

Direction des Statistiques d'Entreprises

E 2008/01

**Désaisonnalisation des séries mensuelles
de créations d'entreprises**

Virginie FABRE

Document de travail



Institut National de la Statistique et des Études Économiques

Institut National de la Statistique et des Études Économiques

*Série des documents de travail
de la Direction des Statistiques d'Entreprises*

E 2008/01

**Désaisonnalisation des séries mensuelles
de créations d'entreprises**

*Virginie FABRE
Responsable des indicateurs conjoncturels de Démographie d'Entreprises*
Division Administration du Répertoire et Démographie d'Entreprises*

Janvier 2008

L'auteur remercie Dominique Ladiray, qui, par ses précieux conseils, a apporté une contribution essentielle tout au long de ce travail.

* Faisait partie de l'Unité « Répertoire et démographie des entreprises » au moment de la rédaction de ce document
Direction des Statistiques d'Entreprises - Timbre E100 - 18, bd Adolphe Pinard - 75015 Paris - France -

*Ces documents de travail ne reflètent pas la position de l'INSEE et n'engagent que leurs auteurs.
Working papers do not reflect the position of INSEE but only their author's views.*

Désaisonnalisation des séries mensuelles de créations d'entreprises

Résumé :

La gestion quotidienne du répertoire inter-administratif des entreprises et des établissements, confiée à l'Insee depuis 1973 constitue une source unique pour suivre de façon conjoncturelle - moyennant les précautions nécessaires dans l'emploi de toute source administrative pour mesurer un phénomène économique - la création d'entreprises de toute nature.

Le suivi des créations a fait récemment l'objet d'une refonte importante en vue d'en harmoniser au plan européen les statistiques annuelles - les seules généralement observables, sauf à disposer d'un outil comme le répertoire français.

Les séries publiées mensuellement en France depuis une quinzaine d'années ont été révisées, pour s'appuyer sur des définitions cohérentes avec les nouvelles séries annuelles harmonisées. Le nouveau dispositif est en place depuis la mi-février 2007, avec la publication des statistiques de janvier 2007, accompagnée des séries rétropolées depuis 1993.

Ces travaux ont comporté un important volet consacré à la correction des variations saisonnières des séries mensuelles, qui fait l'objet du présent document de travail.

La démarche en est originale à plus d'un titre :

Les données brutes sont sensibles non seulement aux fluctuations économiques, mais aussi à certains phénomènes d'origine administrative ou au comportement des acteurs

Pour des raisons liées au calendrier général de l'opération, le choix de la méthode s'est appuyé, en 2006, sur l'étude des anciennes séries de création

Un large usage des méthodes d'analyse des données (ACP, ACH) a aidé à ce choix

La possibilité d'introduire de nouveaux outils de désaisonnalisation dans ces analyses pour en comparer l'effet par rapport à ceux testés en 2006.

A l'issue de ce travail et en prenant en compte les contraintes imposées, une méthode s'avère répondre à tous les critères nécessaires. La méthode de correction de variations saisonnières choisie est celle actuellement utilisée dans la publication des indicateurs de créations d'entreprises.

Mots clés : indicateurs démographies d'entreprises, corrections de variations saisonnières, corrections de jours ouvrables, X12 ARIMA, TRAMO SEATS

Seasonal Adjustment of monthly enterprises' creations

Abstract:

The enterprises and institutions inter-administrative list's daily management entrusted to the INSEE since 1973 represents an unique source in order to follow in a cyclical way – in consideration of the necessary precautions when using any administrative source to measure an economic phenomenon – the creation of any sort of enterprises.

The creation monitoring has lately been the subject of a thorough revision in order to harmonize on the European level annual statistics - the latter being generally the observable ones apart from having a tool like the French list at one's disposal.

The series, which have been being monthly published in France for about fifteen years, have themselves been reviewed so as to rely on coherent definitions for the newly harmonized annual series. The new device has been established since mid-February 2007 with the publishing of January 2007 statistics enclosed with "retropeled" series since 1993.

These works comprised a significant panel dedicated to the correcting of monthly series' seasonal variations which is the subject of the hereby working document.

The thought process is greatly original for many reasons:

Not only were raw data sensitive to economic fluctuations but also to some phenomena of administrative origin or to the behaviour of economic agents.

For reasons linked to the general operation timetable, the method choice was based in 2006 on the study of former creating series.

Wide use of data analysis methods (PC, HCM) helped to make this choice

The possibility of introducing new seasonal adjustment tools in these analysis in order to compare the effect with those tested in 2006.

At the conclusion of this work and by taking into account the imposed limiting factors, a method turn out to answer all the necessary criteria. The chosen correcting seasonal variations' method is the one which is being at the present time used when publishing every month enterprises' creations indicators.

Key words: Business Demography Indicators, seasonal adjustment, trading days adjustment, X12 ARIMA, TRAMO SEATS

AVANT-PROPOS

La gestion quotidienne du répertoire inter-administratif des entreprises et des établissements, confiée à l'Insee depuis 1973 constitue une source unique pour suivre de façon conjoncturelle - moyennant les précautions nécessaires dans l'emploi de toute source administrative pour mesurer un phénomène économique - la création d'entreprises de toute nature.

Le suivi des créations a fait récemment l'objet d'une refonte importante en vue d'en harmoniser au plan européen les statistiques annuelles - les seules généralement observables, sauf à disposer d'un outil comme le répertoire français.

Les séries publiées mensuellement en France depuis une quinzaine d'années ont été elles-mêmes révisées, de façon à s'appuyer sur des définitions cohérentes avec les nouvelles séries annuelles harmonisées. Le nouveau dispositif est en place depuis la mi-février 2007, avec la publication des statistiques de janvier 2007, accompagnée des séries rétropolées depuis 1993.

Ces travaux ont comporté un important volet consacré à la correction des variations saisonnières des séries mensuelles, qui fait l'objet du présent document de travail.

La démarche en est originale à plus d'un titre :

- les données brutes sont sensibles non seulement aux fluctuations économiques, mais aussi à certains phénomènes d'origine administrative ou au comportement des acteurs
- pour des raisons liées au calendrier général de l'opération, le choix de la méthode s'est appuyé, en 2006, sur l'étude des anciennes séries de création
- un large usage des méthodes d'analyse des données (ACP et CAH) a aidé à ce choix
- la possibilité d'introduire de nouveaux outils de dessaisonnalisation dans ces analyses pour en comparer l'effet par rapport à ceux testés en 2006.

Les travaux de Mlle Fabre méritent donc toute l'attention du lecteur.

Michel Euriat

Chef de l'unité « Répertoire et démographie des entreprises et des établissements »

Problématique	9
Introduction	11
1. L'origine des données	11
2. Principe de construction des données sur les créations d'entreprises	13
3. Particularités de ces données	14
Partie 1 : Analyse de la méthode utilisée, choix des alternatives	15
1. Analyse de la méthode utilisée	15
2. Méthodes alternatives	16
a) Alternative à la méthode de correction de variations saisonnières : X12 ARIMA et Tramo Seats	17
b) Alternatives à la méthode de correction de jours ouvrables : modèles économétriques	17
Partie 2 : Résultats	20
1. Qualité de la désaisonnalisation	20
a) Qualité des méthodes de désaisonnalisation adoptées	20
b) Classification Ascendante Hiérarchique des séries désaisonnalisées	21
c) Analyse en Composantes Principales des séries désaisonnalisées	23
2. Cohérence de la désaisonnalisation avec l'information des séries	24
a) Cohérence	24
b) Statistiques de lissage et de qualité	25
c) Classification Ascendante Hiérarchique des caractéristiques des séries désaisonnalisées	28
d) Analyse en Composantes Principales des caractéristiques des séries désaisonnalisées	28
e) Comparaison des 2 « meilleures » séries CVS-CJO : TD2-X12 et TD2-TS avec la série actuelle	32
f) Décomposition précise des 2 séries TD2_X12 et TD2_TS	32
Conclusion	36

Problématique

Les séries de démographie d'entreprises, créations et défaillances, sont des indicateurs importants dans l'appréciation de la situation économique conjoncturelle. Pour interpréter ces indicateurs mensuellement, une correction des variations saisonnières est indispensable. Elle permet de comparer les évolutions entre deux mois consécutifs, ainsi que les chiffres au niveau international, sans que les effets de calendrier n'impactent ces valeurs. Une bonne désaisonnalisation permet d'obtenir des indicateurs lisibles et interprétables. Elle est d'autant plus essentielle que la série des créations d'entreprises suscite un vif intérêt depuis le début des années 2000, suite aux nouvelles politiques engagées sur ce domaine.

L'objectif de ce travail est principalement de construire une méthode de désaisonnalisation adaptée à une série particulière, celle des créations d'entreprises. La désaisonnalisation des séries nécessite de mettre en place deux méthodes : une correction de jours ouvrables et une correction de variations saisonnières (encadré).

La série de créations d'entreprises est désaisonnalisée tous les mois pour être publiée sous forme d'*Informations Rapides*. Cette série est issue du Répertoire des Entreprises et des Établissements (Sirene), géré par l'Insee. Elle est très dépendante du contexte et des règlements administratifs en vigueur ; ainsi une loi ou un changement administratif peuvent influencer le nombre de créations d'entreprises. De plus, le gouvernement ayant fait de la création d'entreprises une de ses actions prioritaires, les aides et dispositifs particuliers favorisant la création d'entreprises se sont multipliés au début des années 2000. Les créations ont ainsi crû de près de 26 % entre 2003 et 2005, ce qui a un impact sur la saisonnalité de la série, fortement modifiée, et donc nécessairement sur la désaisonnalisation. Il semble que la méthode de désaisonnalisation utilisée jusque là ne tienne pas suffisamment compte des particularités de la série et de son évolution au cours des dernières années.

La démarche adoptée dans ce document a été de présenter les données de l'étude, d'analyser la qualité de la méthode de désaisonnalisation utilisée jusqu'alors, de lister les possibilités de modifications disponibles en matière de correction de jours ouvrables et de variations saisonnières. Enfin, une série de tests a été réalisée pour juger de la qualité de ces différentes méthodes et voir celle qui restitue le mieux l'information contenue dans la série brute des créations. L'échéance de ce travail était prévue pour janvier 2007, de façon à coupler le changement de désaisonnalisation avec le changement de définitions et de champ des créations d'entreprises. L'analyse a été faite au cours de l'année 2006, sur les séries de 1993 à 2005.

Définitions

Correction de jours ouvrables :

Cette correction permet d'éliminer des effets « type de jours dans la semaine » (l'intensité des ventes du commerce de détail est plus importante le samedi que les autres jours de semaine par ex.), et des effets « jours fériés et fêtes mobiles » (le mois de mai comporte beaucoup de jours fériés, il y a moins d'activité ce mois-ci). De plus, certains jours fériés ne tombent pas toujours le même mois ou le même trimestre, leur effet sur l'activité économique n'est donc pas purement saisonnier.

Correction de variations saisonnières :

Cette correction permet d'éliminer les fluctuations saisonnières de la série (ex : l'activité aux mois de juillet et août est souvent ralentie).

Le partage entre les effets « jours ouvrables » et les effets « saisonniers » est délicat car ceux-ci ne sont pas toujours indépendants : ils se superposent en partie. Ce phénomène apparaît en présence de fêtes à date fixe mais « à jour flottant » (par exemple le jour de Noël : toujours le 25 décembre, mais pas systématiquement le week-end ou en semaine). Ainsi, le traitement de la saisonnalité n'éliminera pas totalement les effets « jours ouvrables ». A cette difficulté, s'ajoute la présence dans le calendrier français de certaines fêtes à « jour fixe » mais à date mobile (lundi de Pâques, jeudi de l'Ascension, lundi de Pentecôte).

Introduction

1. L'origine des données

La série des créations d'entreprises est constituée à partir des informations issues du répertoire des entreprises et des établissements Sirene (Système Informatique pour le Répertoire des Entreprises et de leurs Établissements). Il enregistre l'état civil de toutes les entreprises et de leurs établissements, quelle que soit leur forme juridique et quel que soit leur secteur d'activité, situés en métropole ou dans les DOM (Guadeloupe, Guyane, Martinique et Réunion). Les entreprises étrangères qui ont une représentation ou une activité en France y sont également répertoriées. L'Insee attribue à chaque entreprise un identifiant de 9 chiffres appelé numéro SIREN et à chaque établissement un identifiant de 14 chiffres, appelé numéro SIRET, composé du numéro SIREN de l'entreprise suivi d'un numéro d'ordre de 5 chiffres, le NIC (Numéro Interne de Classement).

La démographie économique dénombre, à travers le concept de "création" d'entreprises, *tous les mouvements qui conduisent à la prise d'activité économique d'une personne physique ou d'une personne morale*. La création d'entreprises est donc toujours associée à la création d'au moins un établissement économiquement actif dans le champ ICS (Industrie, Construction, Commerce et Services).

Au moment de l'étude, la typologie retenue distinguait trois types de créations : les créations nouvelles, les reprises et les réactivations. Ces trois séries étaient désaisonnalisées de façon indépendante et la série désaisonnalisée globale obtenue comme somme de ces trois séries désaisonnalisées.

- La **création d'une entreprise nouvelle** correspond à la création du premier établissement économiquement actif jusqu'alors inexistant, et donc à l'exploitation de nouveaux moyens de production. On parle aussi de création "pure" ou de création "ex-nihilo". D'un point de vue économique, elle correspond à une réelle augmentation des capacités de production.
- La « **reprise** » intervient lorsqu'une unité légale reprend l'activité, totalement ou partiellement, d'un ou plusieurs établissements économiques d'une autre unité légale.
- La « **réactivation** » est le fait pour une entreprise qui avait cessé son activité de la reprendre. Ce phénomène concerne les personnes physiques dans la mesure où une personne physique garde à vie son immatriculation numéro SIREN. En effet, les personnes morales peuvent se dissoudre alors que les personnes physiques ne disparaissent définitivement qu'en cas de décès.

Le concept de "création" d'entreprises était donc pris dans le sens large d'entreprises nouvellement actives. Il prenait en compte tous les mouvements économiques qui permettaient d'accroître les moyens de production, en les créant ou les réactivant, mais aussi en les maintenant en exploitation en raison d'un changement d'exploitant. Il ne se limitait pas aux nouvelles immatriculations enregistrées dans le répertoire Sirene.

A titre indicatif, ce tableau montre la proportion de chaque type de création dans la série totale et l'évolution de ces créations lors des six dernières années d'observation disponibles au moment de l'étude :

	2 000	2 001	2 002	2 003	2 004	2 005
Créations pures	176 754	177 029	177 995	199 290	223 926	224 828
Reprises	41 652	41 560	40 124	39 983	42 223	40 831
Réactivations	53 666	51 995	52 050	54 298	53 831	51 163
Part des créations pures	65%	65%	66%	68%	70%	71%
Part des réactivations	15%	15%	15%	14%	13%	13%
Part des reprises	20%	19%	19%	18%	17%	16%
Ensemble	272 072	270 584	270 169	293 571	319 980	316 822

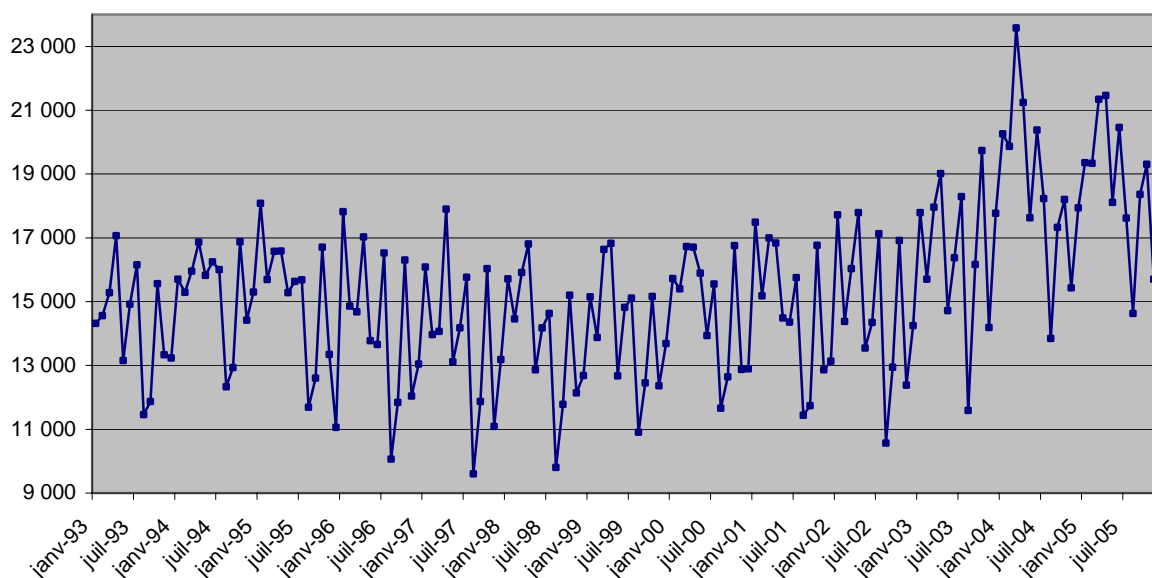
Source : Insee, Répertoire des Entreprises et des Établissements Sirene

L'étude ne prend en compte que la série de **créations pures d'entreprises**. Tout d'abord parce qu'elle représentait près de trois créations sur quatre, mais aussi et surtout parce qu'elle se rapprochait de la future série des « créations d'entreprises ». En effet, depuis la refonte de la démographie des entreprises en janvier 2007, la définition des créations d'entreprises s'appuie, pour faciliter les comparaisons, sur des concepts harmonisés au niveau européen :

une création d'entreprise correspond à l'apparition d'une unité légale exploitante n'ayant pas de prédécesseur. Il n'y a création d'une entreprise que si elle s'accompagne de la mise en œuvre de nouveaux moyens de production. La nouvelle notion de création d'entreprise est plus large que celle de création pure puisqu'elle inclut notamment les réactivations d'entreprises dont la dernière cessation remonte à plus d'un an et les reprises d'entreprises s'il n'y a pas continuité de l'entreprise. Cette continuité est évaluée à la lumière de la localisation de l'entreprise et de l'activité qui y est exercée. Ces changements de concepts ont aussi été l'occasion d'élargir le champ sur lequel porte les créations d'entreprises à l'ensemble des activités marchandes hors agriculture. En effet, la qualité du répertoire (Répertoire des Entreprises et des Établissements - Sirene) sur certains sous-ensembles permet désormais leur prise en compte (activités financières, activité de location de biens immobiliers ou catégories juridiques marginales dont certains établissements publics administratifs ou sociétés civiles).

Ces nouvelles séries ont été publiées en même temps que la nouvelle méthode de désaisonnalisation. L'objectif d'une publication en février 2007 nécessitait de débiter l'analyse de correction de variations saisonnières sans attendre la nouvelle série rétropolée. Compte tenu de ce calendrier, seuls d'ultimes tests de qualité et de cohérence sur les nouvelles séries pouvaient être réalisés avec la méthode de désaisonnalisation choisie. Mais la série de créations pures utilisée représente environ 90 % de ces nouvelles « créations » d'entreprises.

Série brute des créations pures



2. Principe de construction des données sur les créations d'entreprises

Chaque jour, les gestionnaires du répertoire Sirene enregistrent des créations d'entreprises en provenance des Centres de Formalités des Entreprises (CFE). Ces mouvements de créations sont enregistrés dans le répertoire à partir duquel sont construits des fichiers mensuels. Ces fichiers, qui regroupent les mouvements des entreprises et des établissements, sont le point de départ de la statistique de créations d'entreprises.

La statistique de créations d'entreprises est une statistique de "flux". Elle se construit en référence à une période donnée, par exemple le nombre de créations d'entreprises enregistrées au cours d'un mois, d'un trimestre ou d'une année.

Le choix important retenu pour construire les bases de données conjoncturelles sur les créations d'entreprises est celui de dater les mouvements à partir des dates d'enregistrement dans Sirene.

Le choix de la datation des mouvements de créations à partir de la date d'enregistrement dans le répertoire Sirene repose sur une immatriculation rapide des nouvelles entreprises par le circuit des centres de formalités des entreprises. Pour les créations d'entreprises, le délai est très court entre la date d'événement (la date "réelle" de création) et la date d'enregistrement (la prise en compte dans le répertoire Sirene), et donc dans le fichier des créations d'entreprises. De fait, 95 % des créations d'un mois donné sont enregistrées dans Sirene le même mois. En conséquence, exploiter les données en date d'enregistrement a tout son sens et ce choix permet d'avoir des statistiques mensuelles de créations très rapidement, dès les premiers jours du mois suivant.

Les créations en date d'enregistrement dans Sirene constituent donc un très bon indicateur mensuel des créations. En revanche, pour observer les créations de façon journalière, ce choix ne serait pas sans conséquence. En effet, avec l'apparition des 35 heures par exemple, il est possible que moins d'enregistrements soient saisis par les gestionnaires les vendredis ou lors des vacances scolaires.

432 séries étaient désaisonnalisées chaque année à partir de ces fichiers de créations : les séries correspondant aux trois types de créations et la répartition des créations selon 13 secteurs d'activité, soit 16 séries sur la France Entière, et autant pour chacune des 22 régions et des 4 DOM.

3. Particularités de ces données

La série mensuelle issue du répertoire Sirene permet de publier un indicateur conjoncturel le mois suivant le mois de la création.

Depuis janvier 2003, les séries subissent une évolution atypique. Plusieurs causes sont possibles. Des changements impliqués par un nouveau contexte administratif ou une nouvelle législation en sont une, ce qui peut affecter les séries assez fortement. Les fichiers issus d'un répertoire administratif sont souvent liés à l'actualité. Ici par exemple, les perturbations observées dans les courbes désaisonnalisées ont pu être liées à la modification en 2003 des modalités de paiement des charges sociales : avant cette modification, le paiement de ces charges s'effectuait pour tout le trimestre, quel que soit le jour de création de l'entreprise dans le trimestre ; les créateurs avaient donc tendance à créer leur entreprise en début de trimestre. Depuis août 2003, il se fait en tenant compte du jour de création de l'entreprise. Ce qui n'est pas sans conséquence sur la saisonnalité de la série.

Il fallait également prendre en compte les décisions législatives telles que la loi d'initiative économique d'août 2003 qui favorise la création d'entreprises.

L'intervention des gestionnaires dans l'enregistrement d'une création dans le répertoire est une autre cause de l'évolution des séries. Avec les 35 heures et le dispositif d'ARTT, les créations d'entreprises enregistrées le vendredi et le lundi ont pu baisser alors que cette baisse ne reflète pas un changement réel. L'évolution du système informatique utilisé en lien avec le répertoire a rompu d'une certaine façon la série de données. Aucune création n'a lieu les samedis et dimanches. Or, depuis août 2004, le système informatique traite également des créations le dimanche (entre 100 et 300 contre 1 000 à 1 500 les autres jours). Tout se passe donc comme si la série des créations se faisait sur 6 jours au lieu de cinq, le sixième jour étant plus faible.

Pourtant, ces variations n'étaient pas prises en compte dans la correction de jours ouvrables effectuée jusque là. Il s'avérait donc nécessaire de chercher si une méthode de désaisonnalisation pouvait s'adapter à ces séries et faire en sorte que l'évolution de la saisonnalité ne perturbe pas la série désaisonnalisée.

Partie 1 : Analyse de la méthode utilisée, choix des alternatives

1. Analyse de la méthode utilisée

Note méthodologique :

La série brute se décompose en plusieurs éléments de base :

- **la composante tendance cycle** traduit l'évolution de fonds de la série observée sur une longue période. C'est une composante lisse ne traduisant pas d'évolution brusque.
 - Tendance : alternance de périodes de hausse et de baisse dont la longueur varie.
 - Cycle : mouvement lisse et presque périodique de la série que l'on observe souvent autour de la composante tendancielle de long terme.
- **la composante saisonnière (« saisonnalité »)** représente les fluctuations mensuelles ou trimestrielles, à l'intérieur même d'une année, qui se répètent plus ou moins régulièrement d'une année à l'autre.
- **les effets de calendrier**
 - la composante jours ouvrables, liée à la composition journalière du mois ou du trimestre.
 - l'« effet de Pâques », lié au fait que Pâques est en mars ou en avril, et donc au 1^{er} ou au 2nd trimestre.
- **la composante irrégulière (l'« irrégulier »)** : fluctuations résiduelles et erratiques qui ne peuvent être attribuées aux autres composantes « systématiques » : tendance-cycle, saisonnalité, composantes liées aux effets de calendrier (ex. : grèves, erreur d'une mesure...)

Les composantes saisonnières et de calendrier ne sont pas nécessairement présentes dans toutes les séries. Il s'agit de tester leur présence, notamment l'effet du lundi de Pâques, par des programmes adaptés.

La méthode de désaisonnalisation utilisée se faisait jusqu'ici en deux temps :

- on appliquait **un coefficient de correction de jours ouvrables** à la série. Ce coefficient était calculé de la manière suivante :

$$\text{Coefficient CJO du mois } m \text{ de l'année } n = \frac{\text{nombre de jours ouvrables du mois } m}{\text{nombre moyen de jours ouvrables mensuels de l'année } n}$$

Cette méthode de correction dite « proportionnelle » respectait le calendrier français puisqu'elle tenait compte des jours fériés.

- à partir de cette série corrigée de jours ouvrables, on utilisait la **procédure X11 du logiciel SAS** qui permettait une désaisonnalisation quasi automatique. Le programme de la procédure X11 teste cinq modèles ARIMA et choisit le meilleur pour la série proposée. Il se peut qu'aucun des cinq modèles ne satisfassent les tests et les critères imposés par le logiciel¹. Dans ce cas, une adaptation à la main était nécessaire, c'est-à-dire qu'il fallait forcer le logiciel à appliquer un modèle ARIMA (par exemple le modèle : P=2 Q=0 SP=0 SQ=1 DIF=0 SDIF=1), qui satisfait les tests pour près de 90 % des séries sans être nécessairement le modèle le mieux adapté.

Les dernières campagnes de désaisonnalisation en janvier 2005 et 2006 ont révélé quelques problèmes de mise en œuvre. Les modèles et les résultats n'étaient pas stables. La méthode n'est pas uniforme entre les séries puisque certaines séries étaient désaisonnalisées « automatiquement » (sans intervention du responsable) et d'autres nécessitaient une intervention et une modification du programme informatique pour être désaisonnalisées. De plus, des méthodes plus pertinentes et plus récentes étaient alors disponibles.

A priori, la méthode de correction de jours ouvrables utilisée alors n'était pas suffisamment réaliste et la chaîne était de plus en plus difficile à maintenir, en particulier parce que les séries des créations avaient beaucoup évolué notamment depuis 2003.

Des méthodes plus concluantes de correction de variations de jours ouvrables sont aujourd'hui utilisées. Il a fallu aussi s'assurer que la méthode utilisée ne reflétait pas la réalité de la série, et dans un deuxième temps, voir si les nouvelles méthodes s'adaptaient et surtout s'adaptaient mieux aux séries des créations.

Il a fallu faire un choix une fois les résultats obtenus en gardant à l'esprit que la méthode utilisée présentait toutefois certains avantages : la méthode de correction de jours ouvrables était simple et la réalisation entièrement faite sous SAS, ce qui permettait un meilleur savoir faire et une facilité de gestion des éventuels problèmes.

2. Méthodes alternatives

Il fallait dans un premier temps lister les possibilités de changement par rapport à la méthode utilisée. Ces changements étaient de deux types :

- Le premier changement possible consistait à modifier la méthode de corrections de variations saisonnières. Parmi les avancées les plus récentes : X12 ARIMA et TramoSeats permettent une désaisonnalisation automatique. Il était important de se limiter

¹ Le logiciel vérifie qu'il n'existe pas d'autocorrélation significative dans les résidus du modèle, que le modèle n'est pas sur-différencié et que la qualité de l'ajustement sur les 3 dernières années est satisfaisante.

aux méthodes dites « automatiques », c'est-à-dire qui nécessitent le moins possible l'intervention du responsable, notamment à cause du nombre élevé de séries à désaisonnaliser et des délais relativement courts pour réaliser cette opération.

▪ Le second changement possible consistait à modifier la méthode de correction de jours ouvrables : certains jours pouvaient présenter un volume plus important de créations que d'autres jours de la semaine. Or, la méthode de correction de jours ouvrables « manuelle » consistait à attribuer le même poids à tous les jours d'un même mois, ce qui ne semblait pas être une hypothèse très réaliste. Des modèles économétriques permettent de pondérer les jours différemment.

a) Alternative à la méthode de correction de variations saisonnières : X12 ARIMA et TramoSeats

X12 ARIMA a l'avantage d'être une méthode plus pertinente que X11 ARIMA et d'avoir notamment une détection et une correction de points atypiques automatiques. En revanche, la programmation de la procédure X12 n'est pour l'instant pas finalisée sous SAS et nécessite donc de « rompre la chaîne » lors de l'exécution des programmes. Toutefois, on espère que la finalisation sera prochaine et permettra une exécution entièrement sous SAS.

Une désaisonnalisation sous TramoSeats n'a pas été écartée et a été étudiée, mais ne restait envisageable que si elle apportait une amélioration très significative par rapport à X12 ARIMA. En effet, l'utilisation de X12 ARIMA est a priori privilégiée sur celle de TramoSeats, afin notamment de réduire les différences avec l'ancienne série désaisonnalisée, de permettre une continuité dans la chaîne de traitement et programmation et de privilégier la plus grande facilité d'utilisation de SAS.

b) Alternatives à la méthode de correction de jours ouvrables : modèles économétriques

Même si la méthode économétrique de correction de jours ouvrables développée aujourd'hui ne tient pas suffisamment compte de la réalité, elle le fait tout de même plus que celle par correction « manuelle » utilisée jusqu'ici. En effet, elle tient compte du poids de chaque jour et de l'évolution de celui-ci au cours du temps. Cette méthode tient aussi compte du calendrier national (fêtes nationales, jours fériés) et la modélisation reste automatique.

La poursuite du travail consistait à comparer à la fois les logiciels de désaisonnalisation et les méthodes de correction de jours ouvrables. Deux modèles étaient possibles : une méthode appelée TD (Trading Days) et une méthode NC (National Calendar). La première faisant une correction de jours ouvrables sans prise en compte du calendrier français et la seconde le prenant en compte.

Dans chaque méthode, deux modèles ont été envisagés : une régression opposant jours de semaine ouvrés et week-end (régression sur deux variables notée NC2) et une opposant chaque jour au dimanche (régression sur sept variables notée NC7). Les tests sur les logiciels ont permis de comparer la correction de variation saisonnière actuelle (X11 ARIMA) à la désaisonnalisation faite par X12 ARIMA et TramoSeats. Il y avait donc cinq séries pour chacun des deux logiciels, soit 10 séries à tester et à comparer avec la série jusqu'alors utilisée.

Les tests ont consisté à juger de la qualité de désaisonnalisation, à étudier les caractéristiques des séries désaisonnalisées et à synthétiser et comparer les résultats.

- **série X11_CJO** (méthode utilisée) : correction de jours ouvrables « manuelle » et correction de variations saisonnières faite par la procédure X11.

Avec le logiciel de désaisonnalisation SAS :

- **série CJO _ X12** : correction de jours ouvrables « manuelle ».
- **série NC2_ X12** : méthode de correction de jours ouvrables à 2 régresseurs (jours de semaine versus week-end et effet de l'année bissextile) avec prise en compte du calendrier français.
- **série NC7_ X12** : méthode de correction de jours ouvrables à 7 régresseurs (6 contrastes entre chaque jour de la semaine et le dimanche plus un effet de l'année bissextile) avec prise en compte du calendrier français.
- **série TD2_ X12** : méthode de correction de jours ouvrables à 2 régresseurs sans prise en compte du calendrier français.
- **série TD7_ X12** : méthode de correction de jours ouvrables à 7 régresseurs sans prise en compte du calendrier français.

Avec le logiciel de désaisonnalisation TramoSeats:

- **série CJO_TS** : correction de jours ouvrables « manuelle ».
- **série NC2_TS** : méthode de correction de jours ouvrables à 2 régresseurs avec prise en compte du calendrier français.
- **série NC7_TS** : méthode de correction de jours ouvrables à 7 régresseurs avec prise en compte du calendrier français.
- **série TD2_TS** : méthode de correction de jours ouvrables à 2 régresseurs sans prise en compte du calendrier français.
- **série TD7_TS** : méthode de correction de jours ouvrables à 7 régresseurs sans prise en compte du calendrier français.

Explication des méthodes et modèles utilisés :

Méthode TD : on utilise la structure du calendrier normal dans la modélisation des jours ouvrables, sans tenir compte des jours fériés français.

Modèle de régression de jours ouvrables à 2 variables

$$WD_t = (\# \text{ de lundis, mardis, } \dots, \text{ vendredis du mois } t) - 5/2 * (\# \text{ de samedis, de dimanches du mois } t)$$

Et un régresseur pour l'effet de l'année bissextile (voir ci-après)

Modèle de régression de jours ouvrables à 7 variables

$$T_{1,t} = (\# \text{ de lundis du mois } t) - (\# \text{ de dimanches du mois } t)$$

$$T_{2,t} = (\# \text{ de mardis du mois } t) - (\# \text{ de dimanches du mois } t)$$

.....

$$T_{6,t} = (\# \text{ de samedis du mois } t) - (\# \text{ de dimanches du mois } t)$$

Et un régresseur pour l'effet de l'année bissextile (voir ci-après)

Méthode NC : on utilise la structure du calendrier français dans la modélisation des jours ouvrables, en tenant donc compte des jours fériés français.

Modèle de régression de jours ouvrables à 2 variables

$$WD_t = (\# \text{ de lundis, mardis, } \dots, \text{ vendredis travaillés du mois } t) - 5/2 * (\# \text{ de samedis, de dimanches et de jours fériés du mois } t)$$

Et un régresseur pour l'effet de l'année bissextile (voir ci-après)

Modèle de régression de jours ouvrables à 7 variables

$$T_{1,t} = (\# \text{ de lundis travaillés dans le mois } t) - (\# \text{ de dimanches et de jours fériés du mois } t)$$

$$T_{2,t} = (\# \text{ de mardis travaillés dans le mois } t) - (\# \text{ de dimanches et de jours fériés du mois } t)$$

.....

$$T_{6,t} = (\# \text{ de samedis travaillés dans le mois } t) - (\# \text{ de dimanches et de jours fériés du mois } t)$$

Et un régresseur pour l'effet de l'année bissextile (voir ci-après)

Régresseur pour année bissextile :

$$LY_t = \begin{cases} 0.75 & \text{pour le mois de Février d'une année bissextile} \\ -0.25 & \text{pour les autres mois de Février} \\ 0 & \text{sinon} \end{cases}$$

Partie 2 : Résultats

En ce qui concerne les critères d'appréciation pour évaluer les différentes méthodes de désaisonnalisation, dans un premier temps, il s'agissait de juger de la « qualité » de la désaisonnalisation : comment chaque logiciel prend en compte les points atypiques et comment il décompose la série. Dans un second temps, il fallait juger de la « cohérence » des séries : vérifier si l'information de la série désaisonnalisée rend bien compte de l'information contenue dans la série brute (à l'exclusion de la saisonnalité et de la tendance).

1. Qualité de la désaisonnalisation

a) Qualité des méthodes de désaisonnalisation adoptées

Le tableau ci-dessous est issu des résultats obtenus lors de la désaisonnalisation des différentes séries corrigées des jours ouvrables, en distinguant les deux logiciels.

La première partie du tableau présente les résultats issus de TramoSeats. Elle permet de connaître le modèle ARIMA jugé le plus adapté à la série corrigée des jours ouvrables, selon la méthode de correction utilisée. Puis elle donne le nombre de points atypiques détectés sur ces différentes séries. Enfin, elle indique combien de tests ont échoué dans la désaisonnalisation faite par TramoSeats. Ici, toutes les méthodes ont réussi les tests imposés par le logiciel. Ceci permet donc d'« éliminer » des méthodes qui ne pourraient pas être automatisées puisqu'elles ne satisfont pas les tests. Il est aussi intéressant de constater que tous les modèles ARIMA choisis sont identiques, quelle que soit la méthode de correction de jours ouvrables appliquée à la série.

La seconde partie du tableau est extraite des résultats obtenus par la désaisonnalisation sous SAS.

Cette première approche permet de constater que les logiciels ne réagissent pas de la même façon face à des séries identiques.

Logiciel		CJO	TD7	TD2	NC7	NC2
TramoSeats	Model_T	(2,1,0)(0,1,1)	(2,1,0)(0,1,1)	(2,1,0)(0,1,1)	(2,1,0)(0,1,1)	(2,1,0)(0,1,1)
	Outliers	4	2	2	3	1
	TramoFail	0	0	0	0	0
SAS	automd	(0 1 1)(0 1 1)	(0 1 2)(0 1 1)	(2 1 0)(0 1 1)	(0 1 2)(0 1 1)	(2 1 2)(0 1 1)
	Outliers	1	1	1	0	1
	fail	0	1	0	1	0
	idseas	yes	yes	yes	yes	yes
	peakseas	none	none	none	none	none
	peaktd	none	none	sa irr	sa irr	sa
	Irreg	1,01	1,01	1,13	1,25	0,93
	Tcycle	24,75	21,87	24,08	22,79	21,67
	Season	75,36	74,81	71,39	71,33	72,77
	Padj	0,54	0,71	0,80	0,00	0,63
	Tdays	0,00	5,78	4,12	5,63	5,73

Automd : modèle ARIMA choisi par SAS

Outliers : nombre de points atypiques de la série

Fail : nombre de tests qui ont échoué dans la désaisonnalisation

Idseas : détection d'une saisonnalité

Peakseas : existence d'une saisonnalité résiduelle

Peaktd : existence d'un effet de jours ouvrables résiduel

Irreg : part de la composante irrégulière dans la série

Tcycle : part de la composante tendance cycle dans la série

Season : part de la composante saisonnière dans la série

Padj : part de la composante « points atypiques » de la série

Tdays : part des effets de calendrier dans la série

Ainsi, une même série peut satisfaire les tests pour un logiciel et ne pas les satisfaire pour l'autre. Trois séries ne répondent pas aux critères imposés par X12 : les séries TD2, TD7 et NC7 présentent soit un effet de jours ouvrables résiduels (Peaktd), soit un problème avec les statistiques de qualité (Fail). La méthode utilisée par TramoSeats et celle utilisée par SAS diffèrent assez nettement. Les deux logiciels ne vont d'ailleurs pas sélectionner le même modèle ARIMA pour une même série, et ce dans quasiment tous les cas. En outre, les logiciels ne détectent pas le même nombre de points atypiques. TramoSeats semble être beaucoup plus « sensible » sur ce point que SAS.

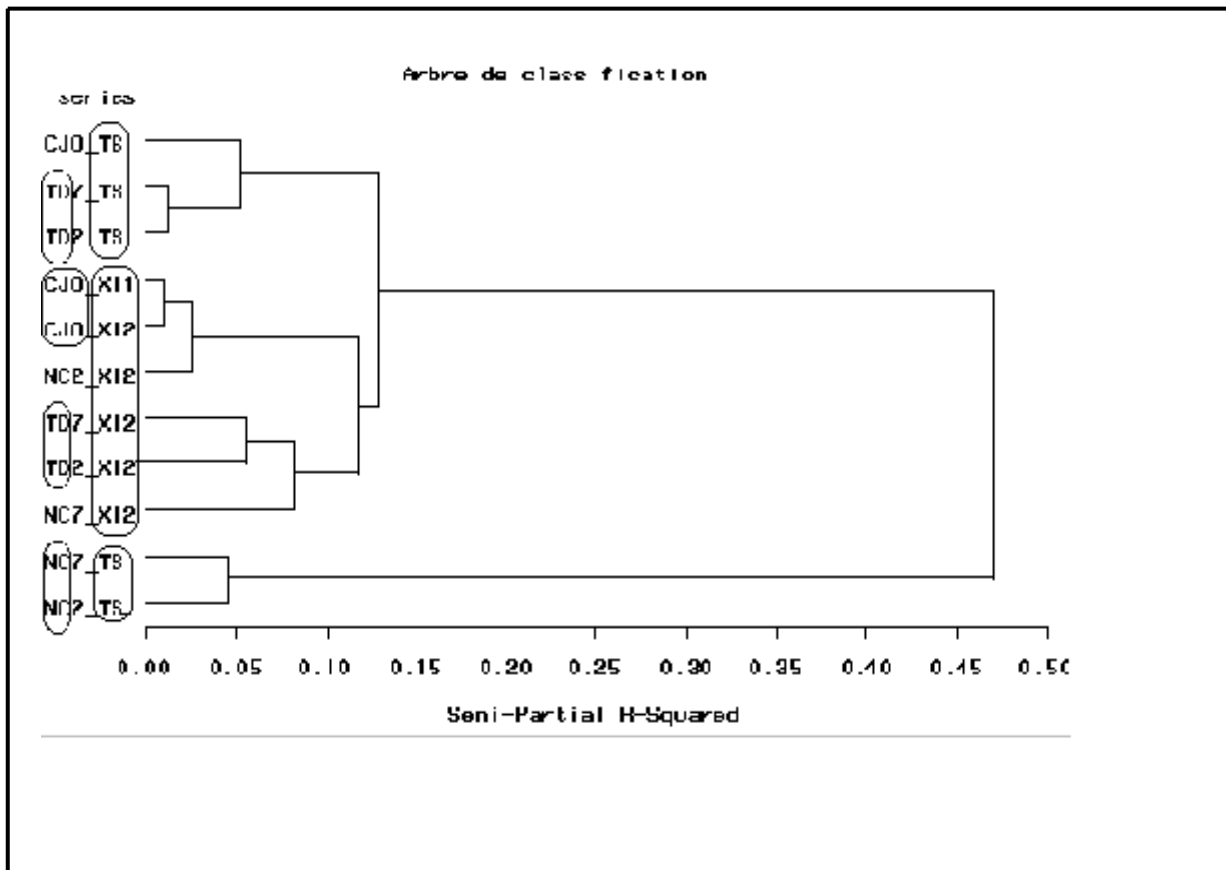
Concernant l'étude de jours ouvrables, on note que 4 à 6% de l'information contenue dans la série brute est imputable à un « effet jours ouvrables ». La composante « jours ouvrables » la plus forte est relevée dans la méthode TD7.

b) Classification Ascendante Hiérarchique des séries désaisonnalisées

La Classification Ascendante Hiérarchique (CAH) est une méthode dite « ascendante », c'est-à-dire que toutes les séries, identifiées par le modèle de correction de jours ouvrables et le logiciel de correction de variations saisonnières, sont considérées au départ comme différentes. Puis par un algorithme, on choisit les deux séries les plus proches, au sens de la distance euclidienne, et on les remplace par une série fictive moyenne. On recommence la procédure jusqu'à ce que toutes les séries soient regroupées dans la même classe. Chaque regroupement est associé à un indice de niveau qui permet de mesurer la perte d'information

que l'on concède en acceptant le regroupement de séries plus ou moins différentes dans une même classe. Comme les classes ainsi formées sont emboîtées, la classification est qualifiée de hiérarchique.

Ici les classes sont constituées à partir des valeurs prises par les différentes séries désaisonnalisées entre janvier 1993 et décembre 2005. Les séries désaisonnalisées regroupées seront donc celles qui auront des valeurs les plus proches durant la période d'observation.



Cette méthode consistant à découper une population donnée vise d'une part à faire en sorte que la différenciation entre les groupes soit maximale, et d'autre part à former des groupes dans lesquels les éléments se ressemblent le plus.

Ainsi, le premier regroupement visible est celui entre les méthodes de correction de jours ouvrables pour un même logiciel de désaisonnalisation ; il y a donc une proximité très nette d'une part entre les méthodes de corrections de jours ouvrables utilisant le calendrier national (NC) et d'autre part entre les méthodes ne l'utilisant pas (TD). Ces deux méthodes s'opposent donc.

La méthode de correction de jours ouvrables dite « proportionnelle » (CJO) donne pour le logiciel X12 des résultats proches de la méthode utilisant le calendrier national, ce qui semble a priori « normal » puisqu'elles tiennent compte toutes deux du même calendrier. En revanche, il semble étonnant que ceci ne soit pas vrai pour TramoSeats.

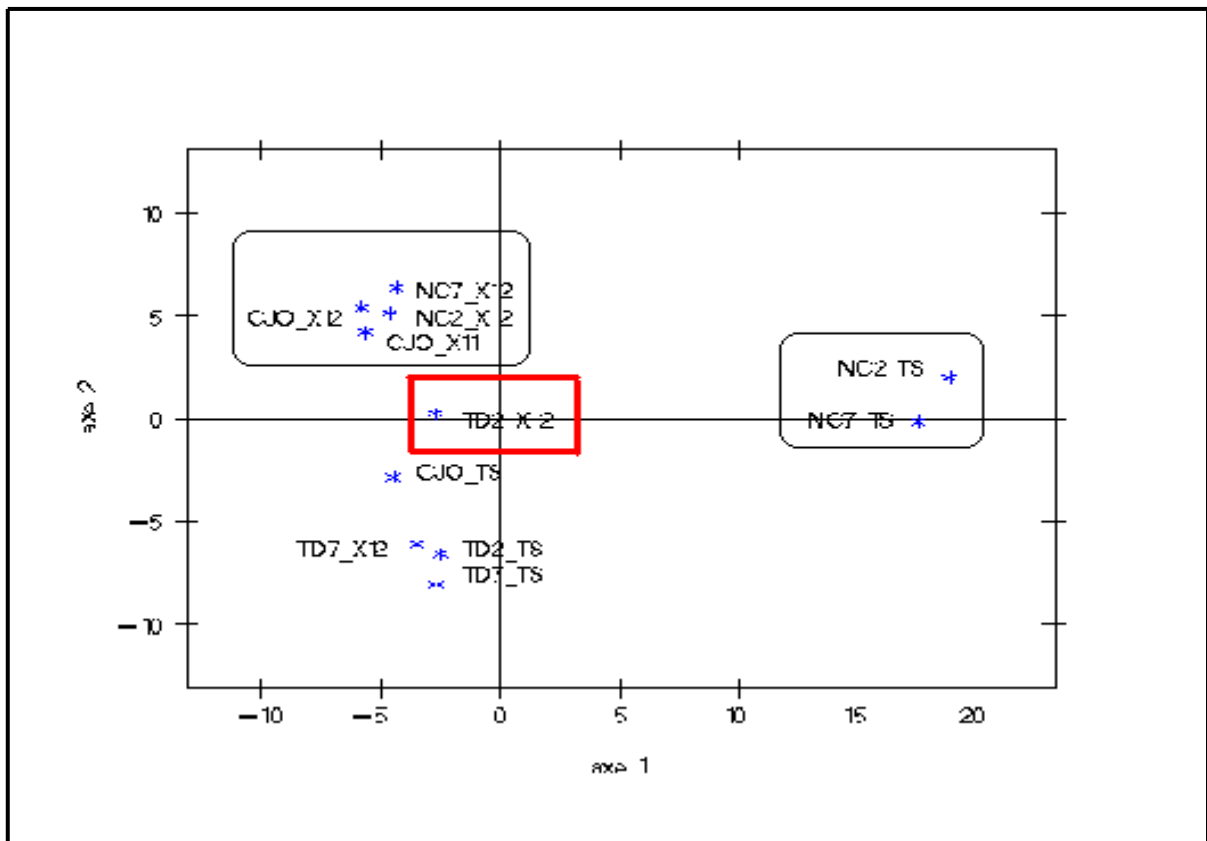
Au final, les méthodes utilisant un calendrier national (2 ou 7 régresseurs) associées au même logiciel de désaisonnalisation sont les méthodes les plus ressemblantes. Le nombre de régresseurs dans la régression calculant les coefficients de correction de jours ouvrables donnent des différences moins importantes que l'utilisation du calendrier national ou non.

A un niveau de regroupement moins fin, le regroupement se fait surtout selon le logiciel utilisé : X12 ou TramoSeats.

c) Analyse en Composantes Principales des séries désaisonnalisées

L'Analyse en Composantes Principales (ACP) est une technique de description statistique conduisant à des représentations graphiques approchées (mais en un certain sens optimales) du contenu d'un tableau de données : description simultanée des liaisons entre variables et des similitudes entre individus. C'est aussi un outil de réduction de la dimensionnalité d'un ensemble de variables continues basé sur la corrélation linéaire entre les variables, utilisable comme intermédiaire de calcul en vue d'analyses ultérieures.

Ici le tableau représenté graphiquement est celui présentant toutes les valeurs prises entre janvier 1993 et décembre 2005 (variables) par chaque méthode de désaisonnalisation (individus). Ainsi, sa représentation graphique permet de regrouper les méthodes dont les résultats sont les plus proches au cours de la période d'observation.



Un premier regroupement net est celui des méthodes combinant l'utilisation du logiciel X12 et une correction de jours ouvrables utilisant le calendrier national. La méthode TD2_X12 est toutefois relativement proche de ce regroupement.

Les méthodes de calendrier national utilisant le logiciel TramoSeats sont elles aussi nettement à part des autres, mais beaucoup plus éloignées de l'origine du graphique, ce qui témoigne d'une spécificité dans la méthode de désaisonnalisation. Ceci a poussé à ne pas choisir ces méthodes tant elles sont éloignées des autres.

La méthode la plus au centre est donc celle qui présente le moins de signes distinctifs par rapport aux autres méthodes et donc le moins de différences avec les méthodes testées. Ici la méthode TD2_X12 bien que plus près du groupement X12 est celle qui est la moins éloignée des autres méthodes TS.

Jusqu'ici, les séries NC2_TS et NC7_TS donnent des résultats assez éloignés de ceux des autres séries. Il restait à vérifier si ces différences persistaient dans l'étude de la cohérence de la désaisonnalisation.

2. Cohérence de la désaisonnalisation avec l'information des séries

a) Cohérence

Il était important que la série désaisonnalisée représente le mieux possible les phénomènes non saisonniers présents dans la série brute. Outre les valeurs de la série désaisonnalisée, l'intérêt se trouvait en termes d'évolution. L'augmentation ou la diminution d'un mois par rapport à l'autre est plus important que le niveau de celles-ci. Il s'agissait donc de recenser pour chacune des séries désaisonnalisées, le nombre de fois où l'information qu'elle véhiculait était contraire aux autres. En effet, il était possible qu'entre deux mois donnés, une méthode présente une évolution positive et l'autre négative. Comme il n'était pas aisé de déterminer quel résultat était juste, il fallait s'attacher plus particulièrement à ce qu'une série désaisonnalisée se différencie le moins possible des autres séries. Il n'était en effet pas concevable, dans le cadre d'une publication nationale, que les résultats soient fortement influencés par la méthode de désaisonnalisation choisie.

Le tableau suivant présente les taux de cohérence entre les signes des taux de croissance de deux séries

	CJO_X11	CJO_X12	NC2_TS	NC2_X12	NC7_TS	NC7_X12	TD2_TS	TD2_X12	TD7_TS	TD7_X12
CJO_TS	78,85	78,85	56,41	78,21	54,49	68,59	75,64	71,79	78,85	72,44
CJO_X11		93,59	58,33	89,10	53,85	73,08	71,15	77,56	67,95	73,08
CJO_X12	93,59		54,49	90,38	50,00	74,36	72,44	78,85	70,51	75,64
NC2_TS	58,33	54,49		56,41	90,38	63,46	61,54	62,82	55,77	54,49
NC2_X12	89,10	90,38	56,41		51,92	80,13	69,23	80,77	67,31	75,00
NC7_TS	53,85	50,00	90,38	51,92		53,85	54,49	53,21	53,85	51,28
NC7_X12	73,08	74,36	63,46	80,13	53,85		72,44	81,41	69,23	74,36
TD2_TS	71,15	72,44	61,54	69,23	54,49	72,44		79,49	86,54	77,56
TD2_X12	77,56	78,85	62,82	80,77	53,21	81,41	79,49		73,72	76,28
TD7_TS	67,95	70,51	55,77	67,31	53,85	69,23	86,54	73,72		83,33
TD7_X12	72,44	73,08	75,64	54,49	75,00	51,28	74,36	77,56	76,28	

Note de lecture : 78.85% des taux de croissance de la série désaisonnalisée CJO_TS sont les mêmes que ceux de la série CJO_X11.

Les méthodes qui présentent le moins de cohérence avec toutes les autres séries désaisonnalisées sont les méthodes : NC2_TS et NC7_TS. Ceci confirme les résultats précédemment obtenus.

Les méthodes TD2 et TD7 du logiciel X12 semblent être celles qui possèdent le plus d'informations en commun avec les autres séries.

b) Statistiques de lissage et de qualité

Il s'agissait ensuite de s'attacher aux caractéristiques des séries décomposées. En d'autres termes, pouvoir juger de la pertinence de décomposition de la série brute par le logiciel. La série se décompose en :

- une série comportant l'information dite « irrégulière »,
- une série comportant la saisonnalité,
- une série comportant les effets de calendrier,
- une série comportant la tendance et le cycle.

Les méthodes de désaisonnalisation et le choix paramètres qui les mettent en œuvre peuvent conduire à des composantes saisonnières plus ou moins stables dans le temps.. Ceci a des conséquences sur les séries désaisonnalisées : plus la saisonnalité est stable, plus la série désaisonnalisée comportera une part d'irrégulier forte et donc, moins elle sera lisse. Or, les utilisateurs attachent une forte importance au fait que la série désaisonnalisée soit « lisse ». Néanmoins, ce critère doit faire l'objet d'une certaine prudence : la composante irrégulière fait partie intégrante de la série désaisonnalisée. Il n'y a donc aucune raison que la série désaisonnalisée soit lisse.

Un certain nombre de mesures ont été proposées dans la littérature pour juger du caractère lisse d'une série.

Dagum ([8]) propose deux mesures de "rugosité" de la série désaisonnalisée.

La première est la norme L_2 de la série différenciée :

$$R_1 = \sum_{t=2}^T (A_t - A_{t-1})^2 = \sum_{t=2}^T (\nabla A_t)^2.$$

La seconde est définie comme la norme L_2 de la différence entre la série désaisonnalisée et la série résultant d'un lissage par une moyenne mobile de Henderson sur 13 termes.

$$R_2 = \sum_{t=1}^T (A_t - H_{13}A_t)^2 = \sum_{t=1}^T [(I - H_{13})A_t]^2.$$

Pfefferman ([14]) suggère une troisième mesure base sur la différence entre la série désaisonnalisée et la tendance-cycle:

$$R_3 = \sum_{t=1}^T (A_t - TC_t)^2.$$

Gomez and Maravall ([11]) préfèrent mesurer le caractère plus ou moins lisse des autres composantes. Pour la composante saisonnière, ils utilisent le critère:

$$Mar(S) = \sum_{t=1}^T [(1 + B + \dots + B^{11})S_t]^2.$$

Pour la tendance cycle, ils utilisent la norme L_2 des différences premières et seconde de la série:

$$Mar1(TC) = \sum_{t=2}^T (\nabla TC_t)^2 \text{ et } Mar2(TC) = \sum_{t=3}^T (\nabla^2 TC_t)^2.$$

Ces différents indicateurs ont été calculés pour chacune des décompositions obtenues, sur les séries complètes et sur les trois dernières années.

X-12-ARIMA propose 12 mesures de qualité, les statistiques 1 à M11 et la statistique Q :

- M1 : contribution de l'irrégulier à la variance totale.
- M2 : contribution de l'irrégulier à la variance de la série brute désaisonnalisée.

- M3 : contribution de l'irrégulier à l'évolution de l'estimation provisoire de la série désaisonnalisée.
- M4 : test sur la présence d'autocorrélation de la composante irrégulière.
- M5 : calcul l'importance des évolutions des composantes tendancielle et irrégulières.
- M6 : mesure l'importance des évolutions annuelles de la composante irrégulière par rapport à celles de la composante saisonnière.
- M7 : permet de voir si la saisonnalité est identifiable ou non par X11.
- M8 et M9 : mesure l'ampleur des variations de la composante saisonnière.
- M10 - M11: mesure l'ampleur des variations de la composante saisonnière sur les dernières années.
- Q : indicateur synthétique de la qualité de la désaisonnalisation, construit comme combinaison linéaire des 11 statistiques (M1-M11)

Ces statistiques ont été normalisées : elles varient entre 0 et 3, mais seules les valeurs en dessous de 1 sont jugées acceptables. Un indicateur synthétique de la qualité de la désaisonnalisation Q est construit comme combinaison linéaire des onze statistiques (M1 à M11). Des définitions précises de ces statistiques sont données dans Ladiray et Quenneville ([12]).

Ces statistiques ont été calculées pour l'ensemble des décompositions obtenues. Il a cependant fallu adapter légèrement ces indicateurs pour pouvoir les calculer aussi pour les séries décomposées avec TramoSeats. Les valeurs alors obtenues pour X-12ARIMA sont un peu différentes (un peu plus grandes) de celles calculées par le logiciel.

Tous les résultats figurent dans le tableau suivant.

Les méthodes NC2 et NC7 de TramoSeats sont ici aussi différentes des autres méthodes. Elles ont un indicateur R1 très fort par rapport aux autres désaisonnalisations ; peut être reste-t-il une part de saisonnalité dans la série corrigée des variations saisonnières.

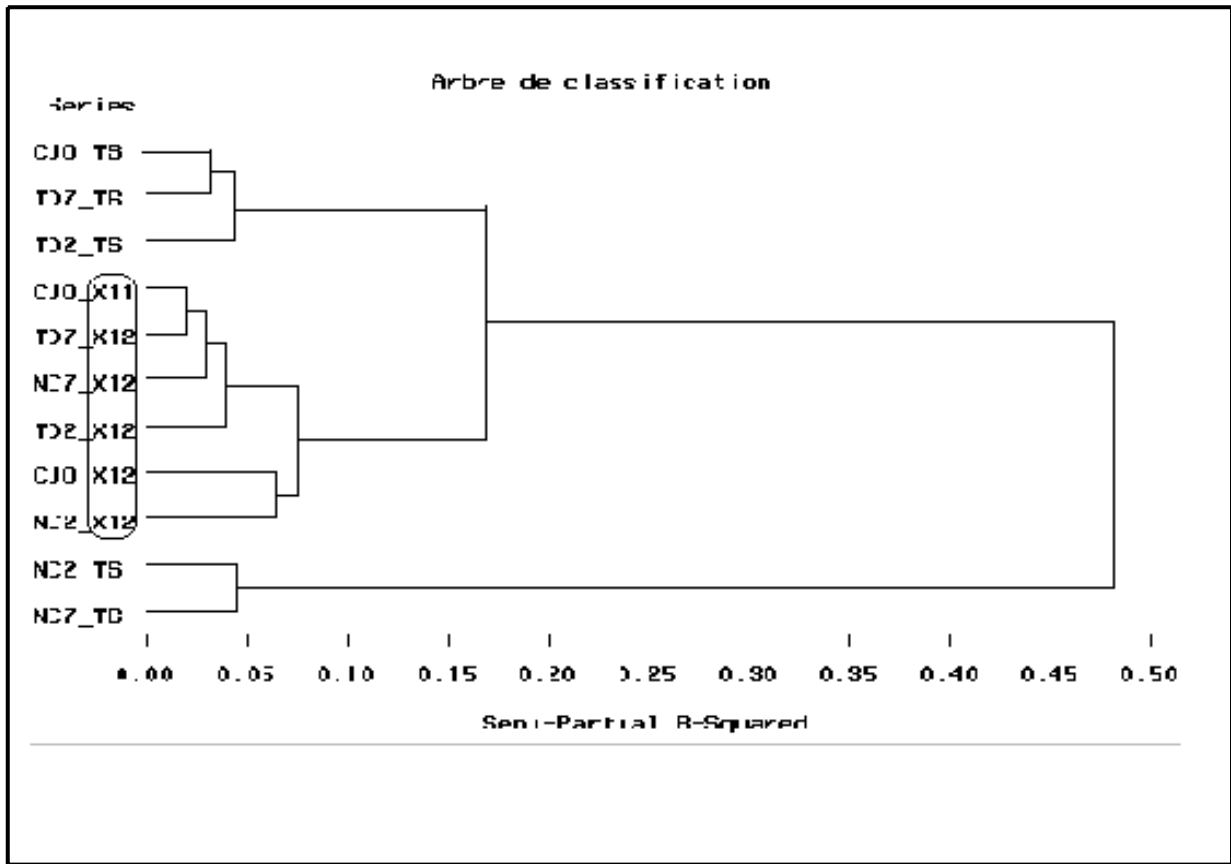
Les séries n'utilisant pas le calendrier national (TD) de TramoSeats ont, au contraire, des séries désaisonnalisées relativement lisses par rapport aux autres séries, ce qui laisse penser que les coefficients saisonniers doivent être peu stables dans le temps.

La statistique Q est un indicateur synthétique de qualité de désaisonnalisation. Il doit être le plus faible possible. C'est plus particulièrement le cas des méthodes CJO_TS, TD2_TS et TD7_TS.

INDICATOR	CJO_TS	CJO_X11	CJO_X12	NC2_TS	NC2_X12	NC7_TS	NC7_X12	TD2_TS	TD2_X12	TD7_TS	TD7_X12
R1 (SA)	358,88	544,91	590,31	1046,76	589,95	964,47	648,42	420,32	729,42	392,02	636,71
R1 (SA), Last 3 years	353,89	595,85	666,36	1087,56	608,54	1026,07	648,49	299,84	822,44	265,40	714,92
R2 (SA)	1,37	2,10	2,26	3,97	2,27	3,63	2,48	1,66	2,76	1,50	2,38
R2 (SA), Last 3 years	1,03	1,85	2,07	3,48	1,92	3,23	2,17	0,79	2,59	0,61	2,21
R3 (SA)	1,93	2,30	2,33	4,05	2,37	3,96	2,58	1,72	2,95	1,67	2,60
R3 (SA), Last 3 years	2,01	2,21	2,14	3,58	2,01	3,97	2,31	0,72	2,72	0,58	2,36
Mar (TC, 1)	98,79	110,58	134,89	116,05	122,55	100,82	115,64	123,93	121,63	110,00	111,56
Mar (TC, 1), Last 3 years	152,10	169,19	224,99	175,08	192,31	149,13	174,56	182,50	201,17	178,99	183,21
Mar (TC, 2)	42,88	37,07	46,19	51,04	40,51	44,76	40,36	58,31	39,73	43,75	35,90
Mar (TC, 2), Last 3 years	44,26	37,63	55,98	47,09	42,14	42,21	37,80	53,56	44,96	44,13	38,39
Mar (S)	0,38	0,37	0,37	0,13	0,52	0,16	0,53	0,34	0,29	0,36	0,32
Mar (S), Last 3 years	0,37	0,36	0,36	0,13	0,66	0,20	0,50	0,37	0,33	0,42	0,35
M1*	0,05	0,27	0,29	1,22	0,30	1,04	0,39	0,07	0,43	0,06	0,36
M2*	0,15	0,21	0,23	0,69	0,24	0,65	0,27	0,14	0,37	0,11	0,27
M3*	0,62	1,84	1,58	4,32	1,72	4,64	2,19	0,56	2,29	0,62	2,25
M4	0,02	0,32	0,17	0,42	0,12	0,57	0,12	0,57	0,02	0,57	0,42
M5	0,53	0,88	0,76	2,07	0,79	2,10	0,99	0,32	1,01	0,34	0,98
M7*	0,15	0,16	0,15	0,15	0,15	0,16	0,17	0,16	0,18	0,14	0,16
M8	3,10	2,92	2,87	1,27	2,63	1,53	2,77	2,74	2,27	2,86	2,76
M9	0,47	0,39	0,35	0,44	0,31	0,41	0,28	0,43	0,34	0,42	0,33
M10	2,85	2,58	2,52	1,33	2,23	1,49	2,55	2,69	2,20	2,72	2,68
M11	1,74	1,76	1,67	0,94	1,63	1,08	1,41	1,63	1,57	1,72	1,69
Q	0,66	0,88	0,81	1,11	0,79	1,13	0,90	0,65	0,88	0,66	0,95

c) Classification Ascendante Hiérarchique des caractéristiques des séries désaisonnalisées

Contrairement à la classification présentée antérieurement, le regroupement des séries se fait ici sur les critères de lissage et de qualité présentés dans le tableau précédent. Ce n'est plus un regroupement de similitude entre les séries désaisonnalisées mais entre les caractéristiques des composantes de la série brute.

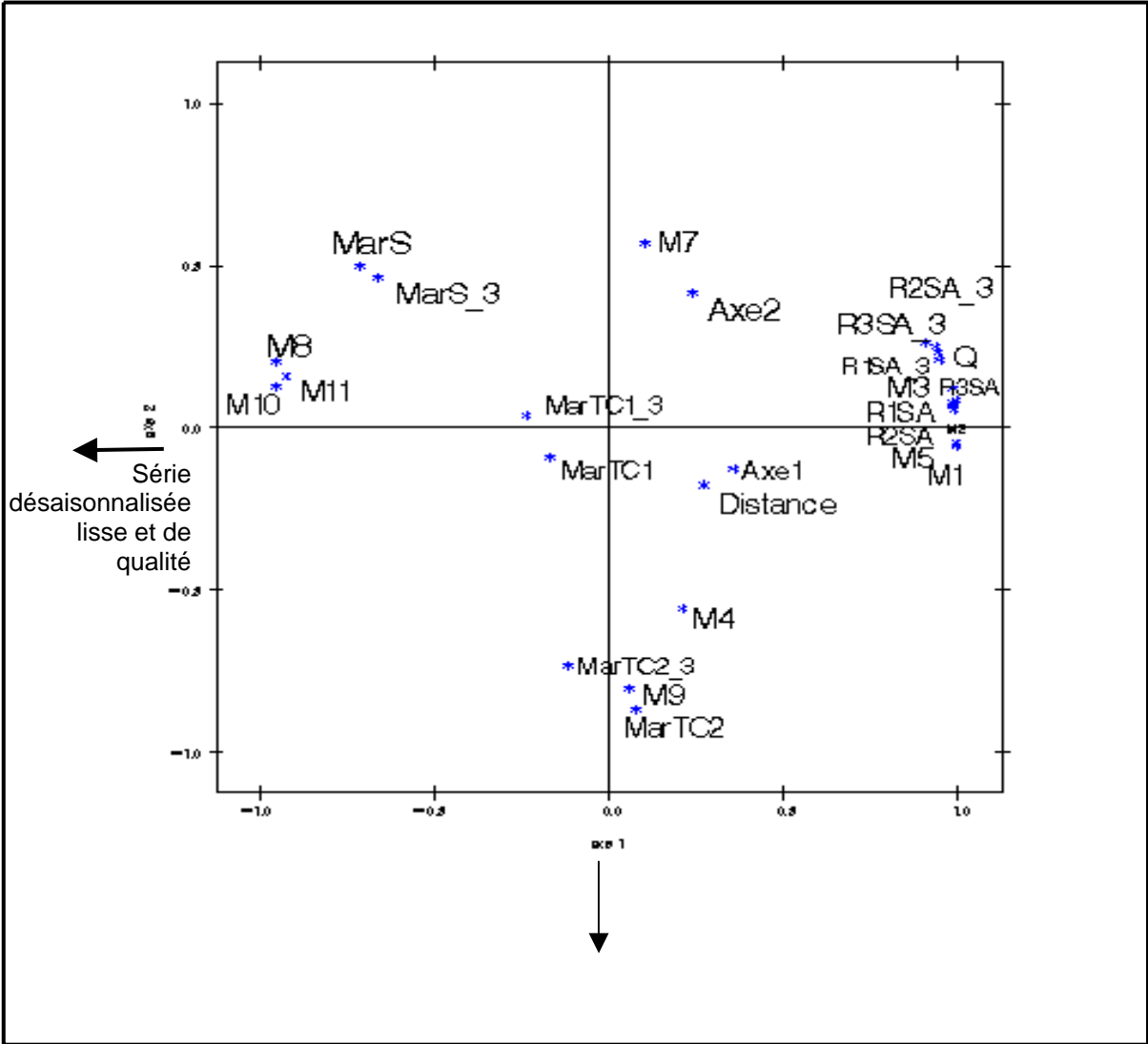


Le bloc X12 présente des similitudes en ce qui concerne les caractéristiques de la décomposition de la série. C'est-à-dire que quelle que soit la méthode de corrections de jours ouvrables utilisée, la décomposition faite par X12 est similaire. En revanche, cela n'est pas le cas pour TramoSeats : les séries corrigées de jours ouvrables tenant compte du calendrier national ne se sont pas décomposées de la même façon que celles n'en tenant pas compte. Cette différence reflète les résultats précédemment observés, à savoir que les séries désaisonnalisées NC2_TS et NC7_TS étaient fortement fluctuantes par rapport aux méthodes TD2_TS et TD7_TS relativement lisses.

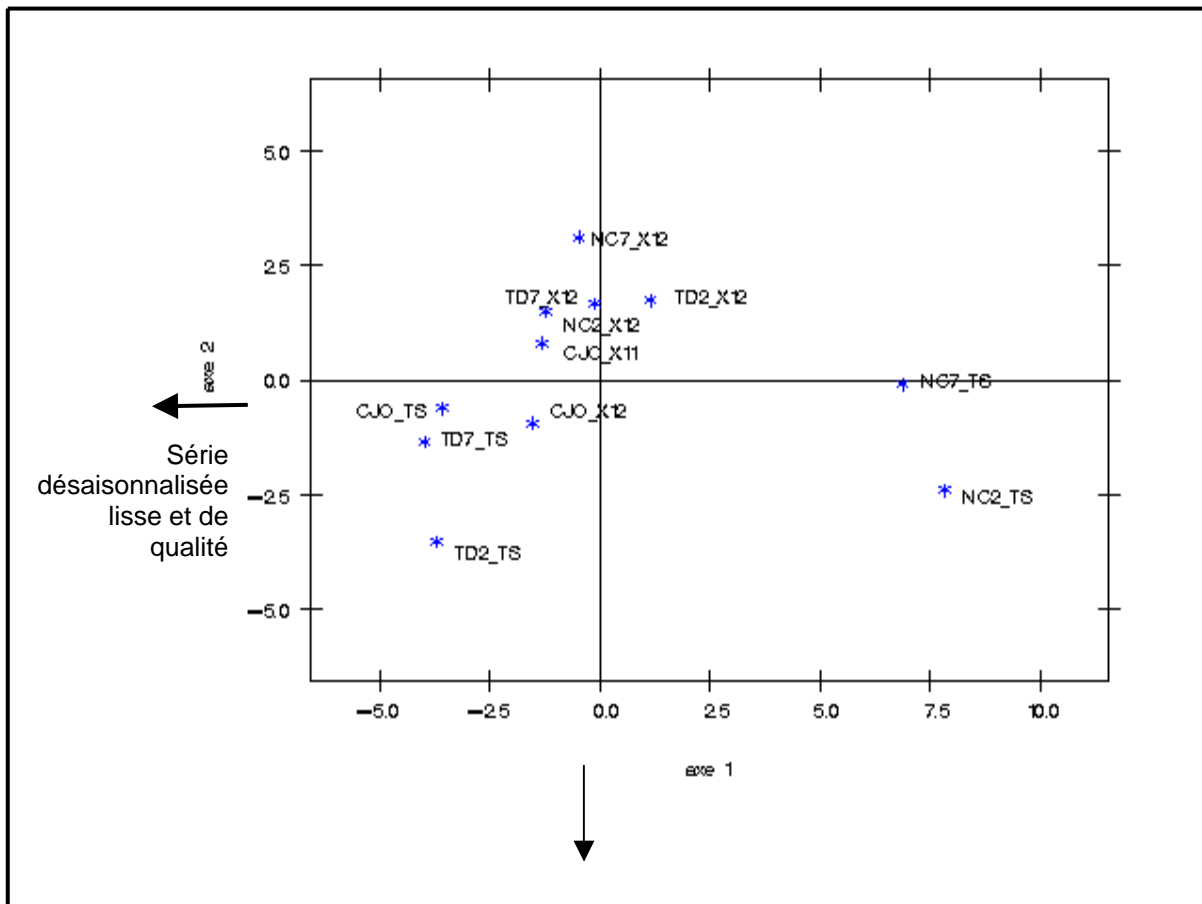
d) Analyse en Composantes Principales des caractéristiques des séries désaisonnalisées

De même, on a réalisé une Analyse en Composantes Principales, en prenant comme distinction entre les séries les statistiques de lissage et de qualité.

Représentations des variables liées aux statistiques de lissage et de qualité :



Représentation des séries en fonctions des statistiques de lissage et de qualité :



Ces deux représentations permettent de mettre en évidence les caractéristiques de chaque série. Par exemple, si l'on souhaite avoir une série désaisonnalisée « lisse », on choisira les séries représentées le plus à gauche sur le graphique. Ici on voit que les désaisonnalisations faites sous TramoSeats, hormis celles utilisant le calendrier national dans sa correction de jours ouvrables, présentent cette particularité.

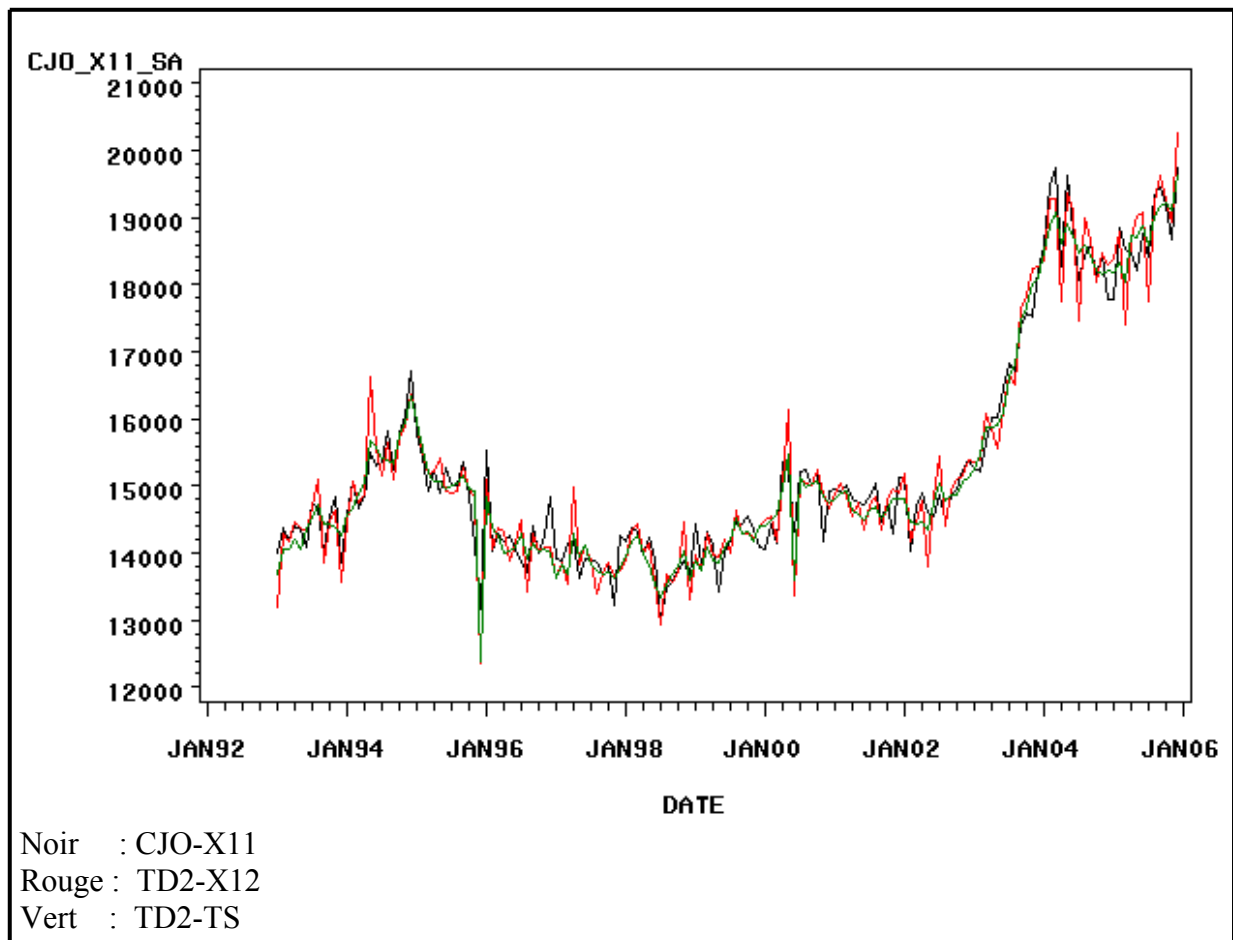
Le groupe X12 est relativement proche du centre ce qui traduit un juste équilibre du lissage.

	Qualité de désaisonnalisation		Cohérence de la désaisonnalisation	
	Décomposition de la série brute	ACP sur la qualité de désaisonnalisation	Similitude entre les séries désaisonnalisées	ACP sur le lissage et la qualité de la série CVS
CJO_X11				
CJO_X12				
CJO_TS		+		+
NC2_X12				
NC2_TS		—	—	—
NC7_X12	—			
NC7_TS		—	—	—
TD2_X12	—	+	+	
TD2_TS				+
TD7_X12	—		+	
TD7_TS				

Le tableau ci-dessus récapitule les résultats obtenus. Le symbole ‘+’ indique la ou les séries qui obtiennent le meilleur résultat dans le test indiqué ; au contraire, le symbole ‘-’ indique de mauvais résultats.

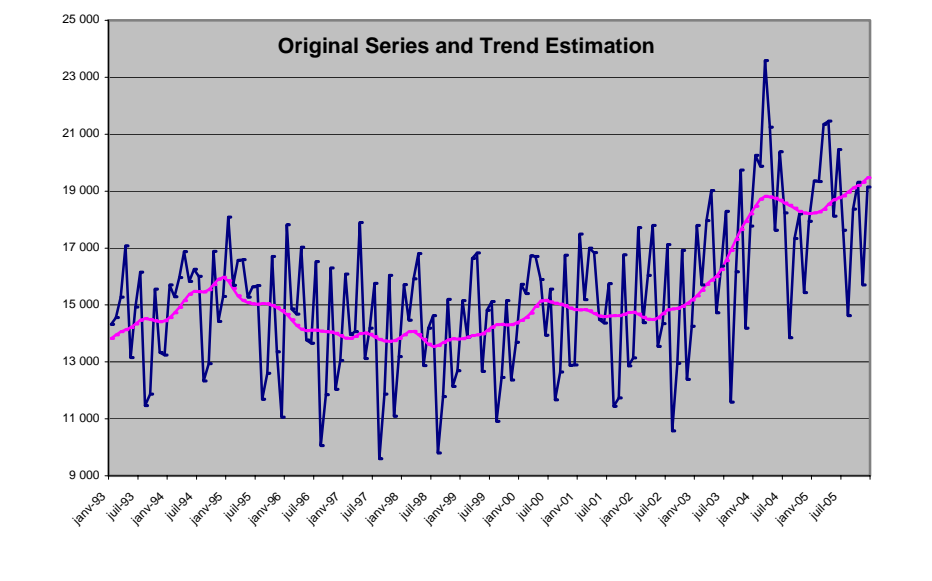
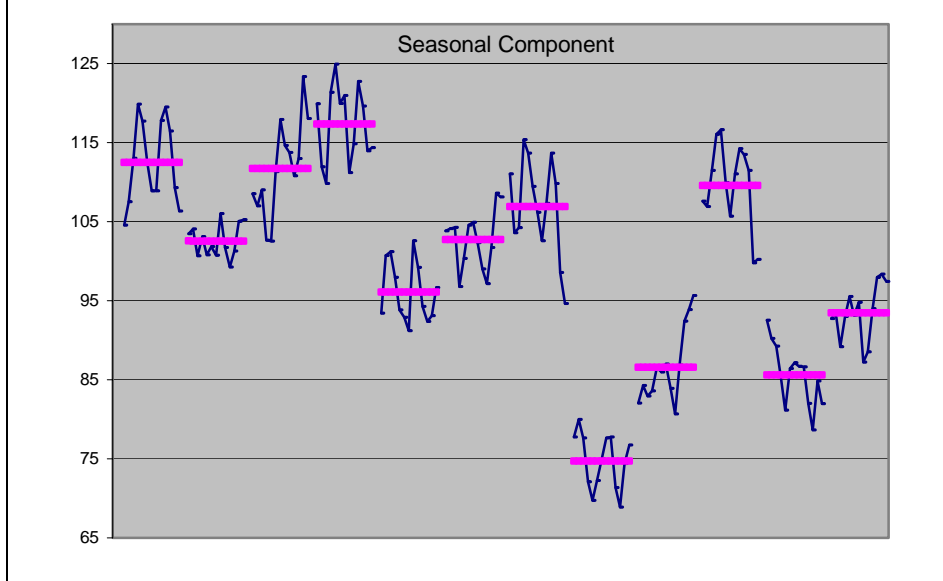
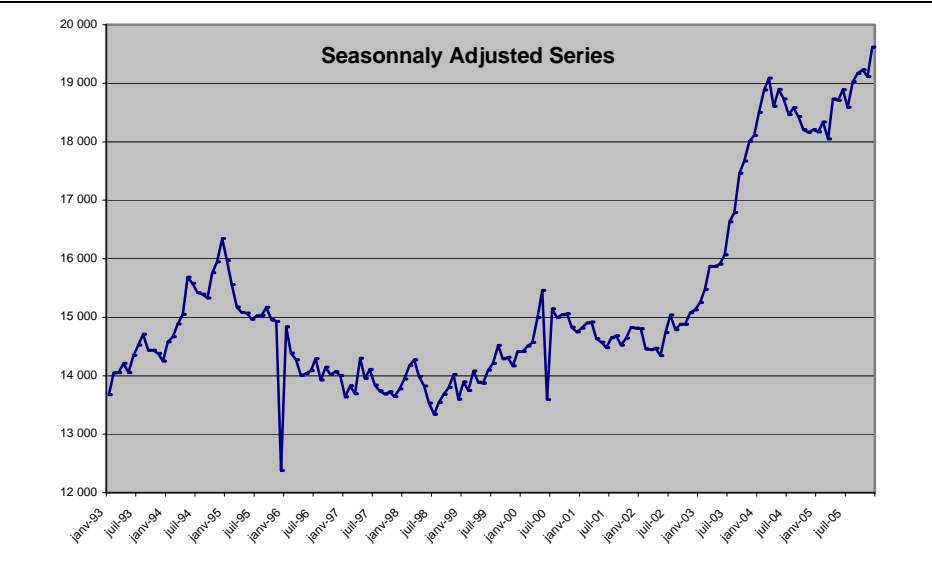
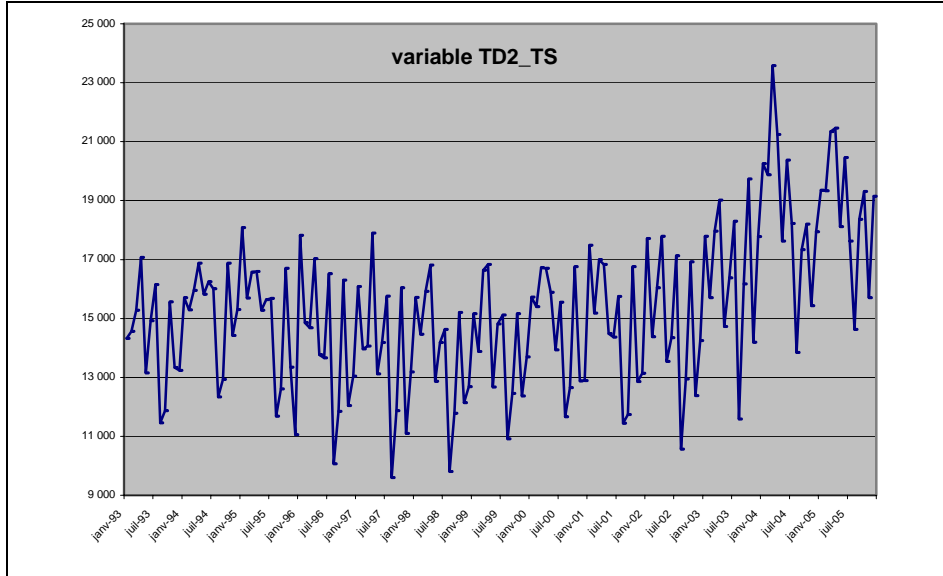
Au vu de ces analyses, les méthodes TD2_X12 et CJO_TS présentaient de bons résultats. Cependant le choix final se portait davantage sur la méthode TD2_X12, le logiciel X12 permettant de garantir une certaine continuité avec la méthode utilisée jusqu’ici.

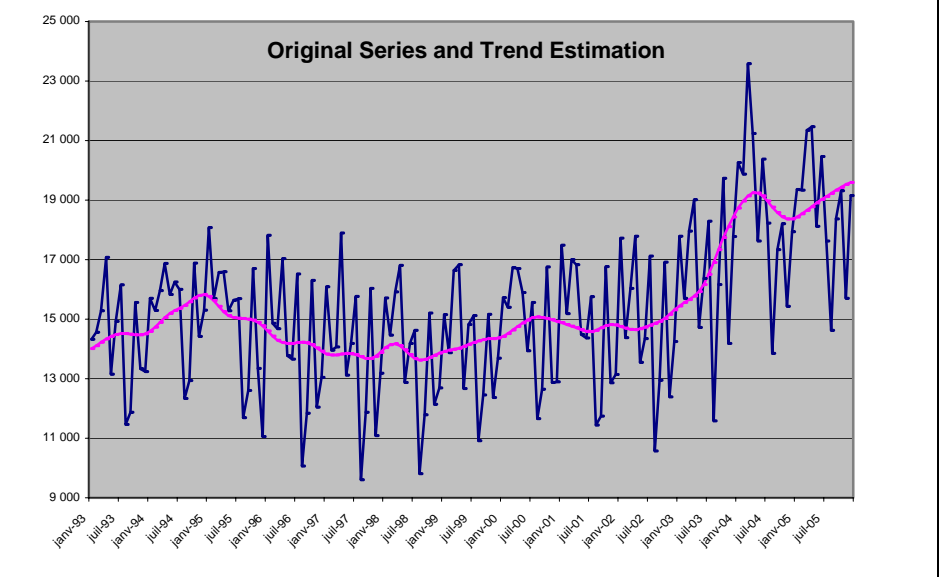
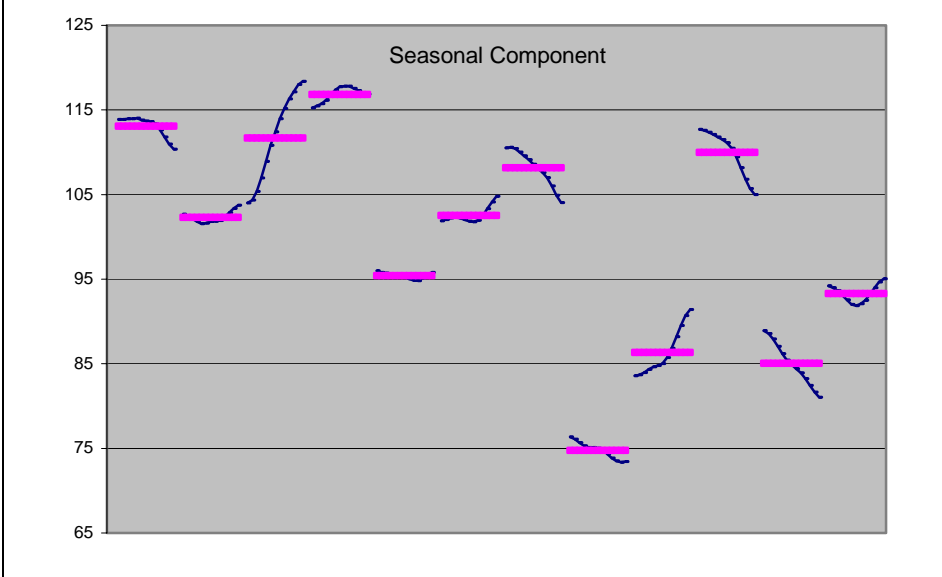
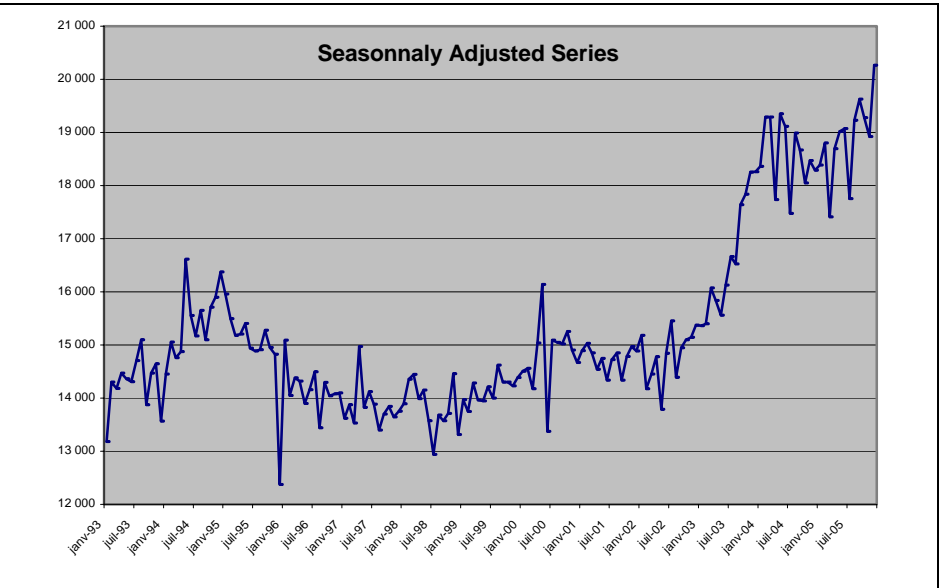
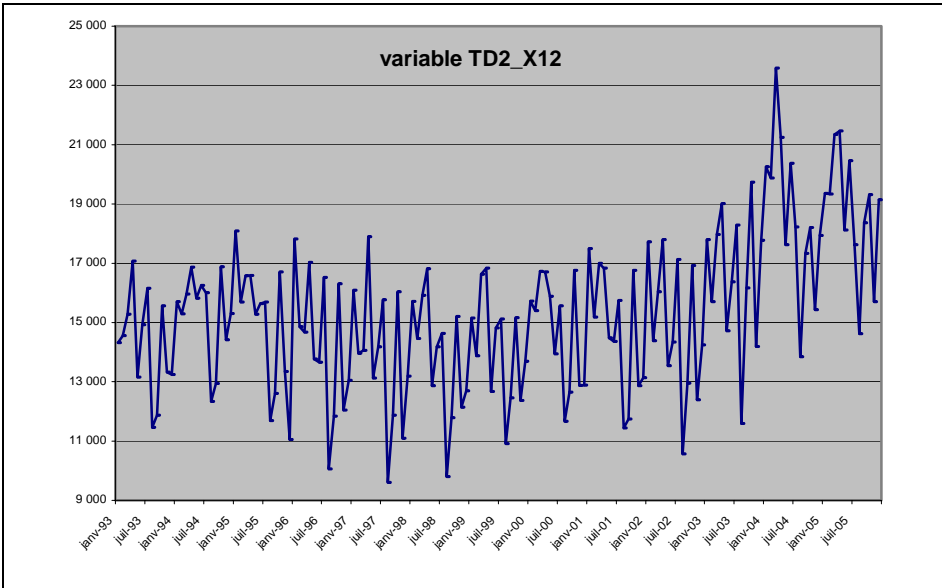
e) Comparaison des 2 « meilleures » séries CVS-CJO : TD2-X12 et TD2-TS avec la série utilisée



Ce que l'on avait constaté grâce aux statistiques de lissage se confirme : la série désaisonnalisée par TramoSeats est plus lisse que celle faite par X12.

f) Décomposition précise des 2 séries TD2 X12 et TD2 TS





Ces graphiques permettent de mettre en évidence toutes les remarques précédentes. La série désaisonnalisée par TramoSeats est beaucoup plus lisse que celle désaisonnalisée par X12, en revanche les coefficients saisonniers sont eux très fluctuants, alors que ceux de X12 sont relativement stables.

On constate que les coefficients saisonniers ont beaucoup évolué durant ces treize années d'observation. Le mois de mars est celui qui a connu la plus grande évolution.

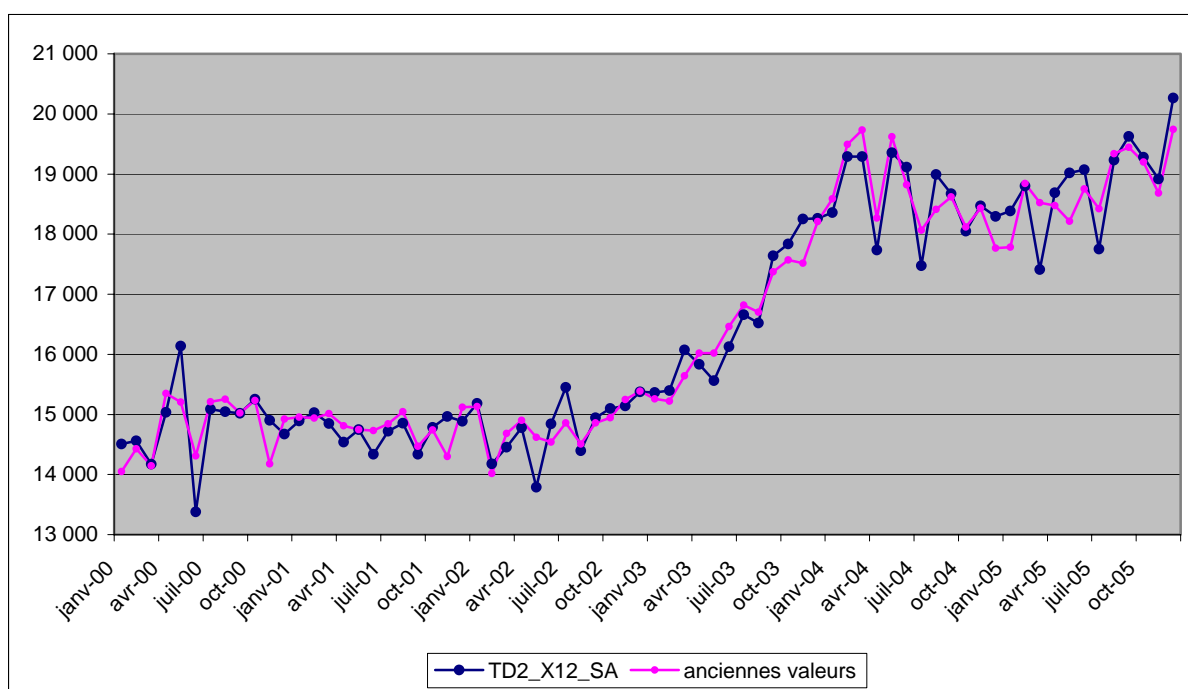
Conclusion

Il s'agissait de tester sur une série qui avait subi de nombreuses variations au cours des dernières années dues principalement à l'entrée en vigueur de nouvelles lois, différentes méthodes de désaisonnalisation, et de retenir celle qui lui était la plus appropriée. Les tests ont donc porté d'une part sur la correction de jours ouvrables (choix d'utilisation d'une méthode économétrique ou non et d'un calendrier national ou non), et d'autre part, sur la correction de variations saisonnières automatisée (choix entre les logiciels de désaisonnalisation disponibles X12 de SAS et TramoSeats).

Le logiciel X12 était, d'un point de vue statistique, plus satisfaisant pour privilégier la stabilité des coefficients saisonniers au cours du temps. Quant à la méthode économétrique, le choix se portait sur l'utilisation de deux ou sept régresseurs pour le calcul des coefficients de correction de jours ouvrables. Les résultats avec deux coefficients semblaient plus appropriés. La question se portait donc sur le choix du calendrier. Compte tenu des contraintes imposées - l'automatisation, la facilité de traitement, une désaisonnalisation proche des autres méthodes - ainsi que des préférences - choix un logiciel SAS pour un meilleur savoir faire -, **la méthode TD2_X12 était celle retenue**. Elle était la plus proche de toutes les autres méthodes testées en ayant notamment de bons taux de similitude avec la majorité des séries testées.

Le graphique suivant confronte la courbe publiée en janvier 2006 et celle qui aurait été publiée avec la méthode TD2_X12 selon le concept de créations pures en vigueur jusqu'à la fin de l'année 2006.

77,6% de l'information était la même pour les deux séries, soit trois évolutions sur quatre allaient dans le même sens. Le changement de méthode n'a donc pas été sans conséquences sur les publications mais la qualité en a été meilleure. En revanche, la série obtenue comporte encore plus d'irrégularité que l'ancienne.



La méthode économétrique de correction de jours ouvrables mise en place permet d'attribuer un poids différent à chaque jour de la semaine, mais il reste constant sur toute la période d'observation. Ce constat semble ne pas refléter complètement la réalité mais pour répondre à ce problème, il faudrait mettre en place une méthode beaucoup plus complexe faisant appel à une désaisonnalisation journalière de la série. Elle aurait toutefois pu être envisagée car les séries de créations d'entreprises ont la particularité d'être des séries journalières. Cependant, cette méthode serait difficilement applicable dans le cadre d'une publication mensuelle.

Bibliographie

- [1] Bell, W. R., Hillmer, S. C. (1983), Modeling Time series with Calendar Variation, *Journal of the American Statistical Association*, 383, 78, 526-534.
- [2] Bell, W. R., Hillmer, S. C. (1984), Issues Involved with the Seasonal Adjustment of Economic Time Series, *Journal of Business and Economic Statistics*, 4, 2, 291-320.
- [3] Cleveland, W. P. (1987), Calendar Adjustment of Time Series, Discussion Paper, Special Studies Section, Board of Governors of the Federal Reserve System, 198.
- [4] Cleveland, W. P., Grupe, M. R. (1983), Modeling time series when calendar effects are present, *Applied Time Series Analysis of Economic Data*, Zellner, A. (editor), U.S. Department of Commerce, U.S. Bureau of the Census, Washington D.C., 57-67.
- [5] Cleveland, W. S. (1983), Seasonal and Calendar Adjustment, *Handbook of Statistics Volume 3: Time Series in the Frequency Domain*, Brillinger D. R. and Krishnaiah P. R. (eds), North-Holland, 39-72.
- [6] Cleveland, W. S., Devlin, S. J. (1980), Calendar effects in monthly time series: Detection by spectrum analysis and graphical methods, *Journal of the American Statistical Association*, 371, 75, 487-496.
- [7] Cleveland, W. S., Devlin, S. J. (1982), Calendar effects in monthly time series: Modeling and adjustment, *Journal of the American Statistical Association*, 379, 77, 520-528.
- [8] Dagum, E.B. (1979), On the Seasonal Adjustment of economic Time Series Aggregates: A Case Study of the Unemployment Rate, *Counting the Labor Force, National Commission on Employment and Unemployment Statistics, Appendix, 2*, 317-344, Washington.
- [9] Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C. and Chen, B. (1998), New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program, *Journal of Business and Economic Statistics*, 16, 127-177.
- [10] Fournier, J-M., Ladiray, D. (1995), Les effets de calendrier dans l'analyse conjoncturelle de la production, *Economie et Statistique*, 285-286, 115-126.
- [11] Gomez, V. and Maravall, A. (1999), *Two Application of TRAMO-SEATS: Examples 2 and 4*, Bundesbank Seminar, Unpublished manuscript.
- [12] Ladiray, D., Quenneville, B. (2001), *Seasonal Adjustment with the X-11 Method*, Lecture Notes in Statistics, Springer-Verlag, New York.
- [13] Maillard, V. (1994), Théorie et pratique de la correction des effets de jours ouvrables, document de travail, Direction des Etudes et Synthèses Economiques, INSEE, Paris, G9405.
- [14] Pfefferman, D., Salama, E. and Ben-Tuvia, S. (1984), *On the Aggregation of Series: A New Look at an Old Problem*, Working paper, Bureau of Statistics, Jerusalem, Israel.
- [15] U.S. Census Bureau (2001), *X-12-ARIMA Reference Manual, Version 0.2.8*, Washington DC, U.S. Bureau of the Census.
- [16] Young, A. H. (1965), Estimating trading-day variations in monthly economic series, Technical Paper, U.S. Department of Commerce, U.S. Bureau of the Census, Washington D.C.