonnalisation, des renseignements supplémentaires peuvent être obtenus auprès de Bastien Larue (tél. 247-84339) de l'unité «Conjoncture, prévisions et modélisation» et auprès de Véronique Elter (tél. 247-84335) de l'unité «Comptes nationaux».

## **Bibliographie**

- [1]\* Findley F. D., Seasonal Adjustment, Central Bureau of Statistics, Washington D. C., pages 1-40
- [2] Fiorentini G. et Maravall A., 1996. *Unobserved Components in ARCH Models: An Application to Seasonal Adjustment*, Journal of Forecasting, Vol. 15, pages 175-201
- [3] Gómez V., Maravall A., Seasonal Adjustment and Signal Extraction in Economic Time Series, Ministerio de Economía, Banco de España, Madrid, pages 1-45
- [4]\* Harhoff C., Seasonal Adjustment, Danmarks Statistik, Danemark, pages 1-68
- [5] Kaiser R. et Maravall A., Combining Filter Design with Model-Based Filtering (with an Application to Business-Cycle Estimation, Universidad Carlos III, Banco de España, Madrid, pages 1-33
- [6]\* Kaiser R. et Maravall A., *Notes on Time Series Analysis, ARIMA Models and Signal Extraction*, Universidad Carlos III, Banco de España, Madrid, pages 1-72
- [7] Ladiray D. et Quenneville B., mars 2001. Seasonal Adjustment with the X-11 Method, Lecture Notes in Statistics 158, Springer, pages 1-256
- [8] Maravall A., janvier 2003. A Class of Diagnostics in the ARIMA-Model-Based Decomposition of a Time Series, Banco de España, Madrid, pages 1-23
- [9]\* Maravall A., mars 2005. An Application of the TRAMO-SEATS Automatic Procedure; Direct versus Indirect Adjustment, Banco de España, Madrid, pages 1-10
- [10] Maravall A. et Pierce D. A., 1987. *A Prototypical Seasonal Adjustment Model*, Journal of Time Series Analysis, Vol. 8, No. 2, pages 177-193
- [11] Maravall A. et Planas C., 1999. Estimation error and the specification of unobserved component models, Journal of Econometrics, 92, pages 325-353
- [12] Maravall A., *Minimum Mean Squared Error Estimation of the Noise in Unobserved Component Models*, Journal of Business and Economic Statistics, 5, pages 115-120
- [13] Maravall A., octobre 1985. On Structural Time Series Models and the Characterization of Components, Banco de España, Madrid, pages 1-17
- [14] Maravall A., *Unobserved Components in Economic Time Series*, Banco de España, Madrid, pages 12-72
- [15] Mélard G., Méthodes de Prévision à Court Terme, Editions de l'Université de Bruxelles, Editions Ellipses

Service central de la statistique et des études économiques

13, rue Erasme B.P. 304 L-2013 Luxembourg

Téléphone 247 84219 Fax 46 42 89

E-mail: info@statec.etat.lu Internet: www.statec.lu

Juillet 2009

ISBN 978-2-87988-0090-7

La reproduction est autorisée à condition de mentionner la source.