

# Un cadre de travail alternatif pour la correction des variations saisonnières de séries temporelles univariées et multivariées

*Stéphane GREGOIR<sup>1</sup>*

*CREST, INSEE*

## Introduction

La demande de séries corrigées des variations saisonnières résulte du fait qu'il est difficile d'analyser les inflexions de moyen-long terme des séries qui présentent des mouvements réguliers en termes de fréquences infra-annuelles. Cependant les statisticiens et les économistes ne sont pas d'accord pour définir ce que doit être le modèle ou la famille de modèles qui décrit ces sources de régularités. En fait, il y a très peu d'études qui modélisent explicitement les composantes saisonnières sur la base d'un modèle économique et des interactions entre les institutions, les influences climatiques et biologiques et les pratiques sociales et religieuses des sociétés.

Dans ces circonstances, les statisticiens ont privilégié une approche descriptive en faisant le constat pragmatique que la difficulté de lecture des séries brutes résultait de la présence de composantes saisonnières persistantes qui du fait de la répétition de profils voisins et de grande amplitude au cours des années cachent les mouvements de moyen-long terme. L'objet de la correction des variations saisonnières correspond celui d'éliminer ces composantes saisonnières en introduisant les plus faibles distorsions possibles aux autres fréquences. L'exemple le plus simple de composante saisonnière persistante à la fréquence  $\omega \in ]0, \pi]$  est une courbe sinusoïdale égale à la combinaison linéaire des fonctions déterministes  $\sin(\omega t)$  et  $\cos(\omega t)$ . Il est aussi possible de recourir à des processus stochastiques qui oscillent à la fréquence recherchée mais dont l'amplitude peut varier aléatoirement au cours du temps en combinant des marches aléatoires complexes à la fréquence  $\omega \in ]0, \pi]$  et leur conjuguée. Le choix de ces formes de processus est un ingrédient important. Le modèle usuellement retenu suppose que les séries étudiées peuvent se décomposer en la somme de processus non corrélés, l'un d'entre eux générant le profil saisonnier. Ceci suppose que dans cette approche descriptive, les sources de la saisonnalité n'interagissent pas avec celles des mouvements de moyen et long terme. Une limite de cette approche est que suivant le cadre d'analyse utilisé, cette décomposition n'est pas toujours possible. Nous proposons de suivre un chemin différent. Nous gardons l'idée d'une décomposition additive mais nous supposons que les innovations de chacune des composantes persistantes sont parfaitement corrélées, ce qui suppose que de mêmes chocs descriptifs peuvent être à l'origine des évolutions de moyen et long terme autant que des profils saisonniers. Cette décomposition est naturelle et existe pour presque tous les processus intégrés simultanément à différentes fréquences ([27], [31]). Différents cadres peuvent être utilisés pour estimer ces décompositions.

---

<sup>1</sup> Je remercie D. Ladiray, G. Laroque, T. Proietti pour leurs commentaires. Ce texte reprend en partie le texte présenté à la Conférence 2004 de la Société Italienne de Statistique à Bari. Toutes les erreurs ou imprécisions qui demeurent me sont imputables.

Nous présentons ici dans un cadre linéaire une méthode pour les estimer lorsque le processus satisfait une représentation SARIMA. L'estimation des composantes saisonnières persistantes repose sur le calcul des prévisions de long-terme du processus stationnarisé démodulé à chacune des fréquences saisonnières d'intérêt.

La demande de séries corrigées des variations saisonnières résulte de besoins variés. Lorsqu'elle provient de praticiens tels que les analystes de court-terme dans les banques ou les administrations, ou des décideurs politiques, elle correspond à une demande simultanée de plusieurs séries caractérisant une situation économique. Ce sont généralement le PIB, la consommation des ménages, l'investissement, l'emploi, les crédits, les salaires,...Elles sont utilisées pour comparer la situation de différents pays ou analyser les mécanismes économiques à l'œuvre en étudiant leurs mouvements joints. Sur la base de cette analyse, de nouvelles stratégies d'investissement ou des actions politiques pourront être mises en œuvre. Vis-à-vis de cette pratique, il semble que les méthodes usuelles de désaisonnalisation ne satisfassent pas les besoins des utilisateurs dans la mesure où les traitements de désaisonnalisation sont univariés. Ces traitements peuvent donc ne pas respecter les liens qui existent entre les séries. C'est une question qui a été l'objet de quelques travaux de recherche dans les années soixante-dix ([47], [23], [46] et [57]), puis est tombée dans l'oubli. Du point de vue pratique de la diffusion de séries corrigées des variations saisonnières, il n'est pas possible de fournir aux différents utilisateurs les séries désaisonnalisées suivant différents ensembles d'information. Il pourrait néanmoins être intéressant de leur donner les séries brutes de manière à leur permettre de faire le traitement adapté à leur besoin ou de diffuser des ensembles standard de séries corrigées des variations saisonnières de façon jointe. Le traitement que nous introduisons ici autorise le traitement joint de quelques séries temporelles.

Les logiciels de traitement de la saisonnalité offrent de plus en plus d'options à leurs utilisateurs pour permettre d'adapter les modèles descriptifs utilisés à un monde économique toujours changeant ou à l'occurrence de points extrêmes. Ce grand nombre d'options peut parfois introduire des instabilités dans les estimations des séries corrigées, en particulier pour les extrémités de séries qui sont d'un point de vue conjoncturel les plus regardées mais qui ne sont pas usuellement traitées d'un point de vue statistique comme les observations plus anciennes. Le cadre linéaire introduit ici utilise le même traitement pour tous les points de l'échantillon et peut facilement être adapté pour permettre l'usage de méthodes d'estimation robuste afin de mieux contrôler l'impact des points aberrants. Il permet aussi de calculer un intervalle de confiance autour de la série corrigée (cet intervalle est associé à l'incertitude dans l'estimation de la représentation satisfaite par le processus et non pas au choix de la définition de la composante saisonnière). Une question importante de cette approche est celle de pouvoir sélectionner le modèle approprié pour décrire les composantes saisonnières persistantes. D'un point de vue statistique, cela correspond à la possibilité de pouvoir tester efficacement la présence de racine unitaire saisonnière pour décrire des profils saisonniers éventuellement fluctuants. Des travaux récents (voir [33]) développent ce type de tests.

Une limite de ce cadre d'analyse est liée au fait que le traitement de correction correspond à l'usage d'un filtre propre à la série. Ce filtre est asymétrique ce qui peut pour certain modèle induire un décalage de phase préjudiciable pour l'analyste conjoncturel. Par ailleurs, il peut aussi présenter des gains plus grands que 1 pour certaines fréquences. Une étude est donc nécessaire pour définir les familles de modèles auxquelles il est prudent de se restreindre ou voir comment il est possible de mélanger l'approche usuelle avec celle introduite ici en particulier pour ce qui concerne les observations de fin d'échantillon dans la mesure où ce nouveau cadre d'analyse permet de construire des filtres asymétriques adaptés au processus de génération des données étudiées. Cependant dans les exemples que nous présentons dans la sixième partie, nous n'observons pas de décalage dans la chronologie des mouvements présents dans les séries (en particulier en ce qui concerne les points de retournement) lorsque nous comparons les séries corrigées des variations saisonnières obtenues par l'approche introduite ici à celles diffusées par les instituts nationaux de statistique. Ceci fait écho au commentaire de Grether et Nerlove dans leur article de 1970 [34] : "It has never been denied, and, indeed, is repeatedly emphasized by Rosenblatt in his own work that the effects noted in the frequency domain are significant only to

the degree to which these same effects, translated into the behavior of the adjusted series over time, affect the interpretation of the movements of that series”.

La section suivante résume brièvement les principes des procédures de correction des variations saisonnières. Quelques résultats d’algèbre sont donnés dans la troisième section, ils permettent de caractériser et de construire la composante persistante d’un processus intégré à une fréquence saisonnière. La quatrième section est consacrée à la décomposition de tout processus univarié ou multivarié intégré à différentes fréquences saisonnières en la somme de processus intégrés à chaque fréquence saisonnière et d’une composante transitoire. L’exemple du modèle « Airline » est analysé en détail. La cinquième partie traite de la mise en œuvre de cette méthode pour des processus SARIMA univariés et des processus multivariés cointégrés à fréquence saisonnière. Des illustrations empiriques sont données dans les deux dernières sections. Une courte conclusion clôt l’article.

## Saisonnalité

En l’absence de consensus sur les modèles structurels qui permettraient aux statisticiens d’éliminer les composantes saisonnières pertinentes<sup>2</sup>, les procédures de correction des variations saisonnières adoptent généralement une approche descriptive. L’hypothèse retenue est celle d’une description adaptée de la série temporelle étudiée  $\{x_t\}$  en une somme de composantes inobservables non corrélées, l’une d’entre elles portant le caractère saisonnier de la série :

$$x_t = tc_t + s_t + i_t$$

où  $\{tc_t\}$  est un processus qui décrit les mouvements de la tendance et des cycles longs,  $\{s_t\}$  décrit les mouvements saisonniers et  $\{i_t\}$  est une composante de bruit sans régularité cyclique marquée. La série corrigée des variations saisonnières est égale à la série initiale à laquelle est soustraite une estimation de la composante saisonnière. La composante saisonnière est identifiée principalement par l’hypothèse de non corrélation à tous les décalages de dates possibles avec les autres composantes. La plupart des méthodes utilisent une modélisation paramétrique de ces composantes, ce qui autorise l’usage des règles d’inférence classique et l’emploi éventuel de procédure optimale d’extraction de signal. Elles diffèrent cependant en leur choix et usages de ces modèles.

Deux classes principales de modèles sont utilisées. D’une part, les modèles de séries temporelles structurelles postulent des composantes paramétrées a priori (voir entre autres [19], [35], [36] et dans un cadre Bayésien, [1], [42]). Sous hypothèse de normalité et à l’aide d’un filtre de Kalman, chaque composante peut être estimée. D’autre part, la famille des modèles ARIMA est utilisée dans deux buts particuliers. Dans une première approche, différents filtres linéaires symétriques de longueur finie sont utilisés pour éliminer les profils saisonniers dans les données. Pour permettre leur emploi en début et en fin d’échantillon, le modèle est utilisé pour estimer les valeurs futures en fin d’échantillon et passées en début d’échantillon ([14],[15]). Dans une deuxième approche, le modèle est utilisé pour dériver une décomposition de la série observée en la somme des trois composantes. En pratique, des modèles SARIMA sont utilisés pour le modèle de la série et chacune des composantes qui le composent, mais avec des contraintes sur la position des racines des polynôme autorégressifs associés et un filtre d’extraction du signal (dit de Wiener-Kolmogorov) en est déduit qui permet d’estimer chacune des composantes moyennant quelques restrictions identifiantes (voir entre autres, [6], [9], [38], [26]).

Ceci appelle quelques commentaires. D’une part, la décomposition évoquée ci-dessus n’est pas toujours possible comme l’ont montré [38]. D’autre part, les procédures de désaisonnalisation usuelles sont en plusieurs étapes et offrent à l’utilisateur une multitude d’options qui peuvent permettre de traiter des ruptures ou des points aberrants mais peuvent aussi rendre difficile

---

<sup>2</sup> Voir [24] pour une illustration de l’inefficacité des procédures usuelles dans un cadre structurel complètement spécifié.

l'analyse des mouvements joints de plusieurs séries, les choix faits pour chaque série n'étant pas nécessairement cohérents. Ceci peut être particulièrement sensible pour les extrémités des échantillons qui sont de fait les points les plus analysés et commentés en pratique. Ces observations corrigées des variations saisonnières sont malheureusement les moins fiables pour au moins deux raisons : elles sont sujettes aux erreurs de mesure liée à leur précocité<sup>3</sup> et elles sont calculées sur la base soit d'un filtre asymétrique différent de celui utilisé pour les points au milieu de l'échantillon, soit d'un filtre symétrique appliqué à des prévisions. Elles sont donc sujettes à révisions, d'autant plus importantes qu'un point aberrant est présent dans leur voisinage<sup>4</sup>. En pratique, les points aberrants sont identifiés dans le cadre du modèle paramétrique qui a été sélectionné. Leur identification est donc contingente à l'échantillon d'observations disponibles et n'est pas nécessairement stable dans le temps. Un point détecté comme aberrant dans un échantillon particulier peut apparaître comme une observation régulière dans un échantillon plus grand. Puisque l'intérêt est dans l'analyse jointe de plusieurs séries temporelles et que la pratique est de les traiter séparément, l'introduction d'un grand nombre d'option sur les détections des points extrêmes ou sur le choix du filtre peut induire des différences entre les séries dues à différents choix d'option et, lorsque les séries doivent être agrégées, peut affecter le comportement de la série agrégées corrigées de variations saisonnières. [50] et [55] recommandent l'usage d'un même filtre linéaire pour toutes les séries apparaissant dans une régression multiple.

Un cadre de travail idéal devrait fournir entre autres une décomposition toujours possible, des séries corrigées aussi proches que possible des séries brutes à toutes les fréquences autres que les fréquences saisonnières, avec un décalage des inflexions aussi faible que possible et la possibilité d'avoir recours à des méthode d'estimation robuste de façon à être moins dépendant de la définition et de la caractérisation des points aberrants.

Le cadre linéaire que nous introduisons ci-après satisfait quelques unes de ces propriétés. De plus, il permet de calculer des intervalles de confiance autour des séries corrigées et de travailler simultanément sur plusieurs séries. Ses limites principales résident dans le fait que le filtre implicite qu'il utilise est asymétrique et peut introduire un décalage de phase.

## Persistance à une fréquence saisonnière

### Algèbre d'opérateurs d'intégration

Nous considérons des processus intégrés à différentes fréquences. Un processus intégré d'ordre 1 à la fréquence  $\omega$  a des trajectoires qui oscillent à cette fréquence mais dont l'amplitude fluctue au cours du temps, elles peuvent donc être vues comme une extension des fonctions sinusoidales qui sont elles d'amplitude fixe. Ces processus intégrés sont tels que l'application d'opérateurs linéaires simples les rend stationnaires du second ordre, processus usuellement analysés à l'aide de leur fonction d'autocovariance ou de leur spectre. Nous introduisons dans un premier temps quelques notations et rappelons des définitions bien connues. Dans la suite, les opérateurs de différence première réels et complexes aux différentes fréquences sont notés comme suit :

$$\Delta_{\omega}(B) = \begin{cases} (1 - e^{-i\omega} B) = \delta_{\omega}(B) & \omega \in \{0, \pi\} \\ (1 - 2 \cos \omega B + B^2) = (1 - e^{-i\omega} B)(1 - e^{i\omega} B) = \delta_{\omega}(B) \delta_{-\omega}(B) & \omega \in ]0, \pi[ \end{cases}$$

où  $B$  est l'opérateur retard. Quand il n'y aura pas d'ambiguïté, nous remplacerons  $\Delta_{\omega}(B)$  et  $\delta_{\omega}(B)$  par  $\Delta_{\omega}$  et  $\delta_{\omega}$ .

<sup>3</sup> toutes les informations nécessaires à leur élaboration ne sont pas disponibles, et l'on doit estimer les informations manquantes.

<sup>4</sup> Lorsque les points aberrants ont une cause bien définie, une analyse dans l'esprit de [7] semble pertinente mais est de fait peu employée du fait de son coût.

**Définition:** Un processus stochastique réel purement non déterministe  $\{x_t\}$  est dit être intégré d'ordre  $h$ ,  $h$  entier, à la fréquence  $\omega \in ]0, \pi[$ , ou  $I_\omega(h)$ , si  $\Delta_\omega(B)^h x_t$  est un processus stationnaire du second ordre tel que le polynôme associé à sa représentation de Wold  $c(B)$  satisfait la condition  $c(e^{i\omega}) \neq 0$  (en conséquence lorsque  $\omega \in ]0, \pi[$ ,  $c(e^{-i\omega}) \neq 0$ ).

**Remarque:** La condition de non nullité du polynôme de la représentation de Wold en  $e^{i\omega}$  peut aussi être présentée comme une condition de densité spectrale strictement positive en  $\omega$ . Elle correspond au fait que le processus n'a pas été sur-différencié.

**Remarque :** Un processus complexe est intégré d'ordre  $h$  à la fréquence  $\omega \in ]0, \pi[$ , si l'application de l'opérateur  $\delta_\omega^h$  donne un processus complexe stationnaire du second ordre de densité spectrale strictement positive en  $\omega$ .

En pratique, les processus peuvent être intégrés simultanément à différentes fréquences. Dans ce cas, le produit des opérateurs de différence première élevés à la puissance appropriée doit être appliqué pour obtenir un processus stationnaire du second ordre. A la suite de [31], nous introduisons un ensemble d'opérateurs d'intégration qui jouent le rôle d'opérateur inverse des opérateurs de différence première à chaque fréquence. Ils sont particulièrement bien adaptés pour introduire la décomposition que nous proposons d'utiliser ici.

**Définition :** L'opérateur intégral  $S_\omega$  associe à toute série réel  $\{y_t, \dots, -1, 0, 1, \dots\}$  la suite

$$S_\omega y \text{ définie par } S_\omega y_t = \begin{cases} \sum_{\tau=1}^t e^{-i\omega(t-\tau)} y_\tau & t \geq 1 \\ 0 & t = 0 \\ -\sum_{\tau=t+1}^0 e^{-i\omega(t+1-\tau)} y_\tau & t < 0 \end{cases}$$

Pour comprendre le rôle joué par ces opérateurs d'intégration, nous pouvons appliquer l'un d'eux à un bruit blanc réel  $\{\varepsilon_t\}$ . Cette opération peut être décomposée en trois étapes. Premièrement, nous construisons le bruit blanc complexe  $\{e^{i\omega\tau} \varepsilon_t\}$ , puis nous construisons le processus complexe intégré d'ordre 1 à fréquence 0  $\left\{ \sum_{\tau=1}^t e^{i\omega\tau} \varepsilon_\tau \right\}$  pour les  $t > 1$  (et l'opération appropriée pour les indices négatifs ou nul), enfin nous effectuons une opération de démodulation à la fréquence  $\omega$  telle que décrite par [29] (la multiplication par  $e^{-i\omega}$  qui déplace le pic de la densité spectrale de 0 en  $-\omega$ ), ce qui donne un processus complexe intégré d'ordre 1 à la fréquence  $-\omega$ . Lorsque l'on veut construire un processus réel intégré d'ordre 1 à une fréquence  $\omega \in ]0, \pi[$ , il faut appliquer  $S_\omega S_{-\omega}$ .

Ces opérateurs satisfont un ensemble de propriétés algébriques, nous en donnons quelques unes pertinentes pour notre propos (voir [31]).

- (a)  $\forall \omega \in [-\pi, \pi], \delta_\omega S_\omega = I$
- (b)  $\forall \{y_t\} \in \mathbf{R}^Z, \forall \omega \in [0, \pi], S_\omega \delta_\omega y_t = y_t - y_0 e^{-i\omega t}$
- (c)  $\forall \{y_t\} \in \mathbf{R}^Z, \forall \omega \in ]0, \pi[, S_\omega S_{-\omega} \Delta_\omega y_t = y_t - \frac{\sin \omega(t+1)}{\sin \omega} y_0 + \frac{\sin \omega t}{\sin \omega} y_{-1}$
- (d)  $\forall (\omega_1, \omega_2) \in [0, \pi]^2, S_{\omega_1} S_{\omega_2} = S_{\omega_2} S_{\omega_1}$
- (e)  $\forall (\omega_1, \omega_2, \omega_3) \in [0, \pi]^3, (S_{\omega_1} S_{\omega_2}) S_{\omega_3} = S_{\omega_1} (S_{\omega_2} S_{\omega_3})$
- (f)  $\forall (\omega_1, \omega_2) \in [0, \pi]^2, S_{\omega_1} S_{\omega_2} = \frac{e^{i\omega_1}}{e^{i\omega_1} - e^{i\omega_2}} S_{\omega_2} + \frac{e^{i\omega_2}}{e^{i\omega_2} - e^{i\omega_1}} S_{\omega_1}$

Des équations (a) et (b) nous observons que les deux types d'opérateurs ne commutent pas. La fonction déterministe que nous introduisons dans (b) est nécessaire pour assurer la propriété selon laquelle les processus intégrés construits avec les opérateurs d'intégration sont égaux à 0 à la date 0. Ceci est une convention. Un choix différent impliquerait l'introduction d'une autre fonction déterministe. Le dernier point (f) permet de remplacer tout produit d'opérateurs d'intégration à différentes fréquences par la somme des ces opérateurs à chaque fréquence. Cette propriété sera utilisée intensivement par la suite.

## Composantes persistantes

La possibilité de décrire de manière parcimonieuse un grand nombre de séries macroéconomiques comme des processus intégrés d'ordre 1 à la fréquence 0 a entraîné l'introduction de nouvelles statistiques descriptives. Parmi celles-ci, la mesure de la persistance des chocs et la décomposition de séries intégrées en la somme d'une série persistante et d'une série transitoire ont été l'objet de nombreux travaux ([4], [10], [11], [55] entre autres). En effet, à la différence des processus stationnaires, les chocs affectant un processus intégré affectent indéfiniment ses valeurs à venir. Cependant, [49] a montré qu'il existait une infinité de décomposition possible et qu'à chacune d'elles pouvait être associée une mesure de persistance. Ce type de mesure n'est donc pas informatif d'un point de vue structurel. En particulier, la composante permanente peut être choisie arbitrairement aussi régulière que souhaité si bien que les fluctuations de la composante transitoire peuvent être les plus importantes en variabilité à tous les horizons finis. D'un point de vue descriptif, de nombreuses décompositions peuvent être considérées. [55] introduit trois situations élémentaires : quand les deux composantes sont orthogonales, quand les innovations des deux composantes sont parfaitement corrélées et une situation intermédiaire. Si la première situation impose des contraintes sur la forme de la densité spectrale du processus stationnarisé, la troisième nécessite une théorie pour fonder le choix des modèles de chaque composante. La deuxième situation correspond à la décomposition de Beveridge-Nelson [4], elle est toujours possible mais peut donner une composante persistante estimée très voisine de la série initiale, ce qui en limite la pertinence. Comme nous l'avons indiqué plus haut, la lecture difficile des séries temporelles saisonnières vient de la présence de composantes au profil saisonnier persistant. Lorsque ces profils persistants sont bien décrits par des processus intégrés aux fréquences saisonnières, une stratégie naturelle consiste dans un premier temps à décomposer les séries intégrées en une somme de composantes intégrées à chaque fréquence à l'aide de la propriété (f) des opérateurs d'intégration, puis dans un second temps à décomposer chacune de ces composantes en une composante permanente et une composante transitoire et enfin dans une dernière étape à éliminer les composantes permanentes. Nous détaillerons la première opération dans la prochaine section, nous traitons ici de la décomposition des processus intégrés à fréquence saisonnière en la somme d'une composante persistante et une composante transitoire.

Nous pouvons en effet transposer le cadre d'analyse aux processus intégrés à fréquence 0 celui des processus intégrés à des fréquences saisonnières. Une mesure de la persistance a déjà été introduite par [48] pour des processus intégrés d'ordre 1 à toutes les fréquences saisonnières. Elle

est égale au ratio de variance introduit par [11] mais appliqué à la différence saisonnière  $(y_t - y_{t-s})$ . Nous suivons une approche différente. Comme nous venons de le rappeler, il existe une multitude de définition de la composante permanente. Nous sélectionnons pour des raisons de simplicité et de flexibilité une décomposition de type Beveridge-Nelson appliquée aux processus intégrés à chaque fréquence saisonnière. Le fait que cette décomposition puisse donner une composante persistante voisine de la série initiale n'est pas une limite ici dans la mesure où nous souhaitons extraire la plus importante composante saisonnière en termes de variance. Des décompositions différentes donneront des séries corrigées des variations saisonnières différentes.

Le principe de notre décomposition repose sur une opération de démodulation à la fréquence  $\omega$  qui déplace la densité spectrale de la série le long de l'axe des abscisses de cette quantité. Appliquée sur un processus intégré d'ordre 1 à la fréquence  $\omega$ , elle donne un processus complexe intégré à la fréquence 0 qui peut être décomposé en un processus persistant et un processus transitoire.

Un processus complexe  $\{x_t\}$  purement non déterministe intégré d'ordre 1 à la fréquence  $\omega \in [0, \pi]$  admet une fois différencié une représentation de Wold de la forme suivante  $\delta_\omega(B)x_t = c(B)\varepsilon_t$  où  $\{\varepsilon_t\}$  est un bruit blanc faible (complexe) de variance  $\sigma^2$ . Une condition additionnelle est nécessaire pour garantir que dans la décomposition que nous allons considérer la composante transitoire est un processus stationnaire du second ordre.

**Condition 1 :**  $c(B) = \sum_{j=0}^{+\infty} c_j B^j$  est tel que  $\sum_{p=0}^{+\infty} p |c_p| < +\infty$

Sous la Condition 1, il existe à la fréquence  $\omega$  une décomposition de type Beveridge-Nelson du polynôme  $c(B)$  comme suit :

$$c(B) = c(e^{i\omega}) + \delta_\omega(B)\tilde{c}_\omega(B)$$

où  $\tilde{c}_\omega(B) = \sum_{p=0}^{+\infty} \tilde{c}_{\omega,p} B^p$  et  $\tilde{c}_{\omega,p} = - \sum_{q=p+1}^{+\infty} c_q e^{i\omega(p-q)}$  tel que  $\{\tilde{c}(B)\varepsilon_t\}$  est un processus complexe

stationnaire du second ordre. Cette équation permet de proposer une définition de la composante persistante dans une décomposition du processus intégré à cette fréquence. En effet, par (a), nous avons

$$\delta_\omega(B)(x_t - c(e^{i\omega})S_\omega \varepsilon_t - \tilde{c}(B)\varepsilon_t) = 0$$

donc

$$x_t = \mu_t + c(e^{i\omega})S_\omega \varepsilon_t + \tilde{c}(B)\varepsilon_t$$

où  $\mu_t = \mu e^{-i\omega t}$  est une fonction déterministe telle que  $\delta_\omega(B)\mu_t = 0$ . Ceci donne une possible décomposition en un processus non-stationnaire à la fréquence  $\omega$  :  $\mu_t + c(e^{i\omega})S_\omega \varepsilon_t$  et un processus stationnaire :  $\tilde{c}(B)\varepsilon_t$ . Par ailleurs, si nous considérons le processus démodulé à la fréquence  $\omega$ , à savoir  $\{e^{i\omega t} x_t\}$ , et calculons la valeur de sa prévision linéaire à long-terme (au terme  $e^{i\omega t}$  près) :  $\lim_{h \rightarrow +\infty} E_t e^{i\omega h} x_{t+h}$ , cette quantité est caractéristique de la persistante des chocs à la fréquence  $\omega$  et est égale à la composante persistante introduite ci-dessus :

$$\begin{aligned}
\lim_{h \rightarrow +\infty} E_t e^{i\omega h} x_{t+h} &= \mu e^{-i\omega t} + c \left( e^{i\omega} \right) \sum_{\tau=1}^t e^{-i\omega(t-\tau)} \varepsilon_\tau \\
&+ \lim_{h \rightarrow +\infty} E_t e^{i\omega h} \left( c \left( e^{i\omega} \right) \sum_{\tau=t+1}^{t+h} e^{-i\omega(t+h-\tau)} \varepsilon_\tau + \tilde{c}_\omega(B) \varepsilon_{t+h} \right) \\
&= \mu e^{-i\omega t} + c \left( e^{i\omega} \right) S_\omega \varepsilon_t + \lim_{h \rightarrow +\infty} E_t e^{i\omega h} \tilde{c}_\omega(B) \varepsilon_{t+h}
\end{aligned}$$

où le dernier terme tend vers 0 puisqu'il s'agit d'un processus stationnaire. Ainsi, nous obtenons

$$\lim_{h \rightarrow +\infty} E_t e^{i\omega h} x_{t+h} = \mu e^{-i\omega t} + c \left( e^{i\omega} \right) S_\omega \varepsilon_t$$

Quelques exemples peuvent illustrer notre propos.

**Exemple 1 :** Un processus persistant à la fréquence  $\omega$  est la marche aléatoire complexe qui satisfait l'équation suivante  $x_t = e^{i\omega} x_{t-1} + \varepsilon_t$  où  $\{\varepsilon_t\}$  est un bruit blanc (complexe). Dans ce cas  $\lim_{h \rightarrow +\infty} E_t e^{i\omega h} x_{t+h} = x_t$ , résultat similaire à ce que l'on obtient pour la marche aléatoire réelle.

**Exemple 2 :** Soit  $\{x_t\}$  un processus stationnaire du second ordre satisfaisant une représentation AR(1) :  $x_t = \rho x_{t-1} + \varepsilon_t, |\rho| < 1$ , alors  $S_\omega x_t$  est un processus intégré à la fréquence  $\omega$ . Par construction pour des indices  $t > 0$ , nous avons

$$S_\omega x_{t+h} = \sum_{\tau=1}^{t+h} x_\tau e^{-i\omega(t+h-\tau)}$$

et donc

$$e^{i\omega(t+h)} S_\omega x_{t+h} = \sum_{\tau=1}^{t+h} x_\tau e^{i\omega\tau}$$

d'où

$$\begin{aligned}
E_t e^{i\omega(t+h)} S_\omega x_{t+h} &= \sum_{\tau=1}^t x_\tau e^{i\omega\tau} + \sum_{\tau=t+1}^{t+h} E_t x_\tau e^{i\omega\tau} \\
&= \sum_{\tau=1}^t x_\tau e^{i\omega\tau} + \left( \sum_{\tau=t+1}^{t+h} \rho^{\tau-t} e^{i\omega\tau} \right) x_t \\
&= \sum_{\tau=1}^t x_\tau e^{i\omega\tau} + \frac{\rho e^{i\omega t} (1 - \rho^{h-1} e^{i\omega(h-1)})}{1 - \rho e^{i\omega}} x_t
\end{aligned}$$

soit

$$\lim_{h \rightarrow +\infty} E_t e^{i\omega h} S_\omega x_{t+h} = S_\omega x_t + \frac{\rho}{1 - \rho e^{i\omega}} x_t$$

## Correction des variations saisonnières par soustraction des composantes persistantes saisonnières

Nous introduisons maintenant la procédure générale pour décomposer toute série intégrée en la somme d'une composante persistante saisonnière et d'une composante persistante non saisonnière. La composante saisonnière est la somme de composantes déterministes et de marches aléatoires saisonnières. Nous considérons des processus dont le processus de génération des données peut être décrit comme suit :

$$y_t = d_t + x_t$$



où  $d_t$  est une fonction déterministe combinaison linéaire des fonctions  $\left\{e^{iv_j t}\right\}_{v_j \in \Omega_d}$  avec  $\Omega_d = \{v_0, \dots, v_l\}$  et  $\{x_t\}$  est un processus intégré d'ordre 1 à différentes fréquences dans  $\Omega_x = \{\omega_0, \dots, \omega_k\}$ ,  $\Omega_d$  et  $\Omega_x$  sont inclus dans l'ensemble  $\left\{\varpi_j = \frac{2j\pi}{s}, j \in \{0, 1, \dots, s\}\right\}$  lorsqu'il y a  $s$  observations par période de référence<sup>5</sup>. Nous notons  $\Delta_x(B) = \prod_{j=0}^k \delta_{\omega_j}(B)$  tel que  $\{\Delta_x(B)x_t\}$  est un processus stationnaire du second ordre. Par hypothèse, nous supposons que  $v_0 = 0$  et  $\omega_0 = 0$ , mais l'analyse peut s'étendre à des situations dans lesquelles ceci n'est pas vérifié. Il en est de même lorsque le processus  $\{x_t\}$  est intégré d'ordre plus grand que 1, comme dans le cas du modèle « Airline » étudié en détail dans la section ci-dessous. Nous considérons dans les deux sections suivantes le cas général pour un processus univarié et un processus multivarié.

### Exemple : le modèle « Airline »

Le modèle « Airline » a été utilisé par [5] pour étudier la série temporelle du nombre de passagers transportés par les compagnies aériennes états-unienues. Il a pour spécification

$$(1-B)(1-B^s)y_t = (1-\theta B)(1-\Theta B^s)\varepsilon_t$$

où  $\{\varepsilon_t\}$  est un bruit blanc faible. Nous considérons des séries trimestrielles ( $s=4$ ). Par les propriétés (a) et (c) des opérateurs d'intégration, nous avons

$$(1-B)(1-B^4)S_0^2 S_\pi S_{\frac{\pi}{2}} S_{-\frac{\pi}{2}} = 1$$

et donc

$$\forall t \in \mathbb{Z}, (1-B)(1-B^4) \left[ y_t - (1-\theta B)(1-\Theta B^4) S_0^2 S_\pi S_{\frac{\pi}{2}} S_{-\frac{\pi}{2}} \varepsilon_t \right] = 0$$

Si nous notons  $\mu_t$  les fonctions déterministes dans le noyau de l'opérateur  $(1-B)(1-B^4)$ , elles ont nécessairement la forme suivante :

$$\mu_t = \mu_{0,2} t + \mu_{0,1} + \mu_{\pi,1} (-1)^t + \mu_{\frac{\pi}{2},1} e^{i\frac{\pi}{2}t} + \bar{\mu}_{\frac{\pi}{2},1} e^{-i\frac{\pi}{2}t}$$

si nous nous restreignons à des fonctions réelles, si bien que

$$y_t = \mu_t + (1-\theta B)(1-\Theta B^4) S_0^2 S_\pi S_{\frac{\pi}{2}} S_{-\frac{\pi}{2}} \varepsilon_t$$

où les paramètres  $\left(\mu_{0,2}, \mu_{0,1}, \mu_{\pi,1}, \mu_{\frac{\pi}{2},1}\right) \in \mathbf{R}^3 \times \mathbf{C}$  peuvent être calculés en fonction des valeurs

prises par  $\{y_t, \varepsilon_t\}$  aux dates  $t=0, -1, -2, -3, -4$ . En utilisant itérativement la propriété (f), nous obtenons que

$$S_0^2 S_\pi S_{\frac{\pi}{2}} S_{-\frac{\pi}{2}} = \frac{1}{4} \left( S_0^2 + \frac{3}{2} S_0 + \frac{1}{2} S_\pi + \frac{1}{1-i} S_{\frac{\pi}{2}} + \frac{1}{1+i} S_{-\frac{\pi}{2}} \right)$$

<sup>5</sup> Le cadre peut être étendu en particulier si l'on souhaite traiter les effets de jours ouvrés en faisant référence à une fréquence particulière.

Nous pouvons donc pour tout  $t > 0$  décomposer le processus  $\{y_t\}$  en la somme de quatre processus intégrés à chaque fréquence, chaque composante étant elle-même la somme d'un processus déterministe et d'un processus stochastique intégré :

$$\begin{aligned} y_t = & \mu_{0,2}t + \mu_{0,1} + \frac{1}{4}(1-\theta B)(1-\Theta B^4) \left( S_0^2 + \frac{3}{2}S_0 \right) \varepsilon_t \\ & + \mu_{\pi,1} + \frac{1}{8}(1-\theta B)(1-\Theta B^4) S_{\pi} \varepsilon_t \\ & + \mu_{\frac{\pi}{2},1} e^{i\frac{\pi}{2}t} + \bar{\mu}_{\frac{\pi}{2},1} e^{-i\frac{\pi}{2}t} + \frac{1}{4}(1-\theta B)(1-\Theta B^4) \left( \frac{1}{1-i} S_{\frac{\pi}{2}} + \frac{1}{1+i} S_{-\frac{\pi}{2}} \right) \varepsilon_t \end{aligned}$$

La décomposition en une composante saisonnière et une composante non-saisonnière qui en découle n'est pas unique dans la mesure où il est toujours possible de déplacer des termes stationnaires d'une composante à l'autre sans en changer la nature. Pour donner un caractère opérationnel à cette décomposition, nous proposons de définir la composante saisonnière à partir des marches aléatoires complexes sous-jacentes. Ceci peut être fait en faisant usage des développements suivants :

$$(1-\theta B)(1-\Theta B^4) = (1+\theta)(1-\Theta) + \theta_{\pi}(B)(1+B)$$

où

$$\theta_{\pi}(B) = (-\theta + \Theta(\theta+1))(1-B+B^2-B^3) + \Theta\Theta B^4$$

et

$$(1-\theta B)(1-\Theta B^4) = (1+i\theta)(1-\Theta) + \theta_{\frac{\pi}{2}}(B)(1+iB)$$

où

$$\theta_{\frac{\pi}{2}}(B) = \Theta(-i\theta+1)(1-iB-B^2+iB^3) + i\theta(1-\Theta B^4)$$

Il s'ensuit qu'une décomposition possible de  $\{y_t\}$  en une composante saisonnière  $\{y_t^s\}$  et une composante non-saisonnière  $\{y_t^*\}$  est donnée par :

$$y_t = y_t^s + y_t^*$$

avec

$$y_t^* = \mu_{0,2}t + \mu_{0,1} + \frac{1}{4}(1-\theta B)(1-\Theta B^4) \left( S_0^2 + \frac{3}{2}S_0 \right) \varepsilon_t + \theta(B) \varepsilon_t$$

où

$$\theta(B) = \frac{1}{8} \left[ 3((1+\theta)\Theta - \theta) + \Theta(1-3\theta)B - 3(1+\theta)\Theta B^2 + \Theta(\theta-3)B^3 + 3\Theta\theta B^4 \right]$$

et

$$\begin{aligned} y_t^s = & \mu_{\pi,1} + \frac{1}{8}(1+\theta)(1-\Theta) S_{\pi} \varepsilon_t \\ & + \mu_{\frac{\pi}{2},1} e^{i\frac{\pi}{2}t} + \bar{\mu}_{\frac{\pi}{2},1} e^{-i\frac{\pi}{2}t} + \frac{1-\Theta}{4} \left( \frac{1-i\theta}{1-i} S_{\frac{\pi}{2}} + \frac{1+i\theta}{1+i} S_{-\frac{\pi}{2}} \right) \varepsilon_t \end{aligned}$$

La composante saisonnière est la somme de processus déterministes saisonniers et de marches aléatoires saisonnières. Le processus  $\{y_t^*\}$  est la série initiale corrigée des composantes saisonnières. Si nous remplaçons  $\varepsilon_t$  dans l'expression de  $y_t^s$  par

$(1-\theta B)^{-1}(1-\Theta B^4)^{-1}(1-B)(1-B^4)y_t$ , nous obtenons une expression du filtre  $\Psi(B)$  implicitement appliqué sur la série initiale pour obtenir la série corrigée des variations saisonnières. Il prend la forme

$$\Psi(B) = \left[ \frac{1}{4} + \frac{3}{8}(1-B) + \frac{\theta(B)(1-B)^2}{(1-\theta B)(1-\Theta B^4)} \right] (1+B+B^2+B^3)$$

C'est filtre asymétrique qui est égal à 1 en 0 et qui est proche de 1 à son voisinage. Il vaut 0 aux fréquences  $\pi$  et  $\frac{\pi}{2}$ , néanmoins la densité spectrale de la série corrigée n'est pas égale à 0 en ces fréquences. Il n'y a pas de sur-ajustement aux fréquences saisonnières. Appliqué à une série stationnaire, ce filtre peut introduire un déphasage. Nous illustrons ces propriétés dans le graphique suivant. Enfin, nous remarquons que nous pouvons toujours calculer cette décomposition et le filtre associé, ce qui n'est pas le cas dans le cas d'une décomposition en composantes orthogonales. [38] montrent qu'une telle décomposition n'existe que si  $\Theta > 0$ . Dans la Figure 1, nous considérons deux jeux de valeurs  $(\theta, \Theta) = (0,25 ; 0,25)$  et  $(\theta, \Theta) = (0,25 ; -0,25)$ . Nous observons que dans le second cas, le gain est très supérieur à 1, mais le déphasage au pour le cycle de moyen et long terme demeure faible.

## Processus univarié

Nous considérons le cas général d'un processus univarié satisfaisant le système d'équations suivant

$$\begin{cases} y_t & = d_t + x_t \\ d_t & = \sum_{j=0}^l \mu_{v_j} e^{iv_j t} \\ \Delta_x(B)x_t & = c(B)\varepsilon_t \end{cases}$$

Nous supposons que  $c(B)$  satisfait la Condition 1. Par la propriété (a) des opérateurs d'intégration, nous pouvons écrire  $\Delta_x(B) \prod_{j=0}^k S_{\omega_j} = 1$  et introduire cette équation dans la représentation de Wold de  $\{x_t\}$ , ce qui donne

$$\Delta_x(B) \left( x_t - c(B) \prod_{j=0}^k S_{\omega_j} \varepsilon_t \right) = 0$$

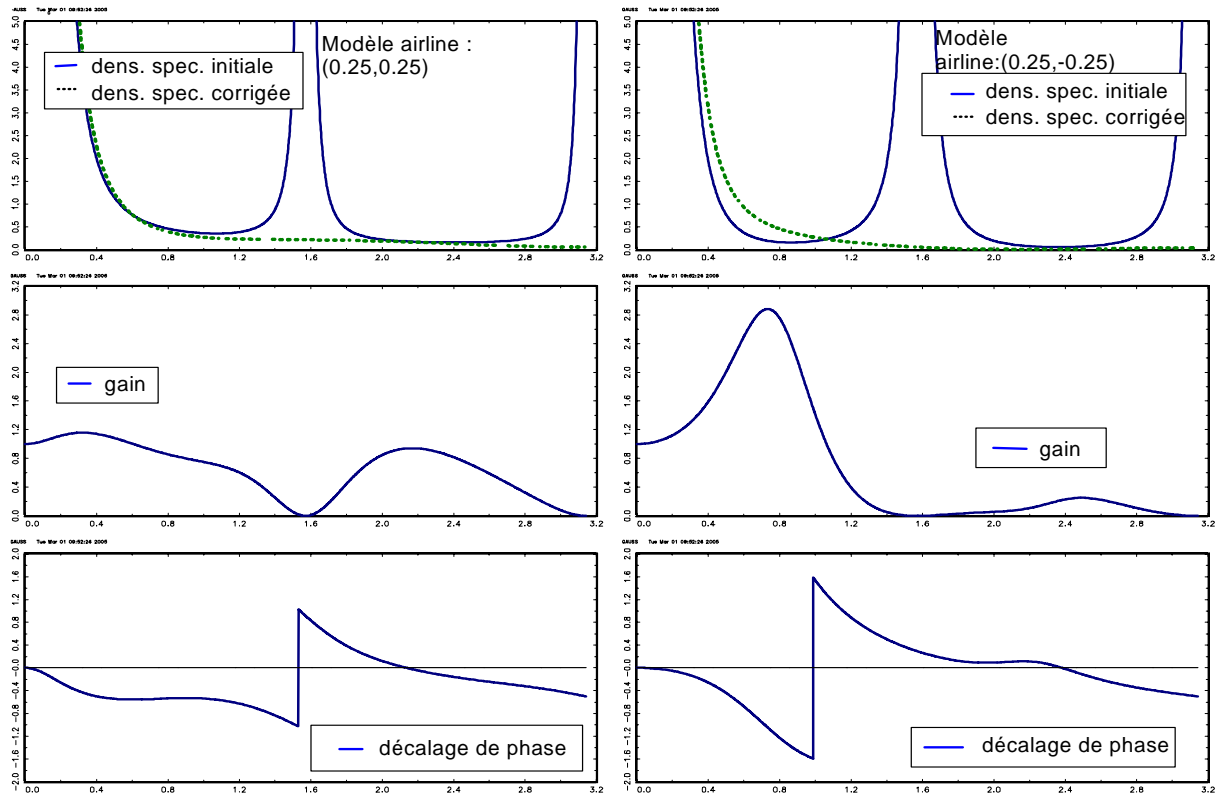


Figure 1

D'une part, il existe une fonction déterministe  $\mu_{x,t} = \sum_{j=0}^k \mu_{x,\omega_j} e^{-i\omega_j t}$  telle que  $\Delta_x(B)\mu_{x,t} = 0$  et donc  $x_t = \mu_{x,t} + c(B) \prod_{j=0}^k S_{\omega_j} \varepsilon_t$ . D'autre part, en appliquant (f) itérativement, nous obtenons qu'il existe  $(k+1)$  nombres complexes  $\{\phi_j\}_{j=0,1,\dots,k}$  tels que

$$\prod_{j=0}^k S_{\omega_j} = \sum_{j=0}^k \phi_j S_{\omega_j}$$

Si nous notons  $\Delta_{x,-j}(B) = \prod_{m=0, m \neq j}^k \delta_{\omega_m}(B)$ , nous obtenons aisément que  $\phi_j = \frac{1}{\Delta_{x,-j}(e^{i\omega_j})}$ . En

utilisant alors la famille des décompositions de Beveridge-Nelson à chaque fréquence :

$$c(B)S_{\omega_j} \varepsilon_t = c(e^{i\omega_j})S_{\omega_j} \varepsilon_t + \tilde{c}_{\omega_j}(B) \varepsilon_t$$

nous pouvons proposer une décomposition du processus en la somme d'une composante sans persistance saisonnière et une composante purement persistante aux fréquences saisonnières par les définitions suivantes :

$$y_t = y_t^* + y_t^s$$

où

$$y_t^* = \mu_{v_0} + \mu_{x,\omega_0} + \phi_0 c(B) S_0 \varepsilon_t + \sum_{j=1}^k \phi_j \tilde{c}_{\omega_j}(B) \varepsilon_t$$

et

$$y_t^s = \sum_{j=1}^l \mu_{v_j} e^{iv_j t} + \sum_{j=1}^k \mu_{x,\omega_j} e^{-i\omega_j t} + \sum_{j=1}^k \phi_j c(e^{i\omega_j}) S_{\omega_j} \varepsilon_t$$

Nous remarquons de la définition des deux composantes ci-dessus qu'il peut exister de nombreux problèmes d'identification habituels dans ce genre de situation. Ils apparaissent clairement à la fréquence 0 mais peuvent aussi être présents aux fréquences saisonnières et sont de nature similaire.

En effet, le terme constant de  $y_t^*$  est défini comme la somme de deux termes, n'importe quel choix de deux nombres ayant la même somme est un jeu de paramètres satisfaisant notre définition. Nous verrons dans la partie suivante comment nous résolvons ce problème en choisissant un estimateur particulier.

Le filtre  $\Psi$  qui est implicitement appliqué au processus  $\{y_t - d_t\}$  pour obtenir le processus corrigé des variations saisonnières est égal à

$$\Psi(B) = 1 - \sum_{j=1}^k \frac{\Delta_{x,-j}(B) c(e^{i\omega_j})}{\Delta_{x,-j}(e^{i\omega_j}) c(B)}$$

Il est donc égal à 1 en 0 et proche de 1 au voisinage de 0. Il est nul en chacune des fréquences saisonnières.

## Processus multivarié

Le cadre que nous avons introduit dans la section précédente peut s'étendre sans difficulté aux processus multivariés, si ce n'est aux coûts de notations supplémentaires. La différence principale vient du fait que nous travaillons alors avec des matrices et que des problèmes de rang doivent être étudiés.

Nous considérons de processus de dimension  $n$  dont le processus de génération des données peut être décrit comme suit :

$$y_t = d_t + x_t$$

où  $d_t$  est une fonction déterministe combinaison linéaire des fonctions  $\{e^{iv_j t}\}_{v_j \in \Omega_d}$  avec

$\Omega_d = \{v_0, \dots, v_l\}$  et  $\{x_t\}$  est un processus de dimension  $n$  tel que la  $r^{\text{th}}$  composante est intégré d'ordre 1 à différentes fréquences dans  $\Omega_x = \{\omega_0, \dots, \omega_k\}$ . Par souci de simplicité, nous supposons que chaque composante est intégrée aux mêmes fréquences. Cette hypothèse peut être relâchée sans difficulté, mais à des coûts de notation importants.  $\Omega_d$  et  $\Omega_x$  satisfont les mêmes

propriétés de la section précédente. Nous notons  $\Delta_x(B) = \prod_{j=0}^k \delta_{\omega_j}(B)$  tel que  $\{\Delta_x(B)x_t\}$  est un processus stationnaire du second ordre de dimension  $n$  dont la matrice polynomiale associée à sa représentation de Wold est notée  $C(B)$  et le processus innovation  $\{\varepsilon_t\}$  :

$$\begin{cases} y_t & = & d_t + x_t \\ d_t & = & \sum_{j=0}^l \mu_{v_j} e^{iv_j t} \\ \Delta_x(B)x_t & = & C(B)\varepsilon_t \end{cases}$$

Comme dans le cas univarié, nous avons besoin d'une condition supplémentaire pour assurer que les processus introduits dans la décomposition de type Beveridge-Nelson multivarié pour décrire la composante transitoire soient stationnaires du second ordre.

**Condition 2 :**  $C(B) = \sum_{p=0}^{+\infty} C_p B^p$  est tel que  $\sum_{p=0}^{+\infty} p (Tr C_p C_p')^{1/2} < +\infty$

Nous pouvons suivre les mêmes lignes de raisonnement que dans le cas univarié. Sous la Condition 2, pour chaque fréquence  $\omega_j \in \Omega_x$ , il existe une décomposition de type Beveridge-Nelson de la forme suivante :

$$C(B) = C(e^{i\omega_j}) + \delta_{\omega_j}(B) \tilde{C}_{\omega_j}(B)$$

où  $\tilde{C}_{\omega_j}(B) = \sum_{p=0}^{+\infty} \tilde{C}_{\omega_j,p} B^p$  avec  $\tilde{C}_{\omega_j,p} = - \sum_{q=p+1}^{+\infty} C_q e^{i\omega_j(p-q)}$  est tel que  $\{\tilde{C}_{\omega_j}(B) \varepsilon_t\}$  est un processus complexe stationnaire du second ordre. Pour chaque fréquence, nous avons :

$$\Delta_{x,-j}(B) x_t = C(B) S_{\omega_j} \varepsilon_t = C(e^{i\omega_j}) S_{\omega_j} \varepsilon_t + \tilde{C}_{\omega_j}(B) \varepsilon_t$$

On note  $r_j$  le rang de  $C(e^{i\omega_j})$ . Si  $r_j < n$ , il existe une matrice complexe de plein rang  $\alpha_{j,\perp}$  de taille  $(n - r_j) \times n$ , tel que  $\alpha_{j,\perp} C(e^{i\omega_j}) = 0$ . Ceci implique que

$$\alpha_{j,\perp} C(B) S_{\omega_j} \varepsilon_t = \alpha_{j,\perp} \Delta_{x,-j}(B) x_t = \tilde{C}_{\omega_j}(B) \varepsilon_t$$

est un processus complexe stationnaire du second ordre alors que  $\{\Delta_{x,-j}(B) x_t\}$  est un processus intégré d'ordre 1 à la fréquence  $\omega_j$ . Dans cette situation, on dit que le processus est cointégré à fréquence  $\omega_j$  ([39], [31], [13]). Cette propriété signifie que les sources de mouvements saisonniers persistants dans les  $n$  composantes du processus sont décrites à l'aide de  $r_j$  composantes. Plus précisément,  $C(e^{i\omega_j})$  étant de rang  $r_j$ , il existe deux matrices de plein rang de dimension  $r_j \times n$  telles que  $C(e^{i\omega_j}) = \alpha_j' \beta_j$  où  $\alpha_j'$  représente la transposée-conjuguée de la matrice  $\alpha_j$ . Dans ces conditions, l'équation de Beveridge-Nelson prend la forme suivante :

$$C(B) S_{\omega_j} \varepsilon_t = \alpha_j' \beta_j S_{\omega_j} \varepsilon_t + \tilde{C}_{\omega_j}(B) \varepsilon_t$$

$\{\beta_j S_{\omega_j} \varepsilon_t\}$  est un processus complexe intégré d'ordre 1 à la fréquence  $\omega_j$  et de dimension  $r_j$  qui génère les mouvements persistants saisonniers dans le processus de dimension  $n$ . Nous pouvons alors aboutir à une décomposition très similaire à ce qui a été obtenu dans le cas univarié. Nous avons

$$y_t = y_t^* + y_t^s$$

avec

$$y_t^* = \mu_{v_0} + \mu_{x,\omega_0} + \phi_0 C(B) S_0 \varepsilon_t + \sum_{j=1}^k \phi_j \tilde{C}_{\omega_j}(B) \varepsilon_t$$

et

$$y_t^s = \sum_{j=1}^l \mu_{v_j} e^{iv_j t} + \sum_{j=1}^k \mu_{x,\omega_j} e^{-i\omega_j t} + \sum_{j=1}^k \phi_j \alpha_j' \beta_j S_{\omega_j} \varepsilon_t$$

Néanmoins le filtre  $\Psi$  qui est implicitement appliqué au processus  $\{y_t - d_t\}$  pour construire le processus corrigé des variations saisonnières  $\{y_t^* - \mu_{v_0}\}$  est plus compliqué du fait de la possible non-inversibilité de  $C(B)$  à la fréquence  $\omega_j$ . Pour obtenir le filtre, nous avons besoin d'une hypothèse additionnelle et d'un théorème de représentation. Sous l'hypothèse que

$\forall \omega_j \in \Omega_j, \text{rang} C(e^{i\omega_j}) = r_j$  et que  $\det C(B) = d(B) \prod_{j=0}^k \delta_{\omega_j}(B)^{r_j}$  où  $d(B)$  est un polynôme dont les racines sont de module strictement plus grand que 1, un théorème de représentation dans [31] permet d'affirmer qu'il existe une matrice polynomiale  $D(B)$  telle que  $D(0) = I_n$  et  $\sum_{p=0}^{+\infty} \text{Tr}(D_p D_p')$   $< +\infty$ , un ensemble de  $(k+1)$  couples de matrices complexes  $(\gamma_j, \theta_j)$  de plein rang de dimension  $(r_j \times n)$  tels que  $\beta_j \gamma_j' = 0$  et

$$D(B) \Delta_x(B) x_t = \sum_{m=0}^k \gamma_m' \theta_m \Delta_{x,-m}(B) x_t + \varepsilon_t$$

Il s'ensuit que

$$\alpha_j' \beta_j S_{\omega_j} \varepsilon_t = \alpha_j' \beta_j \left[ D(B) \Delta_{x,-j}(B) - \sum_{m=0, m \neq j}^k \gamma_m' \theta_m \Delta_{x,-(m,j)}(B) \right] x_t$$

$$\text{où } \Delta_{x,-(m,j)}(B) = \prod_{l=0, l \neq m, l \neq j}^k \delta_{\omega_l}$$

Le filtre multivarié prend alors la forme suivante :

$$\Psi(B) = I_n - \sum_{j=1}^k \frac{1}{\Delta_{x,-j}(e^{i\omega_j})} \alpha_j' \beta_j \left[ D(B) \Delta_{x,-j}(B) - \sum_{m=0, m \neq j}^k \gamma_m' \theta_m \Delta_{x,-(m,j)}(B) \right]$$

Il est égal à  $I_n$  en 0 et proche de cette matrice au voisinage de cette fréquence.

## Avantages et inconvénients

Le cadre de travail que nous venons d'introduire repose sur une approche descriptive linéaire. Il suppose qu'une condition de sommabilité est satisfaite, ce qui exclut des formes d'intégration fractionnaire à la fréquence 0 et aux fréquences saisonnières. C'est une approche pragmatique qui de fait exploite le cadre usuellement utilisé mais dans une approche différente par la nature des liens stochastiques qui existent entre les différentes composantes, celles-ci étant construites sur un même processus des innovations. Cette décomposition est toujours possible. La définition des composantes saisonnières persistantes est intuitive. Elle correspond à l'extension aux fréquences saisonnières de l'approche de Beveridge-Nelson. Corriger des variations saisonnières correspond alors à soustraire des observations ces composantes saisonnières persistantes. Une définition alternative des composantes saisonnières persistante est possible, elle introduirait dans ce cadre une distorsion plus importante entre la série brute et la série corrigée.

Le filtre implicite utilisé est unidirectionnel. Il est égal à 1 à la fréquence 0. Appliqué à une série stationnaire, il peut introduire un déphasage aux fréquences du cycle économique de moyen terme. Néanmoins, l'exemple du modèle « Airline » montre que ce déphasage peut être faible. Le traitement des extrémités d'échantillon est toujours délicat. Il est parfois fait usage de filtres asymétriques construits en cherchant à minimiser l'impact des révisions sous des hypothèses *ad hoc* quant à la nature du processus initial. L'approche développée ici donne un cadre d'analyse et de construction de ce type de filtres. Le filtre dépend alors du processus de génération des données initiales.

# Mise en œuvre de la procédure dans le cadre d'un modèle SARIMA

## Test de racine unitaire

Dans le cas univarié, la mise en œuvre de la méthode suppose premièrement que l'on soit capable de prendre une décision sur le modèle adapté pour décrire la série étudiée en ce qui concerne ses composantes saisonnières persistantes et deuxièmement que l'on soit capable d'en fournir des estimateurs.

D'un point de vue pratique, les statisticiens sont enclins à considérer des situations dans lesquelles l'amplitude des mouvements saisonniers peuvent évoluer au cours du temps. Pour ce faire, ils ont donc recours à des modèles de processus intégrés à des fréquences saisonnières, éventuellement en présence de fonctions déterministes périodiques aux mêmes fréquences. Une stratégie de mise en œuvre suppose donc la possibilité de tester la présence de racine unitaire à des fréquences saisonnières en présence de fonctions déterministes.

Une première approche considère simultanément l'ensemble des racines saisonnières pour une fréquence d'observation donnée, elles correspondent aux racines de l'opérateur de différenciation saisonnière  $1 - B^s$ . Différents tests ont été proposés entre autres par [37], [17], [51] et [8]. Cependant [52] a montré que cette approche peut amener le statisticien à accepter la présence de racines saisonnières sous l'hypothèse nulle alors que le processus de génération des données admet une racine unitaire à fréquence 0 mais aucune aux fréquences saisonnières.

Une seconde approche traite séparément chacune des racines saisonnières (ou couple de racine conjuguée car nous travaillons avec des séries temporelles réelles). La procédure et la loi asymptotique des statistiques de test dépendent de la fréquence, suivant que  $\omega \in \{0, \pi\}$  ou  $\omega \in ]0, \pi[$ . Pour des données mensuelles ou trimestrielles, [39], [2] et [25] utilise une stratégie qui repose sur des spécifications dites de « Dickey-Fuller augmenté » et considèrent des statistiques de test de Student ou de Fisher. [50] étend une approche par test de multiplicateur de Lagrange. [33] dans un cadre de travail qui établit un parallèle entre les tests de racines saisonnières et les tests à fréquence 0 étend l'approche de [18]. Une étude de simulation montre que les tests presque efficaces ainsi construits dominent en termes de puissance ajustée la plupart des tests introduits auparavant. Une propriété d'orthogonalité asymptotique des produits croisés de processus intégrés à des fréquences différentes permet d'utiliser la même approche pour des séries qui présentent des racines unitaires saisonnières à différentes fréquences. En pratique dans des études empiriques, il semble que de nombreuses séries soient décrites de manière parcimonieuse comme des processus intégrés mais à un sous ensemble des fréquences possibles. L'introduction de fonctions sinusoïdales est parfois nécessaire aux fréquences restantes.

Ces tests reposent sur la détection de chocs persistants. Comme l'ont illustré [45] à la fréquence 0 et [22] aux fréquences saisonnières, ils ont une très faible puissance contre des alternatives dans laquelle la série se décompose en un processus stationnaire et une composante déterministe (sinusoïdale ou non) qui connaît une rupture déterministe.

## Test de cointégration à des fréquences saisonnières

Lorsque l'on travaille sur des séries multivariées, la forme du filtre que l'on obtient dépend du rang des matrices  $C \left( e^{i\omega_j} \right)_{j=0,1,\dots,k}$ . A la fréquence 0, cette condition de rang est liée à la possibilité qu'il existe des relations de cointégration entre des variables intégrées, c'est-à-dire des combinaisons linéaires de variables intégrées qui soient stationnaires. La cointégration à fréquence 0 a été introduite par [28], [30] et [20]. L'étude du rang de la matrice  $C \left( e^{i\omega_j} \right)$  lorsque



la fréquence  $\omega_j$  est saisonnière conduit à la notion de cointégration saisonnière introduite par [21] et [39]. Un théorème de représentation qui spécifie la forme des modèles autorégressifs satisfaits dans ces circonstances est donné par [31]. Le test sur le rang de ces matrices et l'estimation de ces vecteurs de cointégration peuvent être faits dans trois cadres. [43], [40] et [13] développent des approches par le maximum de vraisemblance, [12] par une analyse spectrale et [32] par une méthode de composantes principales.

## Démodulation et estimation

### Cas des processus SARIMA univariés

L'estimation de la composante saisonnière persistante est semblable à celle de la tendance dans une décomposition de Beveridge-Nelson standard, mais s'effectue sur le processus démodulé. Nous présentons le calcul lorsque le processus (complexe) est intégré à une seule fréquence i.e. ne possède qu'une seule racine saisonnière  $e^{i\omega}$  et que le polynôme associé à sa forme de Wold satisfait la condition 1. Partant de cette équation :

$$\delta_\omega(B)x_t = c(B)\varepsilon_t$$

nous pouvons décomposer ce processus en une composante persistante et une composante transitoire

$$x_t = \mu_\omega e^{-i\omega t} + c(e^{i\omega})S_\omega \varepsilon_t + \tilde{c}(B)\varepsilon_t$$

où  $\{\tilde{c}(B)\varepsilon_t\}$  est un processus complexe stationnaire et nous avons montré dans la partie 3 que la prévision linéaire de long-terme de ce processus conditionnelle à l'information disponible jusqu'à la date  $t$  prenait la forme :

$$\lim_{h \rightarrow +\infty} E_t e^{i\omega(t+h)} x_{t+h} = \mu_\omega + c(e^{i\omega}) \sum_{\tau=1}^t e^{i\omega\tau} \varepsilon_\tau$$

Si nous notons que

$$\begin{aligned} e^{i\omega(t+h)} x_{t+h} &= e^{i\omega t} x_t + \sum_{\tau=1}^h \left( e^{i\omega(t+\tau)} x_{t+\tau} - e^{i\omega(t+\tau-1)} x_{t+\tau-1} \right) \\ &= e^{i\omega t} x_t + \sum_{\tau=1}^h e^{i\omega(t+\tau)} \delta_\omega(B) x_{t+\tau} \end{aligned}$$

nous pouvons conclure que

$$\mu_\omega + c(e^{i\omega}) \sum_{\tau=1}^t e^{i\omega\tau} \varepsilon_\tau = e^{i\omega t} x_t + \sum_{\tau=1}^{+\infty} e^{i\omega(t+\tau)} E_t \delta_\omega(B) x_{t+\tau}$$

Une nouvelle opération de démodulation nous donne alors une estimation de la composante persistante à la fréquence  $\omega$  basée sur le passé du processus :

$$\mu_\omega e^{-i\omega t} + c(e^{i\omega})S_\omega \varepsilon_t = x_t + \sum_{\tau=1}^{+\infty} e^{i\omega\tau} E_t \delta_\omega(B) x_{t+\tau}$$

L'interprétation de la décomposition de Beveridge-Nelson en termes d'extraction de signal est présentée par [55]. Elle demeure valable.

Dans le cas général, quand le processus est intégré d'ordre 1 à plusieurs fréquences, nous avons

$$\Delta_x(B)x_t = c(B)\varepsilon_t$$

que nous réécrivons en utilisant l'équation de décomposition de produits d'opérateurs d'intégration :

$$\begin{aligned} x_t &= \sum_{j=0}^k \phi_j \Delta_{x,-j}(B) x_t \\ &= \sum_{j=0}^k \phi_j \left( \mu_{\omega_j} e^{-i\omega_j t} + c \left( e^{i\omega_j} \right) S_{\omega_j} \varepsilon_t + \tilde{c}_{\omega_j}(B) \varepsilon_t \right) \end{aligned}$$

L'estimation de la composante à la fréquence  $\omega_j$  est alors obtenue en travaillant sur le processus  $\{\Delta_{x,-j} x_t\}$  :

$$\mu_{\omega_j} e^{-i\omega_j t} + c \left( e^{i\omega_j} \right) S_{\omega_j} \varepsilon_t = \Delta_{x,-j} x_t + \sum_{\tau=1}^{+\infty} e^{i\omega_j \tau} E_t \Delta_x(B) x_{t+\tau}$$

La série corrigée des variations saisonnières est obtenue en ôtant de la série initiale la composante saisonnière :

$$\sum_{j=1}^k \frac{1}{\Delta_{x,-j} \left( e^{i\omega_j} \right)} \left( \Delta_{x,-j} x_t + \sum_{\tau=1}^{+\infty} e^{i\omega_j \tau} E_t \Delta_x(B) x_{t+\tau} \right)$$

Si nous nous plaçons dans la situation dans laquelle le processus  $\{x_t\}$  satisfait une représentation SARIMA. Le calcul de la décomposition de Beveridge-Nelson à fréquence 0 est présenté dans [44]. Nous étendons simplement cette méthode. Le processus de génération des données est

$$\Phi(B) \Phi_s(B^s) \Delta_x(B) x_t = \Theta(B) \Theta_s(B^s) \varepsilon_t$$

où  $d^0 \Phi = P, d^0 \Phi_s = P_s, d^0 \Theta = Q, d^0 \Theta_s = Q_s$ , et toutes les racines de ces polynômes sont de module strictement plus grand que 1. Nous considérons alors la forme compagnon associée aux processus de dimension  $(P + sP_s + Q + sQ_s) \times 1$  :

$$\begin{aligned} z_t &= \left( \Delta_x(B) x_t \quad \dots \quad \Delta_x(B) x_{t-P-sP_s+1} \quad \varepsilon_t \quad \dots \quad \varepsilon_{t-Q-sQ_s+1} \right)' \\ \zeta_t &= \left( \varepsilon_t \quad 0 \quad \dots \quad 0 \quad \varepsilon_t \quad 0 \quad \dots \quad 0 \right)' \end{aligned}$$

qui satisfont l'équation d'évolution

$$z_t = C z_{t-1} + \zeta_t$$

où  $C$  est la matrice de dimension  $(P + sP_s + Q + sQ_s) \times (P + sP_s + Q + sQ_s)$  dont la première ligne est obtenue en développant les produits de polynômes et en réécrivant sous une forme ARMA l'équation satisfaite par le processus  $\{x_t\}$  et les autres lignes sont composées de 1 ou de 0 de manière à assurer la validité de l'équation d'évolution de  $\{z_t\}$ . La prévision linéaire de ce processus à l'horizon  $h$  est donnée par :  $E_t z_{t+h} = C^h z_t$ , d'où si l'on appelle  $K$  un vecteur ligne de dimension  $1 \times (P + sP_s + Q + sQ_s)$  dont la première composante est égale à 1 et toutes les autres sont nulles :

$$\begin{aligned} K \sum_{h=1}^{+\infty} e^{i\omega \tau} E_t \Delta_x(B) x_{t+h} &= K \left( \sum_{h=1}^{+\infty} e^{i\omega h} C^h \right) z_t \\ &= K \left( I_{P+sP_s+Q+sQ_s} - e^{i\omega} C \right)^{-1} e^{i\omega} C z_t \end{aligned}$$

où l'inverse de  $(I_{P+sP_s+Q+sQ_s} - e^{i\omega}C)$  existe par l'hypothèse faite sur la position des racines des polynômes de la forme SARIMA. Il s'ensuit que lorsque le processus satisfait une représentation SARIMA, un estimateur de la composante saisonnière est donnée par :

$$x_t^s = \sum_{j=1}^k \frac{\Delta_{x,-j}x_t + K(I_{P+sP_s+Q+sQ_s} - e^{i\omega_j}C)^{-1} e^{i\omega_j}Cz_t}{\Delta_{x,-j}(e^{i\omega_j})}$$

et la série corrigée des variations saisonnières

$$x_t^* = x_t - \sum_{j=1}^k \frac{\Delta_{x,-j}x_t + K(I_{P+sP_s+Q+sQ_s} - e^{i\omega_j}C)^{-1} e^{i\omega_j}Cz_t}{\Delta_{x,-j}(e^{i\omega_j})}$$

### Cas des processus SARIMA multivariés cointégrés

Lorsque l'on travaille avec plusieurs séries simultanément, le raisonnement précédent demeure valable dans la mesure où nous n'exploitons que des propriétés linéaires. Afin de simplifier la présentation, nous gardons l'hypothèse de processus  $\{x_t\}$  de dimension  $n$  dont les composantes sont toutes intégrées à l'ordre 1 aux mêmes fréquences et satisfaisant le processus de génération des données considéré précédemment. Nous avons donc

$$\Delta_x(B)x_t = C(B)\varepsilon_t$$

et sous la Condition 2, nous obtenons un estimateur des composantes saisonnières persistantes à ôter de la série initiale sous la forme suivante :

$$\sum_{j=1}^k \frac{1}{\Delta_{x,-j}(e^{i\omega_j})} \left( \Delta_{x,-j}x_t + \sum_{\tau=1}^{+\infty} e^{i\omega_j\tau} E_t \Delta_x(B)x_{t+\tau} \right)$$

La difficulté se trouve dans le calcul des prévisions et la spécification du modèle vectoriel suivi par le processus  $\{\Delta_x(B)x_t\}$ . En l'absence de cointégration à toutes les fréquences, l'approche présentée dans le cas d'un processus univarié peut s'étendre aux processus VARMA et nécessite l'usage d'une grande matrice compagnon de dimensions  $(n(P+sP_s+Q+sQ_s)) \times (n(P+sP_s+Q+sQ_s))$  où  $P, P_s, Q$  et  $Q_s$  sont les degrés des polynôme matriciels associés à la représentation VARMA, appliquée à un processus  $z_t$  de dimension  $(n(P+sP_s+Q+sQ_s)) \times 1$ . Lorsqu'il existe des relations de cointégration entre les composantes de  $\{x_t\}$ , une représentation de type modèle à correction d'erreur doit être considérée et la matrice compagnon doit être contrainte de façon à assurer la convergence des estimateurs des prévisions à long-terme. Nous détaillons le cas d'un modèle à correction d'erreur saisonnier avec des relations de cointégration à toutes les fréquences. Le modèle est donné pour le processus  $\tilde{x}_t = \Delta_x(B)x_t$  et a la forme suivante (cf [31]) :

$$\tilde{x}_t = \sum_{j=1}^{\lfloor \frac{s}{2} \rfloor} \left( \gamma_j \theta_j' S_{\omega_j} \tilde{x}_{t-1} + \bar{\gamma}_j \bar{\theta}_j' S_{-\omega_j} \tilde{x}_{t-1} \right) + \gamma_0 \theta_0' S_0 \tilde{x}_{t-1} + \gamma_\pi \theta_\pi' S_\pi \tilde{x}_{t-1} + \sum_{k=1}^p \Phi_k \tilde{x}_{t-k} + \varepsilon_t$$

où  $\{\varepsilon_t\}$  est un bruit blanc multivarié et les matrices  $\gamma_j$  et  $\theta_j$  sont de plein rang et de dimension  $(n \times r_j)$  pour  $j=1, \dots, \left\lfloor \frac{s}{2} \right\rfloor$  et respectivement  $\gamma_0, \theta_0$  et  $\gamma_\pi, \theta_\pi$  sont de plein rang et de dimension  $(n \times r_0)$  et  $(n \times r_\pi)$ . Nous utilisons les relations

$$\forall \omega \in [0, \pi], S_\omega \tilde{x}_t = \tilde{x}_t + e^{-i\omega} S_\omega \tilde{x}_{t-1}$$

pour réécrire la dernière équation sous la forme :

$$\tilde{x}_t = \sum_{j=1}^{\left\lfloor \frac{s}{2} \right\rfloor} \left( \gamma_j \theta_j' e^{-i\omega_j} S_{\omega_j} \tilde{x}_{t-2} + \bar{\gamma}_j \bar{\theta}_j' e^{i\omega_j} S_{-\omega_j} \tilde{x}_{t-2} \right) + \gamma_0 \theta_0' S_0 \tilde{x}_{t-2} - \gamma_\pi \theta_\pi' S_\pi \tilde{x}_{t-2} + \tilde{\Phi}_1 \tilde{x}_{t-1} + \sum_{k=2}^p \Phi_k \tilde{x}_{t-k} + \varepsilon_t$$

où  $\tilde{\Phi}_1 = \Phi_1 + \sum_{j=1}^{\left\lfloor \frac{s}{2} \right\rfloor} (\gamma_j \theta_j' + \bar{\gamma}_j \bar{\theta}_j') + \gamma_0 \theta_0' + \gamma_\pi \theta_\pi'$ . Nous obtenons alors la représentation compagnon :

$$z_t = C z_{t-1} + \zeta_t$$

avec

$$z_t = \left( \tilde{x}_t' \quad \dots \quad (\theta_j' S_{\omega_j} \tilde{x}_{t-1})' \quad (\bar{\theta}_j' S_{-\omega_j} \tilde{x}_{t-1})' \quad \dots \quad (\theta_0' S_0 \tilde{x}_{t-1})' \quad (\theta_\pi' S_\pi \tilde{x}_{t-1})' \quad \tilde{x}_{t-1}' \quad \dots \quad \tilde{x}_{t-p}' \right)'$$

et

$$C = \begin{pmatrix} \tilde{\Phi}_1 & \dots & e^{-i\omega_j} \gamma_j & e^{i\omega_j} \bar{\gamma}_j & \dots & \gamma_0 & -\gamma_\pi & \Phi_2 & \dots & \Phi_p \\ \vdots & \ddots & 0 & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ \theta_j' & 0 & e^{-i\omega_j} I_{r_j} & 0 & \dots & \dots & \dots & \dots & \dots & \vdots \\ \bar{\theta}_j' & \vdots & 0 & e^{i\omega_j} I_{r_j} & \ddots & \dots & \dots & \dots & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \theta_0' & \vdots & \vdots & \vdots & \ddots & I_{r_0} & \ddots & \dots & \dots & \vdots \\ \theta_\pi' & \vdots & \vdots & \vdots & \ddots & \ddots & -I_{r_\pi} & \ddots & \dots & \vdots \\ I_n & \vdots & \vdots & \vdots & \vdots & \vdots & 0 & 0 & \ddots & \vdots \\ 0 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & I_n & 0 & 0 \\ 0 & \dots & \dots & \dots & \dots & \dots & \dots & 0 & I_{n(p-2)} & 0 \end{pmatrix}$$

La formule obtenue dans le cas univarié peut être appliquée avec cette matrice compagnon.

## Première illustration : consommation mensuelle en produits manufacturés des ménages français

Nous illustrons le cadre précédent sur la série mensuelle de consommation des ménages en produits manufacturés. Nous devons dans un premier temps tester la présence de racines unitaires saisonnières ou non dans le processus puis nous utilisons une approximation autorégressive AR(p) du processus pour construire des estimateurs des composantes persistantes. Nous comparons ensuite la série corrigée que nous obtenons à celle proposée par l'INSEE obtenue avec la procédure X11-ARIMA de SAS et sur laquelle des corrections pour jours ouvrés ont aussi été appliquées.

Le test de racine unité utilisé est le test introduit dans [33] qui étend le cadre des tests presque efficaces aux tests de racines saisonnières. Les résultats sont donnés dans le tableau 1 ci-dessous :

Consommation des ménages en produits manufacturés			
Fréquence	Statistique de test	Seuil	Décision
0	67,97	3,11	Acceptation
$\frac{\pi}{6}$	5,15	3,22	Acceptation
$\frac{\pi}{3}$	14,87	3,22	Acceptation
$\frac{\pi}{2}$	3,86	3,22	Acceptation
$\frac{2\pi}{3}$	3,85	3,22	Acceptation
$\frac{5\pi}{6}$	11,37	3,22	Acceptation
$\pi$	1,29	3,11	Rejet

**Tableau 1**

Par souci de simplicité, nous appliquons l'opérateur de différenciation saisonnière sans prendre en compte l'absence de la racine saisonnière à la fréquence  $\pi$  et approchons le processus de génération des données par un modèle autorégressif.

La Figure 2 présente la série brute à traiter et la série corrigée des variations saisonnières obtenue dans le cadre présenté ici. Le Figure 3 permet de comparer les deux corrections, celle proposé par l'INSEE à partir de l'usage de la procédure X11-ARIMA et celle décrite ici en termes d'analyse spectrale. Nous avons estimé la cohérence spectrale et le déphasage spectral entre la série brute et chacune des séries désaisonnalisées (INSEE et nouvelle méthode). Alors qu'en termes de cohérence, la nouvelle méthode présente une meilleure proximité à la série brute, très proche de 1 à toutes les fréquences, cette relation de proximité est inversée lorsque l'on considère le déphasage. La série CVS de l'INSEE apparaît en phase à presque toutes les fréquences avec la série brute à la différence de la série issue de la nouvelle méthode. Le décalage de phase estimé entre les deux séries (brute et corrigée par la nouvelle méthode) est proche d'une période aux basses fréquences. Ceci pourrait se traduire par une chronologie des mouvements de moyens court-terme de la série obtenue par la nouvelle méthode en retard sur la série CVS INSEE. La Figure 4 montre qu'il n'en est rien. Les deux courbes présentent les mêmes inflexions, il n'y pas de différence notable sur la chronologie, excepté peut-être sur la fin de la série. Il faut rappeler que le traitement fait sur les extrémités de série par X11-ARIMA n'est pas homogène à celui mis en œuvre au milieu de l'échantillon à la différence de la méthode introduite ici. De même, la série obtenue dans le cadre introduit ici est moins variable sur presque tout l'échantillon et ceci est renforcé sur la fin de l'échantillon, la variabilité de la série issue de X11-ARIMA est nettement supérieure. Il faut savoir que la série de consommation diffusée par l'INSEE est obtenue par agrégation de plusieurs séries désaisonnalisées indépendamment. Le diagnostic sur la chronologie et la variabilité demeure inchangé lorsque l'on analyse les taux de croissance (Figure 5).

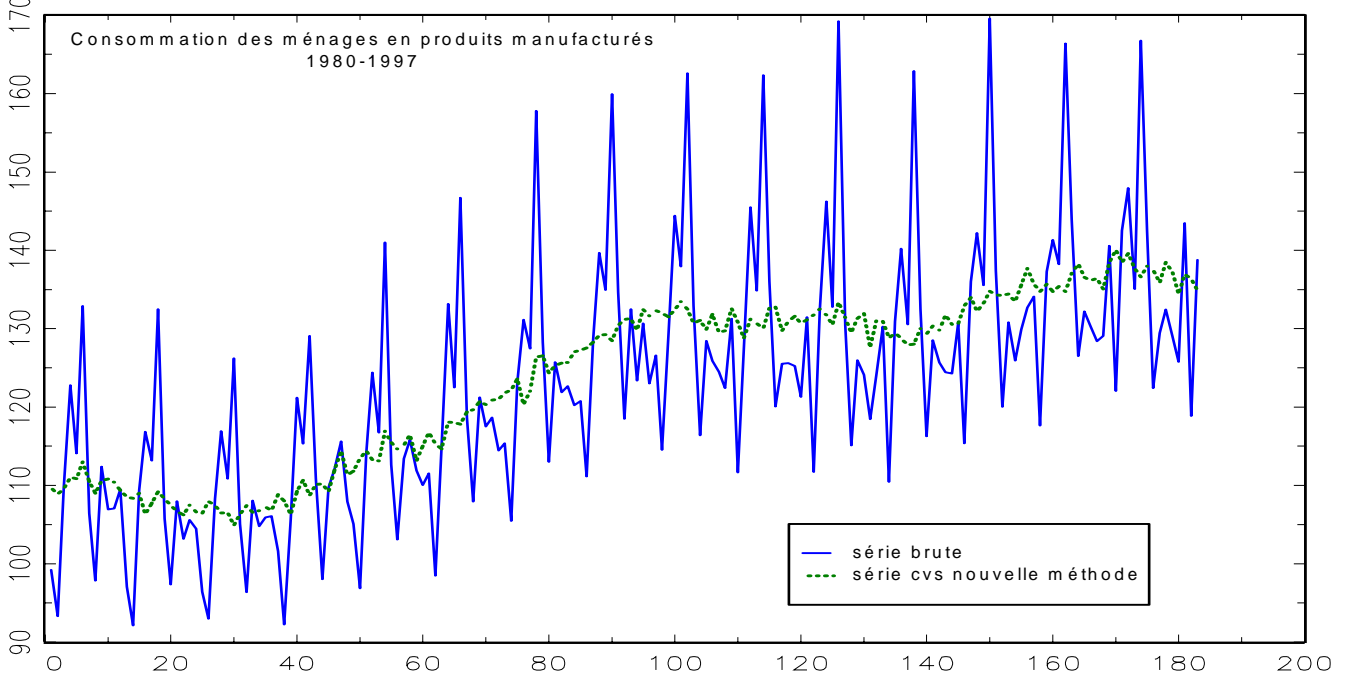


Figure 2

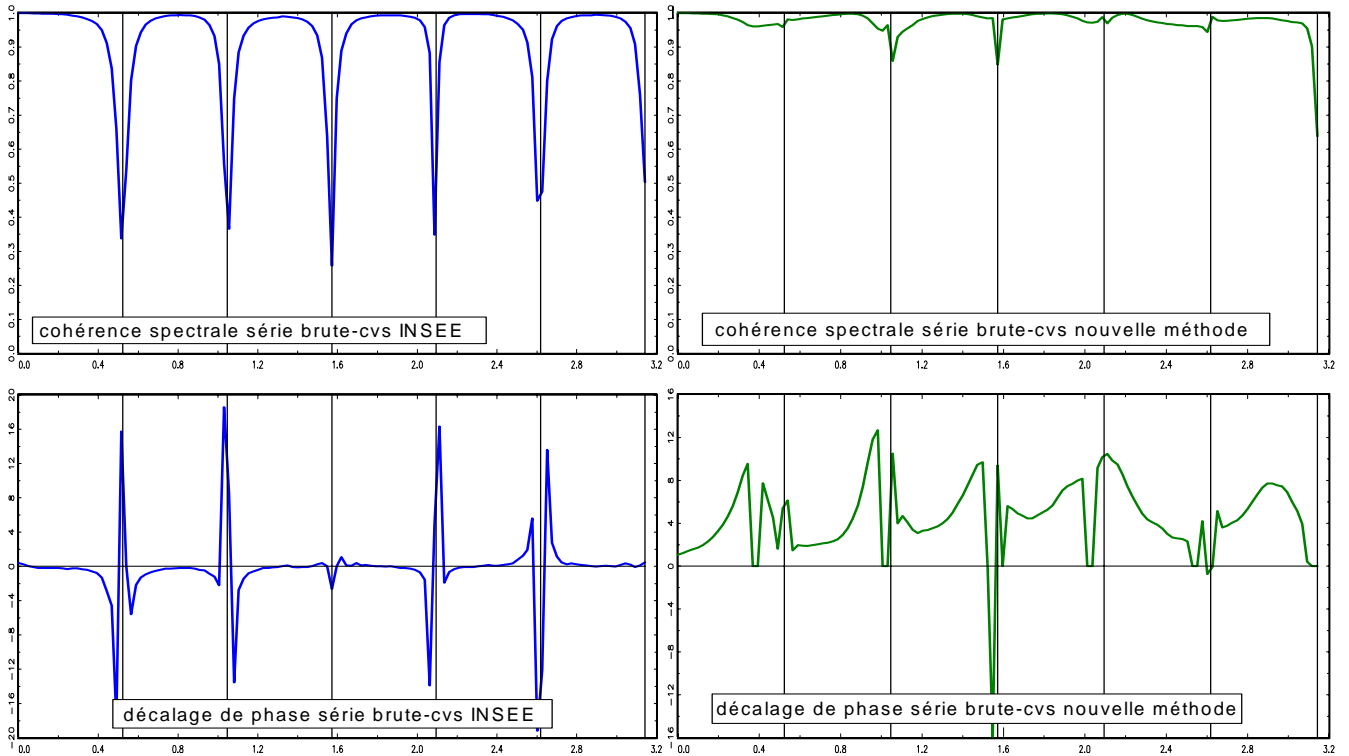


Figure 3

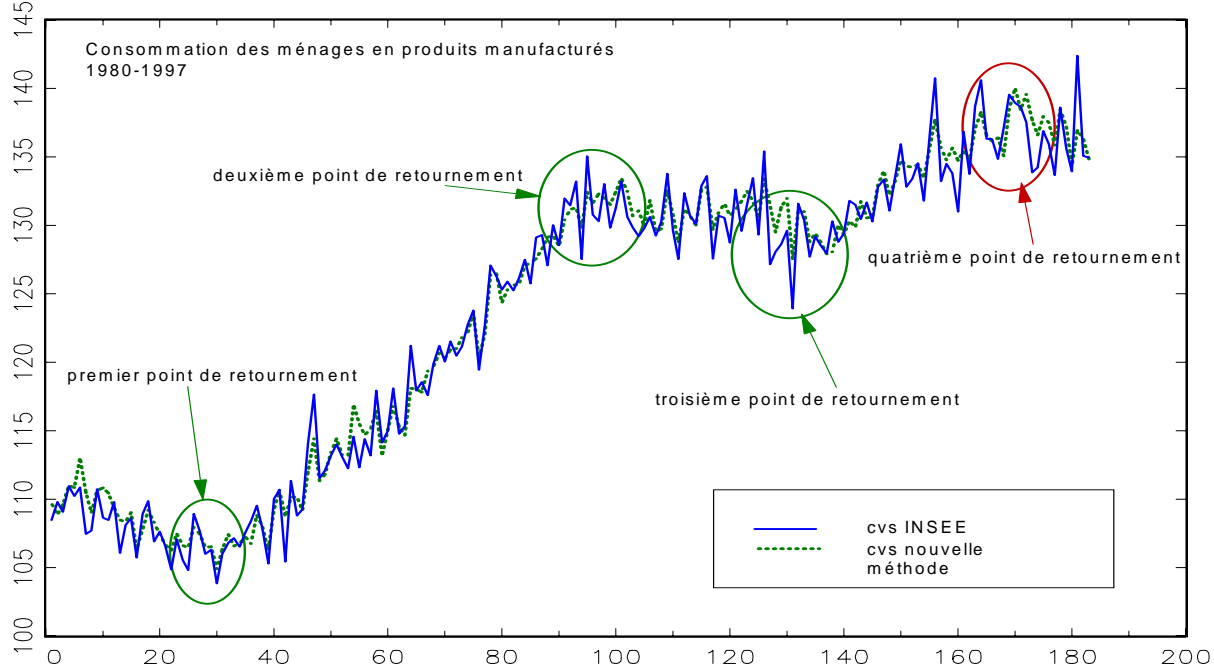


Figure 4

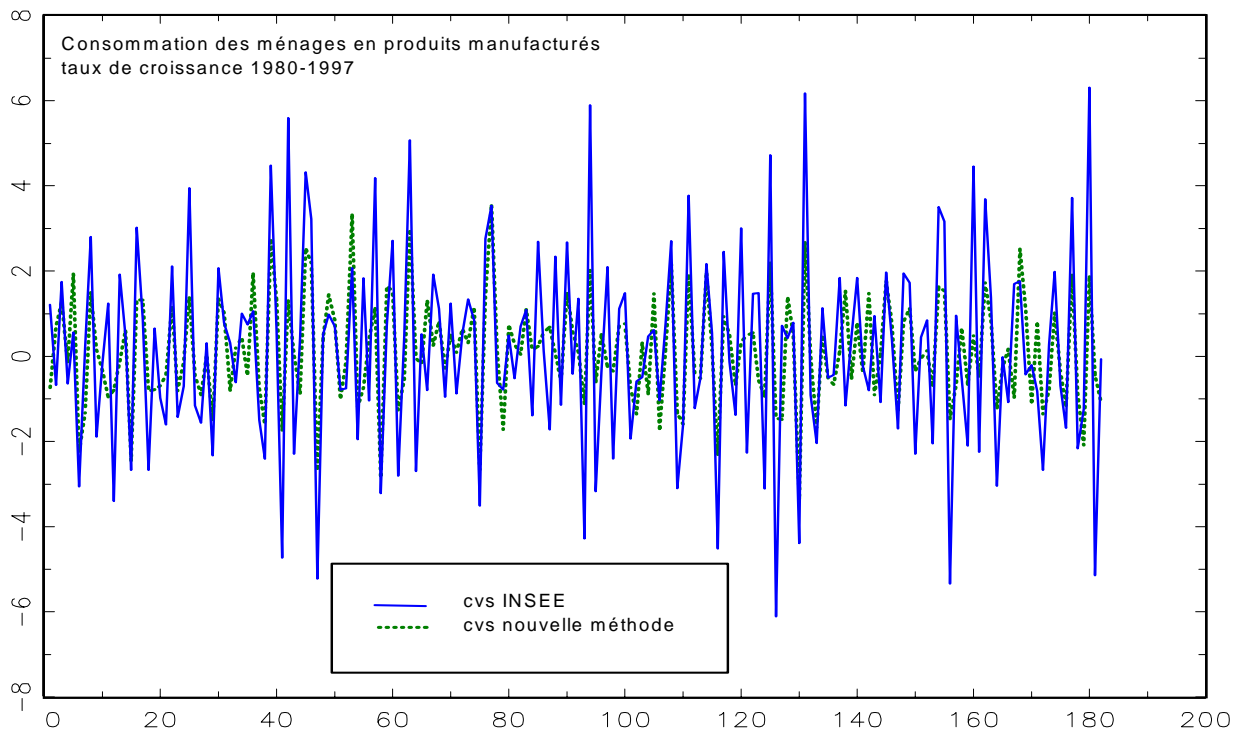


Figure 5

## Seconde illustration : consommation privée trimestrielle au Japon

Nous travaillons sur la série trimestrielle de la consommation privée au Japon. Nous testons dans un premier temps la présence de racines unitaires saisonnières dans le processus de génération des données puis nous utilisons une approximation autorégressive AR(p) du processus pour construire des estimateurs des composantes persistantes. Nous comparons ensuite la série corrigée que nous obtenons à celle proposée par l'INS japonais obtenue avec la procédure X11-ARIMA.

Nous utilisons à nouveau les tests de racine unité introduits dans [33]. Les résultats sont donnés dans le tableau ci-dessous :

Consommation des ménages en produits manufacturés			
Fréquence	Statistique de test	Seuil	Décision
0	81,04	3,11	Acceptation
$\frac{\pi}{2}$	7,20	3,22	Acceptation
$\pi$	4,89	3,11	Acceptation

Nous devons donc appliquer l'opérateur de différenciation saisonnière  $(1 - B^4)$  pour obtenir un processus stationnaire du second ordre et approchons le processus de génération des données par un modèle autorégressif.

La Figure 6 présente la série brute à traiter et les séries corrigées des variations saisonnières obtenue dans le cadre présenté ici et celle proposée par l'INS japonais. La proximité de ces deux dernières séries est visuellement grande. La Figure 7 donne des éléments de comparaison en termes d'analyse spectrale. Le résultat observé dans l'exemple précédent demeure : la méthode introduite ici donne une série corrigée des variations saisonnières plus proche de la série brute en termes de cohérence mais plus éloignée en termes de phase. Le déphasage pour les cycles longs est néanmoins très faible (nul à long terme). Le dernier graphique (Figure 8) permet une comparaison des chroniques de taux de croissance pour les deux séries corrigées des variations saisonnières.

## Conclusion

L'approche que nous avons suivie ici est naturelle dans un cadre linéaire et dans un travail dans l'espace des temps. Elle vise à estimer au mieux les composantes persistantes saisonnières et à les ôter. La définition des composantes persistantes est différente de ce que considère les méthodes usuelles de correction des variations saisonnières. Elle est en ligne avec le cadre utilisé et n'impose pas de contraintes supplémentaires d'identification. Cette approche est valide sous des conditions relativement faibles. Le traitement effectué dépend de la série étudiée. Lorsque l'on traite plusieurs séries simultanément, un ajustement joint permet de traiter les sources de saisonnalité communes et ainsi d'obtenir un meilleur traitement de la saisonnalité mais aussi de respecter les liens qui existent entre les différentes séries aux autres fréquences.

La mise en œuvre de la méthode est relativement souple, plusieurs méthodes d'estimation peuvent être utilisées. En particulier des méthodes robustes peuvent être employées pour limiter les effets des points aberrants et de leur caractérisation. Comme les composantes sont construites à partir d'un modèle estimé et de l'information récente, des méthodes de Monte-Carlo ou de bootstrap ([54]) peuvent être employées pour calculer un intervalle de confiance de la série corrigée des variations saisonnières.



Le problème principal de cette approche réside dans ses propriétés spectrales. Le principe de la correction n'est pas pensé en termes de propriétés spectrales. Le filtre asymétrique qu'elle implique peut entraîner un déphasage à certaines fréquences. Néanmoins, les exemples que nous avons pu analyser ne font pas apparaître un net décalage dans la chronologie des points d'inflexion entre un traitement classique et cette nouvelle approche. Des outils d'analyse doivent être développés pour permettre une appréciation de ces problèmes suivant la nature du modèle considéré comme processus de génération des données. En ce qui concerne le traitement de la fin des échantillons, la méthode introduite ici produit un cadre pour le calcul des filtres asymétriques adapté à chaque modèle considéré et en cela diffère de celui retenu usuellement ([41]) qui cherche à minimiser les révisions en fin d'échantillon sous des hypothèses *ad hoc* quant à la nature du processus à traiter.

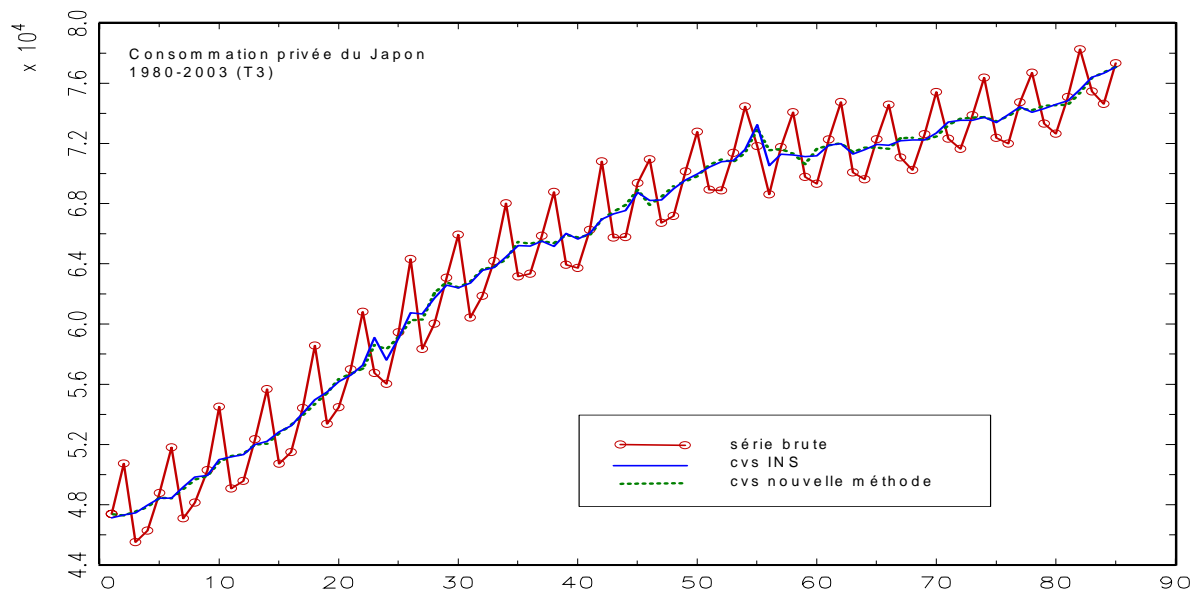


Figure 6

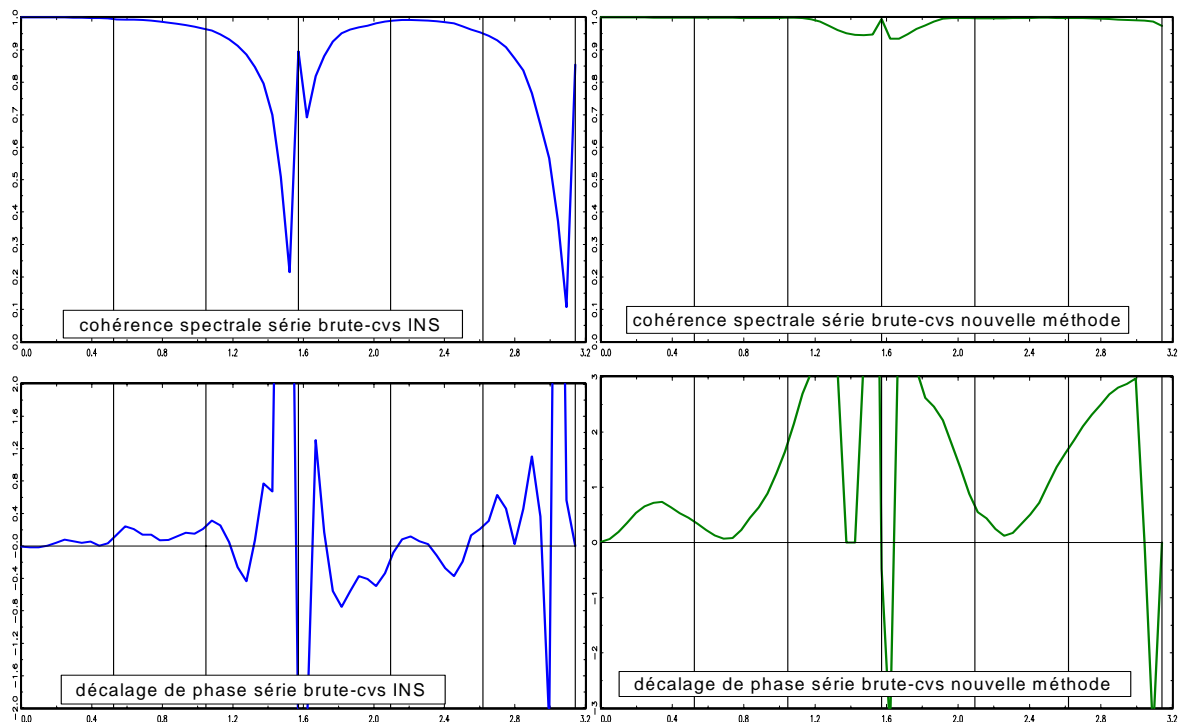


Figure 7

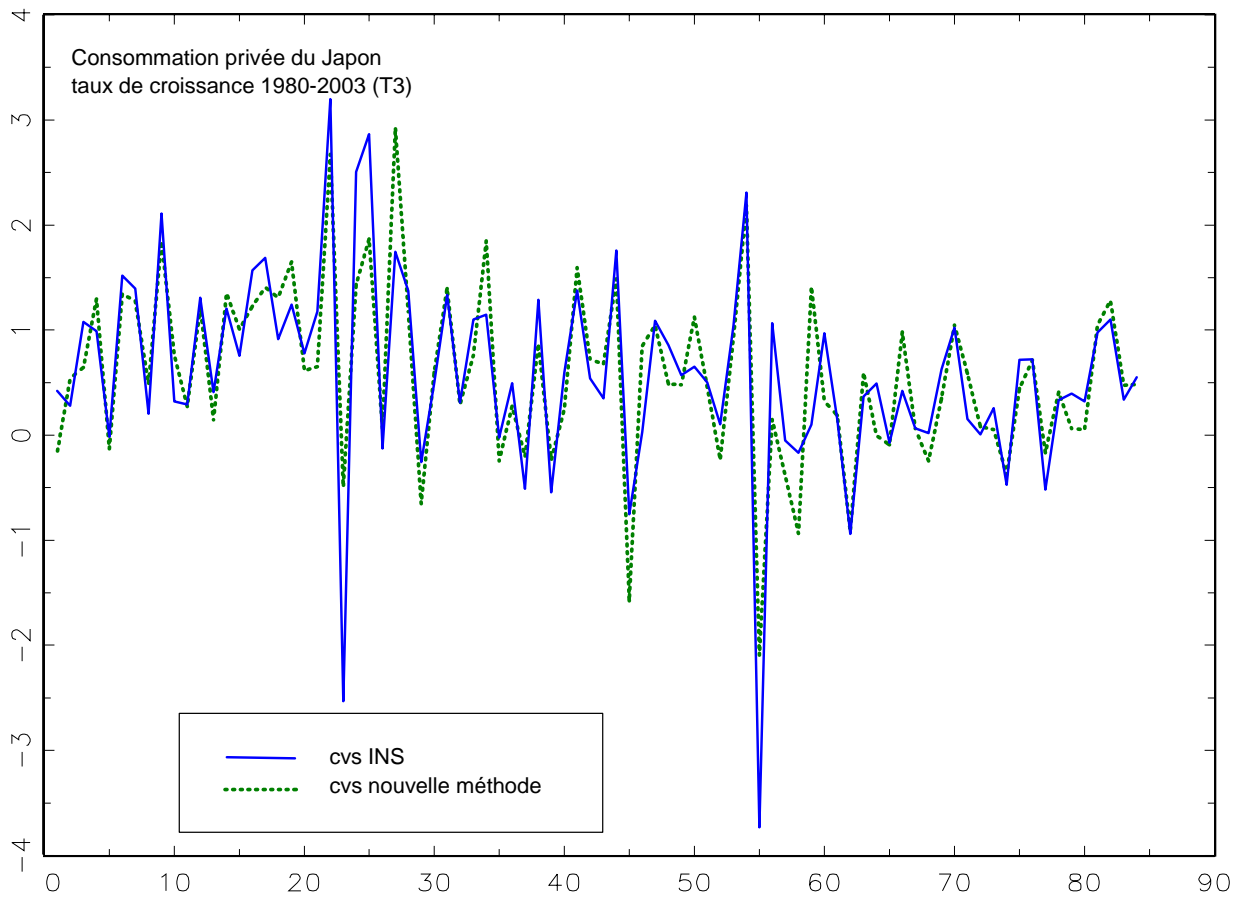


Figure 8

## Bibliographie

- [1] Akaike (1980), 'Seasonal Adjustment by Bayesian Modeling', *Journal of Time Series Analysis*, 1, 1-13
- [2] Beaulieu J.J et J.A.Miron (1993), 'Seasonal Unit Roots in Aggregate U.S. Data', *Journal of Econometrics*, 55, 305-328.
- [3] Bell, W.R. et S.C. Hillmer (1984), 'Issues Involved with the Seasonal Adjustment of Economic Time Series', *Journal of Business and Economic Statistics*, 2, 291-320.
- [4] Beveridge S. et C.R. Nelson (1981) 'A new approach to the decomposition of economic time series into permanent and transitory components with particular attention to the measurement of the "Business Cycle"', *Journal of Monetary Economics*, 7, 151-174
- [5] Box, G.E.P. et G.M. Jenkins (1970), *Time Series Forecasting and Control*, San Francisco, Holden Day.
- [6] Box, G.E.P., S.C. Hillmer et G.C. Tiao (1978), 'Analysis and Modeling of seasonal time series', in *Seasonal Analysis of Economic Time Series*, Ed. A. Zellner, Washington D.C., U.S. Bureau of Census, 309-334
- [7] Box, G.E.P. et G.C. Tiao (1975), 'Intervention Analysis with applications to Economic and Environmental Problems', *Journal of the American Statistical Association*, 70, 70-79.
- [8] Breitung, J. et P.H. Franses (1998), 'On Phillips-Perron-type Tests for Seasonal Unit Roots', *Econometric Theory*, 14, 200-221.
- [9] Burman J.P. (1980), 'Seasonal Adjustment by Signal Extraction', *Journal of the Royal Statistical Society*, Serie A, 143, 321-337.
- [10] Campbell, J.Y. et N.G. Mankiw (1987), 'Permanent and transitory components in macroeconomic fluctuations', *American Economic Review*, 77 (2), 111-117.
- [11] Cochrane J. (1988) 'How Big is the Random Walk in GNP?' *Journal of Political Economy* 96 893-920
- [12] Cubadda G. (1995) "A note on testing for seasonal cointegration using principal components in the frequency domain", *Journal of Time Series Analysis*, 16, 499-508
- [13] Cubadda G. (2001), 'Complex Reduced Rank models for seasonally cointegrated time series', *Oxford Bulletin of Economics and Statistics*, 63, 497-511.
- [14] Dagum E.B.(1975), 'Seasonal Factor forecasts from ARIMA models', International Statistical Institute, Proceedings of 40th session, Vol.3, Warsaw, 206-219
- [15] Dagum E.B.(1978), 'Modelling, Forecasting and Seasonally Adjusting Economic Time Series with the X11-ARIMA method', *The Statistician*, 27, 203-216
- [16] Dickey D.A. et W. Fuller (1979), 'Distribution of the estimators for autoregressive time series with a unit root', *Journal of the American Statistical Association*, 74, 427-431.
- [17] Dickey D.A., D.P. Hasza et W. Fuller (1984), 'Testing for Unit Roots in Seasonal Time Series', *Journal of the American Statistical Association*, 79, 355-367.
- [18] Elliott G., T.J. Rothenberg et J.H. Stock (1996), 'Efficient Tests for an Autoregressive Unit Root', *Econometrica*, 64 (4), 813-836.
- [19] Engle, R.F. (1978), 'Estimating Structural Models of Seasonality', in *Seasonal Analysis of Economic Time Series*, Ed. A. Zellner, Washington D.C., U.S. Bureau of Census, 281-297
- [20] Engle R.F. et C.W.J Granger (1987), 'Co-integration and Error Correction: Representation, Estimation, Testing', *Econometrica*, 55, 251-276
- [21] Engle R.F., C.W.J Granger et J. Hallman (1989), 'Merging Short- and Long-Run Forecasts: An Application of Seasonal Co-Integration to Monthly Electricity Sales Forecasting', *Journal of Econometrics*, 40, 45-62

- [22] Franses P.H. et T.J.Vogelsang (1998), 'On seasonal cycles, unit roots, and mean shifts', *The Review of Economics and Statistics*, 131-240
- [23] Geweke, J.(1979), 'The Temporal and Sectoral Aggregation of Seasonally Adjusted Time Series', in *Seasonal Analysis of Economic Time Series*, Ed. A. Zellner, Washington D.C., U.S. Bureau of Census, 411-427
- [24] Ghysels E. (1988) 'A Study Toward a Dynamic Theory of Seasonality for Economic Time Series', *Journal of the American Statistical Association*, 83, (401),168-172.
- [25] Ghysels E., H.S. Lee et J. Noh (1994), 'Testing for Unit Roots in Seasonal Time Series : Some Theoretical Extensions and a Monte Carlo Investigation', *Journal of Econometrics*, 62 , 415-442.
- [26] Gomez V. et A.Maravall (1996), 'Programs SEATS and TRAMO: Instructions for the User' working paper n°9628, Bank of Spain.
- [27] Gomez V, et J.Breitung (1999), 'The Beveridge-Nelson Decomposition : A Different Perspective with New results', *Journal of Time Series Analysis*, 20 (5), 527-535
- [28] Granger, C.W.J. (1981), 'Some Properties of Time Series Data and their Use in Econometric Model Specification', *Journal of Econometrics*, 16, 121-130
- [29] Granger C.W.J. et H. Hatanaka (1964), *Spectral Analysis of Economic Time Series*, Princeton University Press
- [30] Granger, C.W.J. et A.A. Weiss (1983), 'Time Series Analysis of Error Correcting Models', in *Studies in Econometrics, Time Series and Multivariate Statistics*, Ed. S. Karlin and T. Amemiya, Academic Press NY, 225-278
- [31] Gregoir S. (1999a), 'Multivariate Time Series with Various Hidden Unit Roots: Part I : Integral Operator Algebra and Representation Theory', *Econometric Theory*, 15, 435-468.
- [32] Gregoir S. (1999b), 'Multivariate Time Series with Various Hidden Unit Roots: Part II : Estimation and Testing', *Econometric Theory*, 15, 469-518.
- [33] Gregoir S (2005), 'Efficient Tests for the Presence of a Pair of Complex conjugate Unit Roots in Real Time Series', forthcoming *Journal of Econometrics*
- [34] Grether D.M. et M. Nerlove (1970), 'Some Properties of "Optimal" Seasonal Adjustment', *Econometrica*, 38(5), 682-703
- [35] Harvey, A.C. (1989), *Forecasting Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press.
- [36] Harvey A.C. et P.H.Todd (1983) 'Forecasting Economic Time Series with Structural and Box-Jenkins Models: A case study', *Journal of Business and Economics Forecasting*, 1, 299-307
- [37] Hasza D.P. et W. Fuller (1982), 'Testing for Nonstationarity Parameter Specifications: Seasonal Time Series Models', *The Annals of Statistics*, 10, 1209-1216.
- [38] Hillmer S.C. et G.C. Tiao (1982), 'An ARIMA-model-based Approach to Seasonal Adjustment', *Journal of the American Statistical Association*, 77, 63-70.
- [39] Hylleberg S., R.F.Engle, C.W.J.Granger et B.S.Yoo (1991), 'Seasonal Integration and Cointegration', *Journal of Econometrics*, 44, 215-238.
- [40] Johansen S. et E.Schaumburg (1988) 'Likelihood Analysis of Seasonal Cointegration', *Journal of Econometrics*, 88(2), 301-339
- [41] Kenny P.B. et J.Durbin (1982), 'Local Trend Estimation and Seasonal Adjustment of Economic and social Time Series' (with discussion), *Journal of the Royal Statistical Society, Serie A*, 145, 1-41
- [42] Kitagawa, G. et W. Gersch (1984), 'A smoothness Priors-state Space Modeling of time Series with Trend and Seasonality ', *Journal of the American Statistical Association*, 79, 378-389.
- [43] Lee, H.S. (1992) 'Maximum Likelihood Inference on Cointegration and Seasonal Cointegration', *Journal of Econometrics*, 54, 1-47

- [44] Newbold, P. (1990), 'Precise and efficient computation of the Beveridge-Nelson decomposition of economic time series', *Journal of Monetary Economics*, 26, 453-457
- [45] Perron P. (1989), 'The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis' *Econometrica*, 57(6), 1361-1401.
- [46] Plosser C. (1979) 'A time series analysis of seasonality in econometric models', in *Seasonal Analysis of Economic Time Series*, Ed. A.Zellner, Washington D.C., U.S. Department of Commerce, Bureau of Census, 366-397
- [47] Porter R.D. (1975) 'Multiple time series containing unobserved components', Special Studies Paper, n°65, Federal Reserve Board
- [48] Proietti T (1996 ) 'Persistence of shocks on seasonal processes', *Journal of Applied Econometrics*, 11, 383-398
- [49] Quah D. (1992), 'The Relative Importance of Permanent and Transitory Components: Identification and Some Theoretical Bounds', *Econometrica* 60(1), 107-118
- [50] Rodrigues P.M. (2002), 'On LM type tests for seasonal unit roots in quarterly data', *Econometrics Journal*, 5, 176-195
- [51] Sims, C.A. (1974), 'Seasonality in Regression', *Journal of the American Statistical Association*, 69, 618-626.
- [52] Tanaka K. (1996), *Time Series Analysis*, Wiley-Interscience New-York U.S.A.
- [53] Taylor A.M.R. (2003), 'On the asymptotic properties of some seasonal unit roots tests', *Econometric Theory*, 19, 311-321
- [54] Thombs L.A. et W.R. Schucany (1990) 'Bootstrap Prediction Intervals for Autoregression', *Journal of the American Statistical Association*, 85, (410), 486-492
- [55] Watson M. (1986), 'Univariate detrending methods with stochastic trends', *Journal of Monetary Economics*, 18, 49-75
- [56] Wallis, K.F. (1974), 'Seasonal Adjustment and the Relation between Variables', *Journal of the American Statistical Association*, 69, 18-32.
- [57] Wallis, K.F. (1976), 'Seasonal Adjustment and Multiple Time Series Analysis', in *Seasonal Analysis of Economic Time Series*, Ed. A.Zellner, Washington D.C., U.S. Department of Commerce, Bureau of Census, 366-397

