

## Article

# Un test non paramétrique pour la saisonnalité résiduelle

par Tucker McElroy et Scott Holan

Juin 2009



# Un test non paramétrique pour la saisonnalité résiduelle

Tucker McElroy et Scott Holan<sup>1</sup>

## Résumé

La présence de pics dans le spectre d'un processus stationnaire signale l'existence de phénomènes périodiques stochastiques, tels que l'effet saisonnier. Nous proposons une mesure de ces pics spectraux et un test de détection de leur présence qui s'appuie sur l'évaluation de leur pente et de leur convexité agrégées. Notre méthode est élaborée de manière non paramétrique et peut donc être utile durant l'analyse préliminaire d'une série. Elle peut aussi servir à détecter la présence d'une saisonnalité résiduelle dans les données désaisonnalisées. Nous étudions le test diagnostique au moyen d'une simulation et d'une étude de cas à grande échelle portant sur des données provenant du U.S. Census Bureau et de l'Organisation de coopération et de développement économiques (OCDE).

Mots clés : Tests multiples ; estimation non paramétrique de la densité ; désaisonnalisation ; densité spectrale.

## 1. Introduction

La présence d'un pic dans le spectre d'un processus stationnaire est un indice de comportement périodique, tel que la saisonnalité ou un effet de jour ouvrable. Dans les domaines du génie et de l'économétrie, de nombreux auteurs s'intéressent à la détection de ces pics, car un nœud spectral prononcé exerce une forte influence sur la dynamique du processus stochastique. Un pic est indicatif d'une gamme de fréquences dont la contribution à la variance globale du processus stochastique est relativement importante. Si la force du pic, évaluée par sa hauteur et sa largeur relativement aux valeurs voisines, est suffisamment prononcée, tout modèle de la dynamique du processus ne tenant pas compte des périodicités correspondantes sera spécifié incorrectement. Les ingénieurs ainsi que les économétristes pourraient vouloir extraire ou prévoir le signal, tâche qui, dans l'un et l'autre cas, est sensible à la présence de pics spectraux.

Par pic spectral, nous entendons une région de la densité spectrale dont la masse spectrale est supérieure à celle de ses voisines directes ; une définition plus précise est présentée plus loin. Dans le contexte des applications qui nous intéressent, les pics ont une hauteur finie et correspondent donc à des effets périodiques stochastiques dans un processus stationnaire. Par conséquent, notre principale préoccupation n'est pas la détection des effets périodiques fixes (déterministes) ni celle des phénomènes périodiques non stationnaires (quoique nous fassions certaines extensions à ce cas à la section 3.4) qui, dans les deux cas, correspondent à des pics spectraux de hauteur infinie. L'abondante littérature traitant de la détection des effets fixes est discutée dans Priestley (1981) ; dans nos

applications, au lieu d'être fixes, les aspects périodiques évoluent au cours du temps.

Le présent article est axé sur l'application à la désaisonnalisation. Plus précisément, nous nous intéressons aux pics dits saisonniers, qui peuvent survenir aux fréquences saisonnières  $\pi/6$ ,  $2\pi/6$ ,  $3\pi/6$ ,  $4\pi/6$ ,  $5\pi/6$  et  $6\pi/6$  (en supposant que l'intervalle d'échantillonnage est mensuel). Dans le cas des statistiques fédérales, la détection de la saisonnalité et de la saisonnalité résiduelle pose un problème pratique important et le spectre est un outil logique pour l'effectuer. L'approche basée sur le domaine fréquentiel pour détecter et analyser la saisonnalité est très répandue, parce qu'elle offre un moyen tout naturel de visualiser un comportement quasi périodique. En fait, la saisonnalité est, informellement parlant, caractérisée par la présence d'au moins un pic saisonnier dans le spectre (Nerlove 1964). Les méthodes axées sur le domaine fréquentiel sont maintenant employées dans X-12-ARIMA (Findley, Monsell, Bell, Otto et Chen 1998) et font partie de TRAMO-SEATS (Maravall et Caporello 2004), qui sont les deux programmes de désaisonnalisation les plus répandus mis à la disposition du public. Notons que les méthodes du domaine fréquentiel peuvent être appliquées selon une approche paramétrique (c'est-à-dire fondée sur un modèle) ou non paramétrique. Nous mettons au point un diagnostic non paramétrique appellable pour déterminer l'efficacité de toute procédure de désaisonnalisation, qu'elle soit fondée sur un modèle ou non paramétrique. Comme l'ont souligné Findley, Monsell, Bell, Otto et Chen (1998), l'utilisation de fonctions périodiques fixes seulement pour modéliser la saisonnalité est habituellement inadéquate pour les données économiques (voir aussi la discussion dans Bell et Hillmer 1984).

1. Tucker McElroy, Statistical Research Division, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100. Courriel : tucker.s.mcelroy@census.gov; Scott Holan, Department of Statistics, University of Missouri-Columbia, 146 Middlebush Hall, Columbia, MO, 65211-6100. Courriel : holans@missouri.edu.

La présence de pics spectraux à des fréquences saisonnières dans une série désaisonnalisée peut signifier que les filtres saisonniers sont inadéquats – voir Soukup et Findley (1999) pour une discussion. Au minimum, les filtres de désaisonnalisation devraient éliminer la saisonnalité *non stationnaire* et tout effet périodique fixe, c'est-à-dire les phénomènes qui, dans la série observée, contribuent à la formation d'un pôle saisonnier dans le spectre. Cependant, la grande majorité des désaisonnalisateurs estiment qu'il est également souhaitable d'éliminer certains aspects de la saisonnalité *stationnaire* – d'où l'explosion des travaux de développement de filtres de désaisonnalisation fondés sur un modèle (Bell et Hillmer 1984).

La publication antérieure la plus importante à ce sujet est celle de Soukup et Findley (1999), qui proposent d'utiliser un spectre autorégressif pour trouver des pics « visuellement significatifs » – essentiellement, la valeur du spectre à chaque fréquence saisonnière (ou fréquence de jour ouvrable) est comparée aux valeurs voisines les plus proches, et est considérée comme un pic si l'écart est suffisamment grand. Cette méthode est appliquée à l'heure actuelle dans le programme X-12-ARIMA du U.S. Census Bureau (2002). L'une de ses limites tient au fait qu'elle ne comporte vraiment aucune composante statistique : la signification est non statistique – c'est-à-dire qu'elle n'est pas associée à un test d'hypothèse – et les seuils pour déterminer la « signification visuelle » sont établis de manière *ad hoc*. Dans le présent article, nous proposons un test de signification statistique pour la détection des pics qui peut donc être utilisé pour offrir des preuves statistiques supplémentaires de la présence d'un pic.

Un autre article apparenté est celui de Newton et Pagano (1983), qui élaborent des estimateurs convergents pour les maximiseurs locaux du spectre. Notre approche est légèrement différente, en ce sens que nous connaissons déjà les fréquences d'intérêt (les six fréquences saisonnières), mais que nous cherchons à tester la présence d'un pic statistiquement significatif. Si nous considérons la vraie densité spectrale  $f$  comme une fonction lisse (qui peut être quantifiée au moyen d'une décroissance suffisamment rapide de la fonction d'autocovariance), un pic est une fréquence  $\lambda_0$  telle que

$$\dot{f}(\lambda_0) = 0 \quad \ddot{f}(\lambda_0) < 0, \quad (1)$$

où  $\dot{f}$  et  $\ddot{f}$  désignent les dérivées première et seconde. Manifestement, la dérivée seconde doit être négative *avec une certaine signification* pour que le concept ait un sens. Tout bien réfléchi, il semble naïf d'examiner la géométrie infinitésimale de  $f$  au point unique  $\lambda_0$ , puisque n'importe quel petit pic dans une fonction monotone peut satisfaire (1), tout en étant dissocié des notions plus intuitives de ce qui constitue un pic. Par conséquent, il faut que la convexité soit

négative dans un voisinage raisonnablement grand de  $\lambda_0$ . Cette réflexion aboutit au diagnostic proposé dans le présent article, c'est-à-dire des mesures agrégées de la pente et de la convexité de la densité spectrale, normalisées comme il convient. Mathématiquement, elles prennent la forme d'estimations d'un périodogramme lissé par la méthode du noyau, mais sans que la largeur de bande dépende de la taille de l'échantillon.

À la section 2, nous développons les notions mathématiques de cette méthode et les illustrons au moyen de deux noyaux soigneusement choisis. À la section 3, nous montrons comment formuler les estimateurs statistiques et comment tester les hypothèses statistiques concernant les pics. À la section 4, nous mettons la méthodologie à l'essai ; des simulations fournissent une description, en échantillon fini, de la taille et de la puissance de notre test. Nous démontrons en outre l'utilité de nos méthodes au moyen d'une étude de cas à grande échelle portant sur 130 séries chronologiques provenant du U.S. Census Bureau et de l'Organisation de coopération et de développement économiques (OCDE). Nous appliquons certains concepts tirés de la littérature sur les tests multiples (Hochberg 1988) pour combiner en un seul diagnostic des tests basés sur les fréquences individuelles. À la section 5, nous présentons nos conclusions, et à l'annexe, tous les théorèmes et preuves.

## 2. Mesure de la géométrie locale du spectre

En premier lieu, nous discutons de la géométrie de la densité spectrale (ou spectre) de la série chronologique étudiée. Comme point de départ, nous considérons des mesures de la pente et de la convexité du spectre qui sont entièrement déterministes (voir l'approche de Newton et Pagano 1983) ; plus tard, à la section 3, nous envisagerons des mesures statistiques. À la section 2.1, nous présentons les concepts des mesures de la pente et de la convexité. À la section 2.2, nous discutons de la pertinence de ces mesures pour la détection des pics, tandis qu'à la section 2.3, nous présentons, à titre d'exemples explicites, deux noyaux simples.

Supposons qu'après des transformations appropriées et une différentiation au besoin,  $X_1, X_2, \dots, X_n$  est un échantillon issu d'un processus stochastique stationnaire de moyenne nulle. Nous utilisons la notation  $X = (X_1, X_2, \dots, X_n)'$ . La densité spectrale  $f(\lambda)$ , qui est bien définie à condition que la fonction d'autocovariance  $\gamma_f(h)$  soit absolument sommable, est donnée par

$$f(\lambda) = \sum_{h=-\infty}^{\infty} \gamma_f(h) e^{-ih\lambda} \quad (2)$$

avec  $i = \sqrt{-1}$  et  $\lambda \in [-\pi, \pi]$ . Il s'ensuit que la transformée de Fourier inverse donne

$$\gamma_f(h) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda) e^{ih\lambda} d\lambda, \quad (3)$$

une relation que nous utiliserons à plusieurs reprises dans la suite. Naturellement, cette relation entre  $\gamma_g$  et  $g$  est vérifiée pour toute fonction intégrable  $g$  et non pas uniquement pour une densité spectrale. En outre, si nous désignons la matrice de Toeplitz associée à  $\gamma_g$  par  $\Sigma(g)$ , il s'ensuit que

$$\Sigma_{jk}(g) = \gamma_g(j - k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\lambda) e^{i(j-k)\lambda} d\lambda.$$

Or, d'après (2),  $f$  est une fonction  $d$  fois continûment dérivable si  $\sum_{h=-\infty}^{\infty} |h|^d |\gamma_f(h)| < \infty$ . Dans la suite de l'article, nous supposons que  $f$  est deux fois continûment dérivable (nous représenterons en abrégé cet espace de fonctions par  $C^2$ ).

## 2.1 Mesures de la pente et de la convexité

La géométrie locale d'une fonction dans  $C^2$  peut être décrite au moyen de ses dérivées première et seconde ; une mesure agrégée de ces dérivés s'obtient par intégration sur une bande de fréquences. Alternativement, on peut intégrer contre une fonction  $A$  qui possède un support compact sur cette bande, à condition que  $A$  fournisse une approximation appropriée pour l'intégration sur la bande. Nous désignons cette intégrale au moyen du dispositif général d'une fonctionnelle  $\theta_A$ , où

$$\theta_A(f) = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\lambda) f(\lambda) d\lambda. \quad (4)$$

La fonction  $A$  sera appelée le « noyau » de cette fonctionnelle. Les mesures agrégées de la pente et de la convexité sont donc définies par  $\theta_A(\dot{f})$  et  $\theta_A(\ddot{f})$ , où chaque point désigne une seule dérivée. Ces fonctionnelles donnent une mesure sommaire de la pente et de la convexité de  $f$  sur la bande  $[\mu - \beta/2, \mu + \beta/2] \subset [0, \pi]$ , et les noyaux correspondants seront par conséquent désignés par  $A_{\beta, \mu}$ . Nous considérons les noyaux ayant les propriétés suivantes : i)  $A_{\beta, \mu}$  est une fonction dans  $C^2$  sur  $[-\pi, \pi]$  ; ii)  $A_{\beta, \mu}$  est nul en dehors de la bande  $[\mu - \beta/2, \mu + \beta/2]$  ; iii)  $A_{\beta, \mu}$  est symétrique autour de  $\mu$  sur cette bande ; iv)  $\dot{A}_{\beta, \mu}(\mu \pm \beta/2) = 0$ . La condition (iii) garantit que l'emplacement du pic dans  $f$  ne sera pas modifié par l'emploi du noyau  $A_{\beta, \mu}$ . Il convient de souligner que nous n'imposons pas que l'intégrale totale de  $A_{\beta, \mu}$  soit égale à l'unité, parce que plus tard, nous emploierons une normalisation qui tiendra compte automatiquement de la masse totale du noyau. Maintenant, au moyen de (iv) et d'une intégration par partie dans (4), nous obtenons

$$\begin{aligned} \theta_{A_{\beta, \mu}}(\dot{f}) &= -\theta_{\dot{A}_{\beta, \mu}}(f) \\ \theta_{A_{\beta, \mu}}(\ddot{f}) &= \theta_{\ddot{A}_{\beta, \mu}}(f). \end{aligned} \quad (5)$$

Ces formules sont commodes, parce qu'elles ne requièrent que la connaissance de  $f$ , et non celle de ses dérivées (en supposant que nous puissions calculer  $\dot{A}_{\beta, \mu}$  et  $\ddot{A}_{\beta, \mu}$ ). En nous inspirant de la littérature abondante sur les noyaux dans la régression non paramétrique et l'estimation de la densité spectrale, nous pouvons commencer par utiliser un noyau pair  $A$  défini sur la bande  $[-\pi, \pi]$  qui satisfait (i) et  $\dot{A}(\pm\pi) = 0$ . Alors,  $A_{\beta, \mu}$  est défini par

$$A_{\beta, \mu}(\lambda) = \frac{2\pi}{\beta} A\left(\frac{2\pi}{\beta}(\lambda - \mu)\right),$$

et est nul en dehors de la bande de fréquence  $[\mu - \beta/2, \mu + \beta/2]$ . Manifestement, nous devons imposer que  $\beta \leq 2\mu$  et  $\beta \leq 2(\pi - \mu)$ , pour que  $[\mu - \beta/2, \mu + \beta/2] \subset [0, \pi]$  ; et le noyau  $A_{\beta, \mu}$  satisfait les conditions (i) à (iv). Notons que nous ne pouvons pas construire ces types de mesures si  $\mu$  est égal à 0 ou  $\pi$ . En modifiant les variables, nous voyons que

$$\gamma_{A_{\beta, \mu}}(h) = \exp\{ih\mu\} \gamma_A(h\beta/2\pi), \quad (6)$$

de sorte que les effets de  $\beta$  et de  $\mu$  sont, en un certain sens, séparables. Notons que nous évaluons habituellement  $\gamma_A$  à des valeurs non entières, de sorte que ces relations sont obtenues en étendant (3) à des arguments non entiers. Il peut être troublant de constater que  $\gamma_{A_{\beta, \mu}}$  est une fonction à valeurs complexes, mais en réalité, seule sa partie réelle entrera dans nos estimateurs statistiques. Évidemment, en dernière analyse, nous nous intéressons à  $\dot{A}_{\beta, \mu}$  et  $\ddot{A}_{\beta, \mu}$ , qui sont donnés par

$$\dot{A}_{\beta, \mu}(\lambda) = \frac{4\pi^2}{\beta^2} \dot{A}\left(\frac{2\pi}{\beta}(\lambda - \mu)\right),$$

et

$$\ddot{A}_{\beta, \mu}(\lambda) = \frac{8\pi^3}{\beta^3} \ddot{A}\left(\frac{2\pi}{\beta}(\lambda - \mu)\right).$$

Plus tard, nous considérerons les carrés de ces noyaux, et leurs transformées de Fourier inverses correspondantes. Donc, en supposant que  $[\mu - \beta/2, \mu + \beta/2] \subset [0, \pi]$ , les carrés sont donnés par

$$\dot{A}_{\beta, \mu}^2(\lambda) = \frac{16\pi^4}{\beta^4} \dot{A}^2\left(\frac{2\pi}{\beta}(\lambda - \mu)\right)$$

et

$$\ddot{A}_{\beta, \mu}^2(\lambda) = \frac{64\pi^6}{\beta^6} \ddot{A}^2\left(\frac{2\pi}{\beta}(\lambda - \mu)\right).$$

Enfin, nous notons d'après (4) que nous pouvons réécrire  $\theta_A(f)$  sous la forme

$$\theta_A(f) = \sum_{h=-\infty}^{\infty} \gamma_A(h) \gamma_f(h). \quad (7)$$

Donc, il pourrait être avantageux de déterminer la séquence  $\gamma_A(h)$  à partir du noyau  $A$ . En prenant la transformée de Fourier inverse des noyaux susmentionnés de la pente et de la convexité, nous pouvons construire  $\Sigma(\dot{A}_{\beta, \mu})$ ,  $\Sigma(\ddot{A}_{\beta, \mu})$ ,  $\Sigma(\dot{A}_{\beta, \mu}^2)$  et  $\Sigma(\ddot{A}_{\beta, \mu}^2)$  comme il suit :

$$\begin{aligned} \gamma_{\dot{A}_{\beta, \mu}}(h) &= \frac{2\pi}{\beta} \exp\{ih\mu\} \gamma_A(h\beta/2\pi), \\ \gamma_{\ddot{A}_{\beta, \mu}}(h) &= \frac{4\pi^2}{\beta^2} \exp\{ih\mu\} \gamma_{\dot{A}}(h\beta/2\pi), \\ \gamma_{\dot{A}_{\beta, \mu}^2}(h) &= \frac{8\pi^3}{\beta^3} \exp\{ih\mu\} \gamma_{\dot{A}^2}(h\beta/2\pi), \\ \gamma_{\ddot{A}_{\beta, \mu}^2}(h) &= \frac{32\pi^5}{\beta^5} \exp\{ih\mu\} \gamma_{\ddot{A}^2}(h\beta/2\pi). \end{aligned} \quad (8)$$

Donc, si nous possédons l'information relative au domaine temporel  $\gamma_f(h)$  pour le processus  $\{X_t\}$ , nous pouvons calculer les mesures de la pente et de la convexité en utilisant (7), sachant la séquence des transformées de Fourier inverses des noyaux appropriés. Puisque  $\gamma_f(h)$  est une séquence symétrique, nous ne devons considérer que la partie réelle de  $\gamma_A(h)$  s'il se fait que sa valeur est complexe.

## 2.2 Creux et pics

Les mesures agrégées de la pente et de la convexité spectrales décrites antérieurement représentent les composantes élémentaires des déterminants de la géométrie spectrale locale. Notre objectif général est de déterminer si un intervalle donné du spectre correspond à un pic ou à un creux (ou est monotone). Dans la géométrie de deuxième ordre de l'analyse infinitésimale, un maximum local possède la propriété déterminante que sa dérivée première est nulle et sa dérivée seconde, strictement négative. Il faut donc manifestement examiner, séquentiellement, une mesure de pente et une mesure de convexité définies sur la même bande de fréquences.

Afin de vérifier la présence d'un pic, l'approche séquentielle peut être considérée comme équivalant à formuler des énoncés inférentiels au sujet de  $\theta_{A_{\beta, \mu}}(\dot{f})$  et  $\theta_{A_{\beta, \mu}}(\ddot{f})$ . Soulignons qu'en formulant ces énoncés inférentiels, nous choisissons la valeur de  $\mu$  d'avance, en fonction de l'endroit dans le spectre où nous souhaitons détecter un pic (ou un creux) ;  $\beta$  est choisi en fonction des fréquences que

nous désirons exclure, décision qui dépend de la mesure dans laquelle nous souhaitons que notre point de vue du spectre soit local. Puis, nous disons que  $\mu$  est un pic  $\beta$ -agrégé (par rapport à  $A$ ) du spectre si

$$\theta_A(\dot{f}) = 0 \quad \text{et} \quad \theta_A(\ddot{f}) < 0.$$

L'aspect séquentiel découle de l'idée que, en général, nous commençons par déterminer si  $\theta_A(\dot{f}) = 0$ , puis que nous déterminons la convexité ; cela deviendra plus évident quand nous examinerons les tests statistiques à la section 3.2. De manière comparable, nous définissons un creux  $\beta$ -agrégé quand  $\theta_A(\dot{f}) > 0$ . En ce qui concerne le test d'hypothèse pour un pic, nous avons

$$\begin{aligned} H_0^{(1)} : \theta_A(\dot{f}) = 0 \quad \text{vs} \quad H_a^{(1)} : \theta_A(\dot{f}) \neq 0 \\ H_0^{(2)} : \theta_A(\ddot{f}) = 0 \quad \text{vs} \quad H_a^{(2)} : \theta_A(\ddot{f}) < 0. \end{aligned}$$

L'aspect inhabituel de ce test d'hypothèse est que nous voulons d'abord ne pas rejeter  $H_0^{(1)}$ , puis, conditionnellement à ce test, nous voulons rejeter  $H_0^{(2)}$  en faveur de l'alternative  $H_a^{(2)}$ .

## 2.3 Exemples de noyaux

Il existe une foule de noyaux qui satisfont les conditions (i) à (iv) ; il nous suffit de les emprunter à la littérature sur l'estimation non paramétrique de la densité. Par exemple, les noyaux de Parzen et de Tukey-Hanning (TH) (discutés dans Priestley 1981) conviennent, tandis que ceux de Bartlett et Daniell ne sont pas appropriés, puisque (iv) n'est pas vérifiée. En général, il suffit d'utiliser (8) pour déterminer les transformées de Fourier inverses. À la présente section, nous examinons deux exemples, les noyaux quartique et TH. Leur avantage est qu'il est facile de calculer leurs dérivées première et seconde, et que l'on peut obtenir une forme explicite de leurs transformées de Fourier inverses.

### Exemple 1 : Noyau quartique

Nous commençons par considérer un noyau polynomial de degré quatre, c'est-à-dire quartique. L'imposition de toutes les contraintes (i) à (iv) donne la forme suivante :

$$A(\lambda) = \frac{15}{8\pi^4} (\lambda^4 - 2\pi^2\lambda^2 + \pi^4),$$

$$\dot{A}(\lambda) = \frac{15}{8\pi^4} (4\pi^3 - 4\pi^2\lambda), \quad \text{et}$$

$$\ddot{A}(\lambda) = \frac{15}{8\pi^4} (12\pi^2 - 4\pi^2).$$

En prenant la transformée de Fourier inverse des noyaux de la pente et de la convexité (et leurs carrés), nous obtenons

$$\begin{aligned}\gamma_{\dot{A}}(h) &= \frac{15i}{\pi^5} \left( \frac{\pi^2 \sin \pi h}{h^2} + \frac{3\pi \cos \pi h}{h^3} - \frac{3 \sin \pi h}{h^4} \right), \\ \gamma_{\ddot{A}}(h) &= \frac{15}{\pi^5} \left( \frac{\pi^2 \sin \pi h}{h} + \frac{3\pi \cos \pi h}{h^2} - \frac{3 \sin \pi h}{h^3} \right), \\ \gamma_{\dot{A}^2}(h) &= \frac{225}{\pi^9} \left( -\frac{2\pi^4 \sin \pi h}{h^3} - \frac{18\pi^3 \cos \pi h}{h^4} \right. \\ &\quad \left. + \frac{78\pi^2 \sin \pi h}{h^5} + \frac{180\pi \cos \pi h}{h^6} - \frac{18 \sin \pi h}{h^7} \right), \\ \gamma_{\ddot{A}^2}(h) &= \frac{225}{4\pi^{11}} \left( \frac{\pi^4 \sin \pi h}{h} + \frac{6\pi^3 \cos \pi h}{h^2} \right. \\ &\quad \left. - \frac{24\pi^2 \sin \pi h}{h^3} - \frac{54\pi \cos \pi h}{h^4} + \frac{54 \sin \pi h}{h^5} \right),\end{aligned}$$

auxquels nous appliquons (8) et obtenons

$$\begin{aligned}\gamma_{\dot{A}_{\beta, \mu}}(h) &= \frac{30i}{\beta} \exp\{ih\mu\} \left( \frac{\sin k}{k^2} + \frac{3 \cos k}{k^3} - \frac{3 \sin k}{k^4} \right), \\ \gamma_{\ddot{A}_{\beta, \mu}}(h) &= \frac{30}{\beta^2 \pi} \exp\{ih\mu\} \left( \frac{\sin k}{k} + \frac{3 \cos k}{k^2} - \frac{3 \sin k}{k^3} \right), \\ \gamma_{\dot{A}_{\beta, \mu}^2}(h) &= \frac{1\,800\pi}{\beta^3} \exp\{ih\mu\} \left( \frac{2 \sin k}{k^3} + \frac{18 \cos k}{k^4} \right. \\ &\quad \left. - \frac{78 \sin k}{k^5} - \frac{180 \cos k}{k^6} + \frac{180 \sin k}{k^7} \right), \text{ et} \\ \gamma_{\ddot{A}_{\beta, \mu}^2}(h) &= \frac{1\,800}{\beta^5 \pi} \exp\{ih\mu\} \left( \frac{\sin k}{k} + \frac{6 \cos k}{k^2} \right. \\ &\quad \left. - \frac{24 \sin k}{k^3} - \frac{54 \cos k}{k^4} + \frac{54 \sin k}{k^5} \right),\end{aligned}$$

où  $k = h\beta/2$ . Notons que  $\gamma_{\dot{A}_{\beta, \mu}^2}(0) = 240\pi/(7\beta^3)$  et  $\gamma_{\ddot{A}_{\beta, \mu}^2}(0) = 360/(\beta^5 \pi)$  découlent de l'application de la règle de L'Hôpital. Ces formules nous permettent de construire les matrices de Toeplitz appropriées pour le diagnostic (comme il est exposé plus bas, à la section 3.1, il suffit de considérer la partie réelle de ces séquences).

#### Exemple 2 : Noyau TH

Nous pouvons obtenir une forme semblable au noyau quartique en utilisant une fonction cosinus. Le choix qui suit satisfait toutes les conditions énoncées sur un noyau :

$$\begin{aligned}A(\lambda) &= \frac{1}{2\pi} (1 + \cos \lambda), \\ \dot{A}(\lambda) &= \frac{1}{2\pi} (-\sin \lambda), \text{ et} \\ \ddot{A}(\lambda) &= \frac{1}{2\pi} (-\cos \lambda).\end{aligned}$$

Cette fonction est identique à la fenêtre des décalages de Tukey-Hanning, quoique ici, nous l'appliquons comme une

fenêtre spectrale (voir Priestley 1981). Dans la suite de l'exposé, nous la désignerons par noyau TH. En prenant la transformée de Fourier inverse des noyaux de la pente et de la convexité (et leurs carrés), nous obtenons

$$\begin{aligned}\gamma_{\dot{A}}(h) &= \frac{i}{4\pi^2} \left( \frac{\sin \pi(h+1)}{h+1} - \frac{\sin \pi(h-1)}{h-1} \right), \\ \gamma_{\ddot{A}}(h) &= -\frac{1}{4\pi^2} \left( \frac{\sin \pi(h+1)}{h+1} + \frac{\sin \pi(h-1)}{h-1} \right), \\ \gamma_{\dot{A}^2}(h) &= \frac{1}{16\pi^3} \left( \frac{2 \sin \pi h}{h} - \frac{\sin \pi(h+2)}{h+2} - \frac{\sin \pi(h-2)}{h-2} \right), \text{ et} \\ \gamma_{\ddot{A}^2}(h) &= \frac{1}{16\pi^3} \left( \frac{2 \sin \pi h}{h} + \frac{\sin \pi(h+2)}{h+2} + \frac{\sin \pi(h-2)}{h-2} \right).\end{aligned}$$

L'application de (8) donne alors

$$\begin{aligned}\gamma_{\dot{A}_{\beta, \mu}}(h) &= \frac{i}{2\beta} \exp\{ih\mu\} \left( \frac{\sin(k+\pi)}{k+\pi} - \frac{\sin(k-\pi)}{k-\pi} \right), \\ \gamma_{\ddot{A}_{\beta, \mu}}(h) &= -\frac{\pi}{\beta^2} \exp\{ih\mu\} \left( \frac{\sin(k+\pi)}{k+\pi} + \frac{\sin(k-\pi)}{k-\pi} \right), \\ \gamma_{\dot{A}_{\beta, \mu}^2}(h) &= \frac{\pi}{2\beta^3} \exp\{ih\mu\} \left( \frac{2 \sin k}{k} - \frac{\sin(k+2\pi)}{k+2\pi} \right. \\ &\quad \left. - \frac{\sin(k-2\pi)}{k-2\pi} \right), \text{ et} \\ \gamma_{\ddot{A}_{\beta, \mu}^2}(h) &= \frac{2\pi^3}{\beta^5} \exp\{ih\mu\} \left( \frac{2 \sin k}{k} - \frac{\sin(k+2\pi)}{k+2\pi} \right. \\ &\quad \left. + \frac{\sin(k-2\pi)}{k-2\pi} \right),\end{aligned}$$

où  $k = h\beta/2$ . Notons que  $\gamma_{\dot{A}_{\beta, \mu}^2}(0) = \pi/\beta^3$  et  $\gamma_{\ddot{A}_{\beta, \mu}^2}(0) = 4\pi^3/\beta^5$  découlent de l'application de la règle de L'Hôpital (en utilisant la convention que  $\sin(0)/0 = 1$ ). Ces formules nous permettent de construire les matrices de Toeplitz appropriées pour le diagnostic (de nouveau, comme nous l'exposons plus bas à la section 3.1, il suffit de considérer la partie réelle de ces séquences).

### 3. Méthodologie statistique

Évidemment, d'habitude, nous ne connaissons pas le spectre  $f$  si bien qu'il faut généralement produire des estimations d'après les données. À la présente section, nous décrivons des estimations statistiques des mesures de la pente et de la convexité qui sont convergentes et simples à calculer dans le domaine temporel. Sous certaines hypothèses supplémentaires faibles, ces estimations sont asymptotiquement normales, ce qui sera avantageux quand nous effectuerons les tests d'hypothèse. À la section 3.1, nous définissons les estimations statistiques et discutons de leurs

propriétés asymptotiques. À la section 3.2, nous discutons de l'application à la détection individuelle des pics et à la section 3.3, nous donnons une extension à leur détection conjointe, qui facilite une application importante dans la désaisonnalisation. À la section 3.4, nous discutons des extensions à des données non stationnaires avec tendance.

### 3.1 Estimateurs de la pente et de la convexité

Nous commençons par noter que la forme quadratique (pour toute fonction intégrable  $g$ ) est

$$\frac{1}{n} X' \Sigma(g) X = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\lambda) I(\lambda) d\lambda,$$

où  $I$  désigne le périodogramme. Bien que ce dernier soit habituellement défini par les fréquences de Fourier  $(2\pi j/n ; j = 1, \dots, \lfloor n/2 \rfloor)$ , nous le définissons comme une bande continue de fréquences de la façon suivante

$$I(\lambda) = \frac{1}{n} \left| \sum_{t=1}^n X_t e^{-it\lambda} \right|^2 = \sum_{h=1-n}^{n-1} R(h) e^{-ih\lambda}, \lambda \in [-\pi, \pi] \quad (9)$$

avec  $R(h)$  égal à la fonction d'autocovariance (non centrée) d'échantillon. Cela donne un moyen élégant de passer du domaine temporel au domaine fréquentiel qui est bien décrit dans la littérature sur les séries chronologiques (voir Taniguchi et Kakizawa 2000). En outre, ce genre d'intégrales du périodogramme sont généralement convergentes, c'est-à-dire que  $\theta_g(I) \xrightarrow{p.s.} \theta_g(f)$  quand  $n \rightarrow \infty$ , sous les conditions faibles discutées plus bas (notons que la non-convergence du périodogramme est résolue par l'agrégation spectrale contre la fonction  $g$ , comme il est montré en annexe). Par conséquent, nous obtenons des estimations statistiques des mesures de la pente et de la convexité  $f$  en utilisant une approche de substitution, c'est-à-dire en remplaçant simplement  $f$  par  $I$  dans  $\theta_{A_{\beta, \mu}}$ . En particulier,

$$\begin{aligned} \hat{\theta}_{A_{\beta, \mu}}(\dot{f}) &= -\theta_{A_{\beta, \mu}}(I) = -\frac{1}{n} X' \Sigma(\dot{A}_{\beta, \mu}) X, \text{ et} \\ \hat{\theta}_{A_{\beta, \mu}}(\ddot{f}) &= \theta_{A_{\beta, \mu}}(I) = \frac{1}{n} X' \Sigma(\ddot{A}_{\beta, \mu}) X. \end{aligned} \quad (10)$$

Cette définition comprend l'utilisation de (5), qui explique le signe négatif de la mesure de la pente. Afin de calculer l'estimation, nous utilisons la représentation dans le domaine temporel (exprimée sous une forme quadratique). Cette représentation est commode parce que nous devons uniquement déterminer une longueur appropriée pour les séquences  $\gamma_{\dot{A}_{\beta, \mu}}(h)$  et  $\gamma_{\ddot{A}_{\beta, \mu}}(h)$ , former les matrices de Toeplitz  $\Sigma(\dot{A}_{\beta, \mu})$  et  $\Sigma(\ddot{A}_{\beta, \mu})$ , puis calculer les formes quadratiques. Notons que les transformées de Fourier inverses de  $\dot{A}_{\beta, \mu}$  et  $\ddot{A}_{\beta, \mu}$  ne doivent être déterminées qu'une seule fois (voir la section 2.3 pour des exemples explicites) et peuvent être produites d'avance,

puis appliquées de manière répétée à de nombreuses séries chronologiques différentes.

Afin de calculer la représentation dans le domaine temporel des mesures de la pente et de la convexité données par (10), nous utilisons (8) ; par exemple voir les formules dans les exemples 1 et 2. Évidemment, il résultera en général que  $\Sigma(\dot{A}_{\beta, \mu})$  et  $\Sigma(\ddot{A}_{\beta, \mu})$  sont complexes. Cependant, même si  $\Sigma(g)$  (où  $g$  peut être  $A_{\beta, \mu}, \dot{A}_{\beta, \mu}$  ou  $\ddot{A}_{\beta, \mu}$ ) est une matrice de Toeplitz complexe,  $X' \Sigma(g) X$  sera toujours réel. Partant de (8), il est facile de voir que  $\Sigma(g) = M + iN$  où  $M$  est réelle, symétrique et Toeplitz, et que  $N$  est réelle, antisymétrique et Toeplitz. D'où  $X'NX = 0$  pour tout vecteur  $X$ , de sorte que  $X' \Sigma(g) X = X' M X$ . Par conséquent, pour calculer les mesures statistiques de la pente et de la convexité, nous pouvons prendre la partie réelle de  $\gamma_A(h)$  dans (8).

Ces estimations statistiques sont non seulement convergentes, mais aussi asymptotiquement normales sous certaines conditions supplémentaires (discutées à l'annexe). Cependant, pour construire une normalisation appropriée, il sera nécessaire d'estimer leur variation. La variance asymptotique de  $\theta_g(I)$  est  $\theta_{g^2}(f^2)$  (si  $g$  est appuyée sur  $[0, \pi]$ ), qui peut être estimée de manière convergente par la voie de  $\theta_{g^2}(I^2)/2$ . (Le facteur de 2 est requis, puisque l'intégrale de  $I^2$  tend vers l'intégrale correspondante de  $2f^2$  - voir Chiu (1988)). Nous pouvons représenter cela dans le domaine temporel de la façon suivante. Soit  $R = \{R(1-n), \dots, R(0), \dots, R(n-1)\}'$  un vecteur de dimension  $2n-1$  des autocovariances d'échantillon et soit  $\Sigma(g^2)$  de dimension  $2n-1$  dans la formule suivante :  $R' \Sigma(g^2) R / 2 = \theta_{g^2}(I^2) / 2$ . Cette relation peut être vérifiée facilement en utilisant (9). Donc, nous normalisons  $\theta_g(I)$  par la racine carrée de  $\theta_{g^2}(I^2) / 2$ . D'où une mesure statistique normalisée de la pente et de la convexité sont données par

$$-\psi_{\dot{A}_{\beta, \mu}}(I) = -\frac{\theta_{\dot{A}_{\beta, \mu}}(I)}{\sqrt{\theta_{\dot{A}_{\beta, \mu}}^2(I^2)/2}} = -\frac{1}{n} \frac{X' \Sigma(\dot{A}_{\beta, \mu}) X}{\sqrt{R' \Sigma(\dot{A}_{\beta, \mu}^2) R / 2}}$$

et

$$\psi_{\ddot{A}_{\beta, \mu}}(I) = \frac{\theta_{\ddot{A}_{\beta, \mu}}(I)}{\sqrt{\theta_{\ddot{A}_{\beta, \mu}}^2(I^2)/2}} = \frac{1}{n} \frac{X' \Sigma(\ddot{A}_{\beta, \mu}) X}{\sqrt{R' \Sigma(\ddot{A}_{\beta, \mu}^2) R / 2}},$$

où les dimensions des matrices  $\Sigma$  sont soit  $n$  soit  $2n-1$  comme il est approprié. Les propriétés asymptotiques de  $\psi_{\dot{A}_{\beta, \mu}}(I)$  et de  $\psi_{\ddot{A}_{\beta, \mu}}(I)$  sont discutées à l'annexe. En résumé,  $-\sqrt{n} \psi_{\dot{A}_{\beta, \mu}}(I)$  et  $\sqrt{n} \psi_{\ddot{A}_{\beta, \mu}}(I)$  sont toutes deux marginalement asymptotiquement de loi  $N(0, 1)$  sous  $H_0^{(1)}$  et  $H_0^{(2)}$ , respectivement, et sous les hypothèses discutées à l'annexe. Les simulations indiquent que la normalisation de la variance ne converge que lentement et que sa corrélation avec le numérateur cause un certain

degré de non-normalité dans les petits échantillons. D'après l'histogramme de la distribution simulée sous une hypothèse nulle de bruit blanc gaussien avec  $n = 360$  et 10 000 répétitions (figure 1), la concordance avec la loi normale est proche, sauf aux extrémités dans les queues. À la section 4, nous examinons ce comportement plus en détail au moyen d'études par simulation.

### 3.2 Applications au test de la présence d'un seul pic

Nous considérons maintenant l'application au test de la présence d'un pic. Rappelons que nous avons une hypothèse nulle initiale  $H_0^{(1)}$  qui ne doit pas être rejetée pour pouvoir poursuivre. Cela peut s'interpréter comme signifiant qu'il n'existe pas de preuve suffisamment convaincante pour conclure que la dérivée première (pente) de la densité spectrale diffère de manière significative de zéro. Or, nous savons que  $-\sqrt{n}\psi_{\hat{A}_{\beta, \mu}}(I)$  est asymptotiquement de loi  $N(0, 1)$  sous  $H_0^{(1)}$  ainsi que les hypothèses discutées à l'annexe. Si nous supposons en outre qu'une valeur suffisamment petite  $x$  est obtenue pour la statistique de test, nous ne pourrions rejeter  $H_0^{(1)}$  avec aucun degré de confiance. Dans ce cas, nous pouvons considérer l'hypothèse  $H_0^{(2)}$ , que nous cherchons à rejeter, ce que nous testons au moyen de  $\sqrt{n}\psi_{\hat{A}_{\beta, \mu}}(I)$ . Bien que  $-\sqrt{n}\psi_{\hat{A}_{\beta, \mu}}(I)$  et  $\sqrt{n}\psi_{\hat{A}_{\beta, \mu}}(I)$  soient asymptotiquement corrélées (voir le théorème 1 de l'annexe), nous considérerons les tests de la pente et de la convexité comme s'ils étaient exécutés séparément. (Cette corrélation peut être estimée et utilisée pour déterminer la distribution du diagnostic de convexité conditionnellement au diagnostic de pente; cependant, l'interprétation des valeurs  $p$  devient confuse. Pour simplifier, nous traitons les tests séparément, un à la fois, et nous ne tenons pas compte explicitement de la corrélation.) Notre procédure de test s'exécute alors de la façon suivante :

1. Faire le test bilatéral de  $H_0^{(1)}$  en utilisant  $-\sqrt{n}\psi_{\hat{A}_{\beta, \mu}}(I)$ .
2. Poser que  $p$  est la valeur  $p$  associée avec la première valeur de la statistique de test  $x = -\sqrt{n}\psi_{\hat{A}_{\beta, \mu}}(I)$ , avec  $x$  et  $p$  reliées par  $p = 2\Phi(-|x|)$ .
3. Si  $p > 0,05$  (ou un autre seuil de tolérance pré-établi), poursuivre; sinon, conclure qu'aucun pic n'est présent.
4. Exécuter le test unilatéral inférieur de  $H_0^{(2)}$  en utilisant  $\sqrt{n}\psi_{\hat{A}_{\beta, \mu}}(I)$ .
5. Rejeter  $H_0^{(2)}$  et conclure qu'il existe un pic si  $\sqrt{n}\psi_{\hat{A}_{\beta, \mu}}(I) < \Phi^{-1}(\alpha)$ , où  $\alpha$  est le niveau du test de convexité.

### 3.3 Test conjoint de la présence de plusieurs pics : application à la désaisonnalisation

Considérons maintenant la situation où nous voulons tester simultanément la présence de plusieurs pics spectraux. Il est clair que nous pourrions concevoir un noyau comportant plusieurs nœuds, un à chaque pic, mais il serait simplement la somme de plusieurs diagnostics de pic spectral individuels. L'inconvénient serait qu'un pic spectral significatif à un endroit pourrait annuler un creux spectral significatif ailleurs. Par conséquent, nous préférons un test permettant d'examiner un ensemble de diagnostics spectraux dans un paradigme de tests multiples.

Par exemple, considérons le contexte de la détection de la présence de pics spectraux dans des données désaisonnalisées. Six pics saisonniers présentent un intérêt, mais nous devons nous limiter à cinq d'entre eux à cause de problèmes d'alias, c'est-à-dire de repliement du spectre (le pic de fréquence  $\pi$  ne peut pas être identifié). Si un ou plusieurs pics spectraux sont significatifs, nous devons rejeter notre procédure de désaisonnalisation (puisque'elle n'a pas réussi à éliminer tous les pics); par conséquent, nous sommes dans des conditions de tests multiples et nous utiliserons une méthode permettant de contrôler le taux d'erreur de type I global (FWER pour *familywise error rate*) proposé par Hochberg (1988) et décrit dans Benjamini et Hochberg (page 294, 1995). En nous limitant à la question de la convexité, nous avons les hypothèses nulles  $H_0^{(2)}$  pour chacune des cinq fréquences saisonnières. Dans nos conditions, la procédure de Hochberg (1988) consiste à calculer les valeurs  $p$  pour le test de convexité à chacune de ces cinq fréquences saisonnières et à les classer selon  $p_{(1)} \leq p_{(2)} \leq p_{(3)} \leq p_{(4)} \leq p_{(5)}$ , avec les hypothèses nulles correspondantes désignées par  $H_{(i)}$ . Pour un taux d'erreur de type I global spécifié de niveau  $\alpha$  (par exemple  $\alpha = 0,05$ ), soit  $k$  la valeur de  $i$  la plus grande pour laquelle  $p_{(i)} \leq i/(6 - i)\alpha$ ; alors, nous rejetons toutes les  $H_{(i)}$  pour  $i \leq k$ .

En utilisant cette procédure, nous devrions commettre des erreurs de type I – c'est-à-dire déterminer la présence d'au moins une fréquence saisonnière ayant une convexité négative alors qu'il n'en existe aucune – une proportion de fois environ égale à  $\alpha$  (si nous nous limitons à examiner  $H_0^{(2)}$ , l'hypothèse de convexité). L'avantage de l'approche du taux d'erreur de type I global FWER de Hochberg (H-FWER) est qu'elle améliore spectaculairement la puissance statistique comparativement aux autres méthodes. La validité de cette méthode requiert l'indépendance des statistiques de test prises en considération et, pour cette raison, nous prenons cinq noyaux  $A_1, \dots, A_5$  – centrés sur les fréquences saisonnières  $\pi/6, \dots, 5\pi/6$  respectivement – qui ont un support disjoint. Alors, le théorème 1 peut être généralisé pour obtenir l'indépendance asymptotique de cinq



statistiques de test de convexité (voir la discussion qui suit le théorème 1 à l'annexe). Naturellement, nous exécutons aussi cinq tests distincts de la pente à chaque fréquence saisonnière, où il faut que nous n'aboutissions à un rejet dans aucun cas pour pouvoir poursuivre

Enfin, nous notons qu'en pratique, une désaisonnalisation est rarement rejetée sur la base d'une masse spectrale significative à la cinquième fréquence saisonnière de  $5\pi/6$  (Findley 2006). Cela tient en partie à la difficulté d'attribuer une interprétation à cette fréquence. Par conséquent, le praticien pourrait être plus intéressé par un « test à quatre pics » qui est axé sur les quatre premières fréquences saisonnières ; ce test s'obtient par une modification évidente de la procédure H-FWER décrite plus haut.

### 3.4 Extension à des données non stationnaires

La méthodologie que nous venons d'exposer s'appuie sur l'hypothèse que les données sont un échantillon d'un processus stationnaire. Cependant, dans le contexte de la désaisonnalisation, les données désaisonnalisées sont habituellement intégrées une fois ou deux fois. Dans ce cas, on différencierait les données désaisonnalisées une ou deux fois avant d'appliquer les diagnostics. Or, les opérateurs de différenciation  $1 - B$  et  $(1 - B)^2$  appliqués sont essentiellement des filtres passe-haut, qui en principe atténuent les pics spectraux résiduels proches de la fréquence zéro (en particulier, la première fréquence saisonnière à  $\pi/6$ ). Donc, il serait peut-être souhaitable d'appliquer le diagnostic au pseudo-spectre, ce qui peut se faire si le support du noyau est borné de façon à ne pas contenir les pôles de la densité spectrale.

Supposons que les données observées sont maintenant  $Y_{-d}, \dots, Y_n$  pour  $d$  l'ordre de différenciation de la tendance (donc habituellement  $d = 1$  ou  $2$ ). Lorsque les données observées sont différenciées, nous obtenons l'échantillon  $X$ , qui est strictement stationnaire. La pseudo-densité spectrale du processus  $\{Y_t\}$  est  $g(\lambda) = f(\lambda) |1 - e^{-i\lambda}|^{-2d}$ , où  $f$  est le spectre de  $\{X_t\}$ . Ce pseudo-spectre pourrait être estimé au moyen de  $\hat{g}(\lambda) = I(\lambda) |1 - e^{-i\lambda}|^{-2d}$ , où  $I$  est le périodogramme de  $X$  comme auparavant ; il s'agit de l'approche de recoloration de Nerlove (1964). Alors,  $\theta_A(g)$  est bien défini à condition que  $A(\lambda) |1 - e^{-i\lambda}|^{-2d}$  soit une fonction intégrable ; essentiellement, nous devons nous assurer que la fréquence zéro est exclue du support du noyau  $A$ . Puisqu'en pratique  $A$  est centré autour des fréquences saisonnières, nous pouvons facilement trouver le moyen d'appliquer cette condition. L'estimateur correspondant est alors

$$\hat{\theta}_A(\hat{g}) = \theta_{\hat{A}b}(I),$$

où  $b(\lambda) = |1 - e^{-i\lambda}|^{-2d}$ . L'estimateur est bien défini si  $\hat{A}b$  est intégrable ; en outre, les propriétés asymptotiques

discutées à l'annexe pour le cas stationnaire s'étendent à ce cas également, à condition que  $\hat{A}b$  soit borné.

Cette extension pourrait paraître plus séduisante à certains chercheurs. Cependant, son prix est que la transformée de Fourier inverse de  $\hat{A}b$  doit être déterminée, ce qui requiert un travail mathématique supplémentaire. Dans les études par simulation et les illustrations des données de la section 4, nous différencions les données désaisonnalisées, mais nous n'appliquons pas le facteur de correction  $b$  dans le noyau.

## 4. Études empiriques

Ayant développé les aspects théoriques du diagnostic spectral, nous nous tournons maintenant vers sa performance en pratique. Pour commencer, nous présentons certains résultats obtenus par simulation, qui donnent une idée des propriétés de taille et de puissance de la statistique de test dans les échantillons finis. Puis, nous étudions empiriquement la taille et la puissance en appliquant les diagnostics spectraux à une suite de 130 séries chronologiques (65 séries du U.S. Census Bureau et 65 séries de l'OCDE) ; nous considérons les séries originales ainsi que les séries désaisonnalisées et nous comparons les résultats à ceux du diagnostic de signification visuelle et aux diagnostics de contrôle de la qualité M7 et M8 de X-12-ARIMA (U.S. Census Bureau 2002). D'autres études empiriques sont décrites dans Evans, Holan et McElroy (2006).

### 4.1 Étude par simulation

Afin d'évaluer la performance de nos diagnostics, nous avons exécuté plusieurs simulations. Le premier ensemble de simulations a pour but d'examiner la taille (niveau) pour le diagnostic de la présence d'un pic unique. Pour cette simulation, nous considérons les diagnostics de pente et de convexité séparément. Même si, en pratique, quand nous considérons le diagnostic de pente, nous souhaitons ne pas rejeter l'hypothèse nulle  $H_0^{(1)}$ , ici nous voulons étudier empiriquement les propriétés distributionnelles et nous imposons donc la définition habituelle de taille pour cette étude. Donc, nous simulons un bruit blanc gaussien qui satisfait les hypothèses du théorème 1, ainsi que  $\psi_{\hat{A}}(I) = \psi_{\hat{A}}(I) = 0$ , de sorte que  $H_0^{(1)}$  et  $H_0^{(2)}$  sont vraies. Naturellement, il existe de nombreux processus pour lesquels  $H_0^{(1)}$  et  $H_0^{(2)}$  sont vraies simultanément – par exemple, tout processus dont la densité spectrale est localement uniforme ; cependant, en raison de considérations asymptotiques, il suffit de considérer le bruit blanc. Pour une (grande) taille d'échantillon de  $n = 360$ , en utilisant le noyau TH avec  $\mu = \pi/6$  et  $\beta = \pi/6$  (ce qui correspond à un noyau centré sur l'intervalle  $[0, \pi/6]$ ), 10 000 répétitions

produisent une distribution empirique des diagnostics normalisés,  $\psi_{\lambda}(I)$  et  $\psi_{\lambda}(I)$ , dont les histogrammes sont présentés à la figure 1. À partir de maintenant, représentons par  $\delta$  et  $\alpha$  les niveaux associés aux tests de pente et de convexité, respectivement. Notons que, dans ce cas, nous définissons le niveau comme étant la probabilité de rejeter  $H_0^{(j)}$  ( $j=1, 2$ ) quand  $H_0^{(j)}$  est vraie. Même si, en pratique, dans le cas de l'hypothèse de pente, nous souhaitons ne pas aboutir au rejet, nous suivons la définition stricte du niveau et supposons (pour les besoins de cette simulation) que l'hypothèse nulle  $H_0^{(1)}$  pour la pente est vérifiée. De même, l'hypothèse nulle pour la convexité est  $H_0^{(2)}$ . Les hypothèses de pente et de convexité sont évaluées indépendamment. Le tableau 1 résume les résultats en utilisant les deux noyaux décrits à la section 2, pour diverses tailles d'échantillons ; les niveaux  $\delta$ ,  $\alpha$  indiqués s'appliquent au seuil nominal de 5 %. En outre, d'autres choix de  $\mu$  et  $\beta$  (non présentés ici) ont donné des résultats semblables. Comme l'illustre la présente étude, dans les échantillons de petite taille, nous observons une asymétrie de la distribution qui semble être due à la corrélation entre  $\theta_{\lambda}(I)$  et  $\theta_{\lambda}(I^2)$ . En outre, notons que dans le cas du test de convexité, la taille est plus grande pour le noyau quartique que pour le noyau TH.

Ensuite, considérons la puissance empirique de notre diagnostic de la présence d'un seul pic. Dans ces conditions, nous évaluons la puissance basée sur un test conjoint de la pente et de la convexité. Plus précisément, nous souhaitons ne pas rejeter  $H_0^{(1)}$ , tout en rejetant simultanément  $H_0^{(2)}$ , au niveau  $\delta = \alpha = 0,05$ , et donc identifier correctement les pics spectraux. Puisque notre hypothèse nulle composite est qu'il n'existe pas de pic, l'hypothèse alternative inclut les processus tels que le processus  $AR(2)$  donné par

$$(1 - 2\rho \cos \omega B + \rho^2 B^2)X_t = \varepsilon_t \quad (11)$$

avec la variance de bruit blanc  $\sigma^2$ , associée à une fréquence fixe donnée  $\omega \in [0, \pi]$ . Le spectre associé au processus (11) est donné par  $f(\lambda) = \sigma^2 |1 - 2\rho \cos \omega e^{-i\lambda} + \rho^2 e^{-2i\lambda}|^{-2}$ , qui est maximisé à  $\lambda_0 = \cos^{-1}(\cos \omega(1 + \rho^2)/2\rho)$ . Par conséquent, nous pouvons explorer la puissance d'une procédure de dépistage de pic en procédant à une simulation selon (11) avec divers choix de  $\rho$ ,  $\omega$  et  $\sigma$ . Le tableau 2 donne les résultats de 10 000 simulations, pour diverses tailles d'échantillons, d'après le modèle cyclique  $AR(2)$  donné en (11) avec un pic à  $\mu = \pi/6$  et la largeur de bande fixée à  $\beta = \pi/6$ . La force du pic est paramétrisée au moyen de  $\rho$ , que nous faisons varier de 0,85 à 0,95 ; il est clair que  $H_0^{(1)}$  et  $H_0^{(2)}$  sont toutes deux vraies pour ce modèle. Autrement dit, il existe des pics spectraux, de

différentes hauteurs, à  $\lambda = \pi/6$ . Ce modèle  $AR$  a été choisi parce qu'il donne une paramétrisation commode de la localisation et de la forme du pic spectral. De plus, ce choix de  $\beta$  est compatible avec les conditions de désaisonnalisation, car il fournit la largeur de fenêtre maximale tout en évitant le chevauchement des pics spectraux. Comme prévu, la puissance de notre diagnostic augmente avec la taille de l'échantillon et la pointicité, variant de 0,227 (noyau quartique) dans les petits échantillons possédant un faible pic spectral à  $\approx 0,95$  (noyau TH) dans les échantillons plus grands ayant un pic spectral plus prononcé (voir le tableau 2). Notons que, dans cette procédure, la valeur de la variance d'innovation est fixée à 1, mais qu'elle est négligeable en raison de la normalisation du diagnostic. En résumé, les noyaux quartique et TH possèdent tous deux des propriétés de taille et de puissance convenables. Généralement, le noyau quartique semble avoir une taille et une puissance supérieures, si bien qu'on le préférerait pour des spectres de cette forme (notons que la puissance plus faible du noyau TH est due en partie au fait que sa taille est trop petite). En outre, il semble que les valeurs plus faibles de  $\beta$  (résultats non présentés) requièrent une plus grande taille d'échantillon ; un  $\beta$  plus petit correspond à des « conditions d'observation » plus fines du pic spectral, qui nécessiterait une plus grande quantité de données pour traiter la résolution.

Bien que le scénario du test de la présence de pics individuels constitue le fondement de notre cadre de test conjoint, comme nous l'avons mentionné, ce cadre représente une méthodologie importante pour les applications aux statistiques fédérales. L'application vraiment importante est l'exploration de la saisonnalité résiduelle en vue de déterminer si la désaisonnalisation est efficace. Il est donc particulièrement intéressant de savoir comment notre approche de tests multiples se comporte dans la simulation. Par conséquent, pour étudier la taille et la puissance associées à notre test conjoint, nous avons simulé 10 000 répétitions à partir d'un processus de bruit blanc gaussien et d'un modèle  $AR(25)$  obtenu par un ajustement à la série sur l'emploi courant (hommes occupés, de 16 à 19 ans). L'objectif de notre étude de puissance était de construire un processus  $AR(p)$  (à cause de la facilité de son utilisation dans les simulations et de ses propriétés théoriques désirables en tant qu'estimateur spectral paramétrique – voir Parzen 1983) avec des pics spectraux (stationnaires) qui sont réalistes, ou qui s'approchent de ce qu'on pourrait trouver en pratique. Donc, nous obtenons notre modèle  $AR(25)$  – ajusté par la méthode du maximum de vraisemblance en utilisant le critère  $AIC$  – qui possède la même dynamique saisonnière (comportement spectral local) que la série sur l'emploi courant (CES pour *Current Employment Series*).

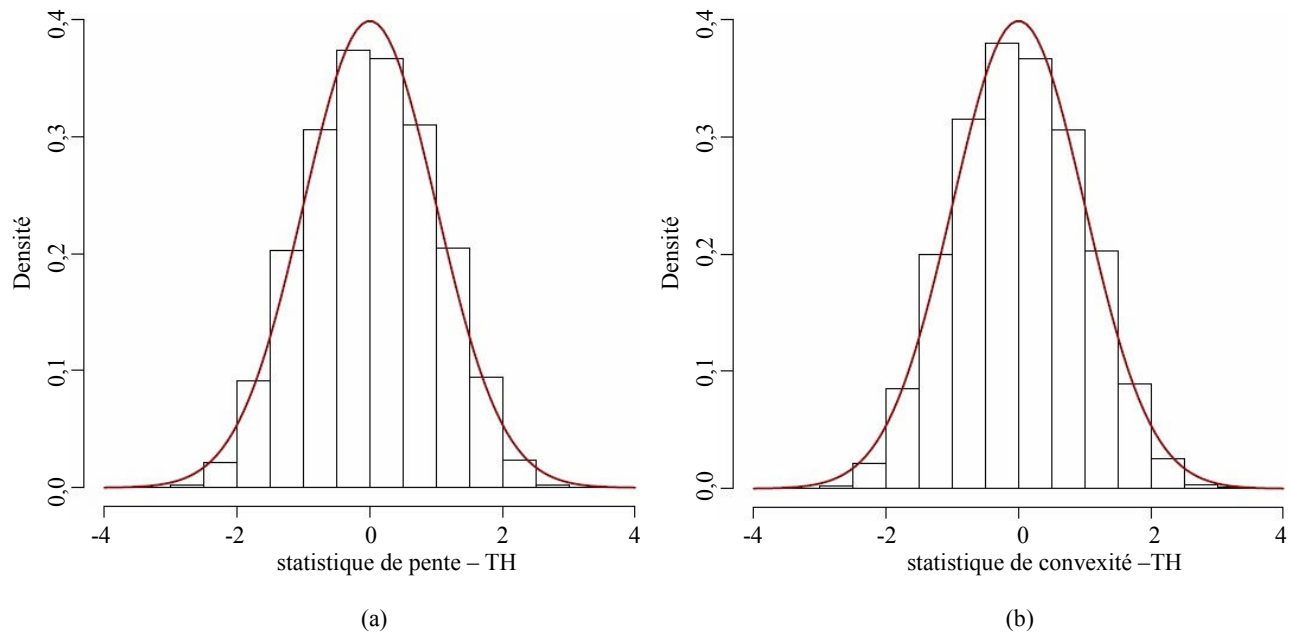


Figure 1 Histogramme de la distribution de  $-\sqrt{n}\psi_{\lambda_{\beta,\mu}}(I)$  (a) et de  $\sqrt{n}\psi_{\lambda_{\beta,\mu}}(I)$  (b) sous une hypothèse nulle de bruit blanc gaussien en utilisant le noyau TH. La taille d'échantillon est  $n = 360$  avec 10 000 répétitions.

Tableau 1

Résultats de la simulation de la taille pour le diagnostic d'un pic unique. Ici  $\mu = \beta = \pi/6$  et 10 000 répétitions sont utilisées. Les diagnostics de pente et de convexité ont été étudiés séparément pour les noyaux quartique et TH

Taille pour un pic unique $\mu = \beta = \pi/6$												
$n$	Pente						Convexité					
	Noyau quartique			Noyau TH			Noyau quartique			Noyau TH		
	Moyenne	Écart-type	Niveau $\delta$	Moyenne	Écart-type	Niveau $\delta$	Moyenne	Écart-type	Niveau $\alpha$	Moyenne	Écart-type	Niveau $\alpha$
120	0,003	0,903	0,007	-0,011	0,903	0,008	-0,065	0,852	0,032	0,025	0,888	0,018
144	-0,004	0,920	0,014	-0,011	0,927	0,015	-0,077	0,882	0,042	0,006	0,892	0,025
180	-0,003	0,942	0,022	0,002	0,920	0,017	-0,071	0,892	0,043	0,005	0,902	0,028
288	0,003	0,954	0,027	-0,002	0,950	0,025	-0,072	0,921	0,051	-0,006	0,926	0,033
360	0,003	0,962	0,032	-0,009	0,954	0,031	-0,056	0,922	0,051	0,006	0,951	0,040

Tableau 2

Résultats de la simulation de la puissance pour le diagnostic d'un pic unique. Ici  $\mu = \beta = \pi/6$  et 10 000 répétitions sont utilisées. L'hypothèse alternative est donnée par le modèle  $AR(2)$  défini par (11). Les diagnostics de pente et de convexité ont été étudiés simultanément pour les noyaux quartique et TH en utilisant  $\delta = \alpha = 0,05$  pour les deux tests (voir la section 4.1)

Puissance pour un pic unique $\mu = \beta = \pi/6 - (\delta, \alpha) = (0,05, 0,05)$						
$n$	Noyau quartique			Noyau TH		
	$\rho = 0,85$	$\rho = 0,90$	$\rho = 0,95$	$\rho = 0,85$	$\rho = 0,90$	$\rho = 0,95$
120	0,227	0,438	0,758	0,147	0,335	0,670
144	0,287	0,532	0,856	0,208	0,431	0,799
180	0,354	0,643	0,923	0,272	0,567	0,901
288	0,447	0,755	0,949	0,372	0,706	0,950
360	0,601	0,872	0,937	0,537	0,859	0,948

Pour évaluer la taille, nous avons considéré un test basé sur la convexité seulement ainsi qu'un test basé simultanément sur la pente et la convexité. Les tests portant sur la convexité seulement (C) sont exécutés aux seuils de signification nominaux  $\alpha$  de 0,05 et 0,10, en utilisant la méthode H-FWER pour contrôler le taux d'erreur de type I global. Les tests fondés simultanément sur la pente et la convexité (S, C) ont été exécutés de la façon suivante :

1. Exécuter les tests multiples de convexité,  $H_0^{(2)}$ , en utilisant la méthode H-FWER pour contrôler le taux d'erreur global au niveau  $\alpha$  (qui est égal à 0,05 ou 0,10).
2. Pour tout pic significatif découvert à l'étape 1, exécuter le test de pente individuel,  $H_0^{(1)}$ , au niveau  $\delta$  (qui est égal à 0,10 ou 0,25). Notons qu'ici, nous souhaitons ne pas rejeter  $H_0^{(1)}$  afin de déclarer n'importe quel « pic » comme étant statistiquement significatif.
3. Déclarer qu'il existe un pic statistiquement significatif si l'étape 1 permet de découvrir une fréquence saisonnière dont la convexité agrégée dans le spectre est significative et que l'étape 2 n'aboutit pas simultanément à la découverte d'une pente agrégée significative pour la fréquence saisonnière correspondante.

Les résultats de cette simulation sont résumés au tableau 3. Un aspect de cette procédure qui nécessite une explication plus approfondie est l'étape 2, où  $\delta$  (le niveau pour le test de pente) est pris égal à 0,10 et à 0,25. Bien que la partie test de pente de la procédure soit exécutée pour des pics individuels, il semble raisonnable de vouloir être prudent. La question est que, même si certains tests de pente individuels sont rejetés, nous pouvons encore poursuivre dans d'autres cas. Donc la situation observée ici diffère de l'hypothèse classique d'« absence de pics » qui peut être

rejetée si un seul pic est découvert. Évidemment, puisque nous effectuons chaque test d'hypothèse de pente sur la base de pics individuels, tout niveau  $\delta$  supérieur à 0,05 sera considéré comme plus modéré.

Si, dans le cas de la procédure combinée (S, C), nous ne pouvons pas nous attendre à ce que la taille s'approche de la valeur nominale (parce que l'utilisation du test de pente déséquilibre le taux d'erreur de type I), cette taille n'est pas extrêmement exacte non plus dans le cas de l'utilisation du test de convexité seulement (C), comme le montre l'examen du cas  $\alpha = 0,10$  avec  $n = 288, 360$ . Ici, pour le noyau quartique, la convexité est de trop grande taille, tandis que dans le cas d'un pic unique, le test de convexité possède une taille exacte (tableau 1) pour ces tailles d'échantillon. Notons que la méthode H-FWER produit uniquement une procédure approximativement correctement dimensionnée ; un autre facteur est que les tests pour les cinq pics ne sont qu'asymptotiquement indépendants. Toutes ces raisons font que la taille empirique observée au tableau 3 diffère quelque peu des niveaux nominaux.

Pour étudier la puissance, nous avons considéré la même procédure en trois étapes décrite plus haut. Cependant, pour cette simulation, nous n'avons envisagé que le test conjoint pente-convexité et examiné quatre paires de niveaux  $(\delta, \alpha)$ , à savoir  $(\delta, \alpha) = (0,10, 0,05), (0,25, 0,05), (0,10, 0,10)$  et  $(0,25, 0,10)$ . Les résultats de cette simulation (tableau 4) révèlent une puissance énorme, même pour des tailles d'échantillon aussi petites que  $n = 120$ . Il s'agit d'une propriété extrêmement importante, car  $n = 120$  est représentatif des tailles d'échantillon observées en pratique lorsque l'on procède à la désaisonnalisation (par exemple données mensuelles couvrant 10 années). Pour des échantillons de taille  $n = 144$ , une puissance supérieure à 90 % est réalisée.

**Tableau 3**

**Résultats de la simulation de la taille pour le diagnostic de pics multiples. Ici, 10 000 répétitions ont été utilisées. Le test de convexité a été étudié séparément avec le taux d'erreur de type I global contrôlé à  $\alpha = 0,05$  et à  $\alpha = 0,10$  en utilisant la méthode H-FWER. En outre, les diagnostics de pente et de convexité ont été étudiés simultanément en utilisant la méthode H-FWER pour la convexité en contrôlant le taux d'erreur de type I global à  $\alpha = 0,05$  et à  $\alpha = 0,10$ , tandis que la pente a été évaluée à  $\delta = 0,5$  et  $\delta = 0,10$ . Les noyaux quartique et TH ont tous deux été utilisés pour les deux tests (voir la section 4.1)**

n	Taille pour H-FWER pour pics multiples											
	C = 0,05		(S, C) = (0,10, 0,05)		(S, C) = (0,25, 0,5)		C = 0,10		(S, C) = (0,10, 0,10)		(S, C) = (0,25, 0,10)	
	Quartique	TH	Quartique	TH	Quartique	TH	Quartique	TH	Quartique	TH	Quartique	TH
120	0,006	0,002	0,006	0,002	0,005	0,002	0,076	0,047	0,070	0,044	0,076	0,046
144	0,009	0,002	0,011	0,002	0,008	0,003	0,087	0,053	0,090	0,051	0,086	0,050
180	0,019	0,005	0,020	0,006	0,017	0,005	0,107	0,062	0,097	0,059	0,093	0,057
288	0,031	0,009	0,026	0,008	0,025	0,008	0,117	0,069	0,116	0,069	0,112	0,068
360	0,042	0,012	0,045	0,019	0,035	0,015	0,140	0,087	0,133	0,084	0,129	0,087

Tableau 4

Résultats de la simulation de puissance pour le diagnostic de pics multiples. Ici, 10 000 répétitions ont été utilisées. Les diagnostics de pente et de convexité ont été étudiés simultanément en utilisant la méthode H-FWER pour la convexité en contrôlant le taux global d'erreur de type I à  $\alpha = 0,05$  et à  $\alpha = 0,10$ . La pente a été évaluée à  $\delta = 0,25$  et  $\delta = 0,10$ . Les noyaux quartique et TH ont tous deux été utilisés pour les deux tests (voir la section 4.1)

<i>n</i>	Puissance pour H-FWER pour pics multiples							
	(0,10, 0,05)		(0,25, 0,05)		(0,10, 0,10)		(0,25, 0,10)	
	Quartique	TH	Quartique	TH	Quartique	TH	Quartique	TH
120	0,877	0,897	0,860	0,881	0,997	0,997	0,996	0,998
144	0,943	0,952	0,942	0,950	0,999	1,00	0,999	1,00
180	0,989	0,992	0,989	0,992	1,00	1,00	0,999	1,00
288	1,00	1,00	0,999	0,999	1,00	1,00	0,999	1,00
360	1,00	1,00	0,998	0,999	1,00	1,00	0,998	0,999

## 4.2 Études de cas

Nous avons également examiné 130 séries chronologiques, 65 provenant du U.S. Census Bureau et 65, de l'OCDE. Ces séries comprennent 35 séries américaines sur la fabrication, 10 séries américaines sur le logement, 10 séries américaines sur les importations et exportations et 10 séries américaines sur le commerce de détail, ainsi que 22 séries allemandes, 15 séries pour la zone euro, 11 séries françaises et 17 séries britanniques provenant de l'OCDE et portant sur les secteurs de la fabrication, du commerce de détail, du commerce de gros, du commerce extérieur, du chômage et de l'industrie. Pour chaque série, nous avons calculé les tests de pics saisonniers pour les données brutes (log-transformées) et les données désaisonnalisées (log-transformées) – en utilisant la spécification  $x_{11}$  de X-12-ARIMA – avec le noyau quartique ainsi que le noyau TH. Nous avons appliqué la procédure H-FWER en contrôlant le taux d'erreur de type I global à  $\alpha = 0,05$  et  $0,10$  et où le seuil pour les tests de pente était  $\delta = 0,25$  et  $\delta = 0,10$  à chaque pic (voir la section 4.1). Notons que  $\delta = 0,25$  et  $\delta = 0,10$  produisent des résultats semblables, si bien que, par souci de concision, nous ne présentons ici que les résultats pour  $\delta = 0,10$ . Les résultats pour  $\delta = 0,25$  peuvent être obtenus sur demande auprès du premier auteur. Pour les données brutes ainsi que désaisonnalisées, nous n'avons effectué qu'une seule différentiation de la tendance (comme cela est le cas pour le diagnostic de signification visuelle décrit plus bas) avant d'appliquer les tests de dépistage des pics saisonniers.

En outre, nous présentons les statistiques M7 et M8, ainsi que les résultats du diagnostic de signification visuelle (SV) avant et après la désaisonnalisation. La statistique de contrôle de la qualité M7 mesure la quantité de saisonnalité stable relativement à la saisonnalité mobile dans la série originale, les valeurs supérieures à 1 indiquant que la saisonnalité dans la série n'est pas identifiable (Lothian et Morry 1978); de même, la statistique M8 mesure la taille

des fluctuations dans la composante saisonnière, avec une interprétation similaire. Nous avons également considéré le test non paramétrique robuste de Kruskal-Wallis (U.S. Census Bureau en 2002) pour la présence de saisonnalité sous l'hypothèse de stabilité. Le test SV est fondé sur une estimation spectrale  $AR(30)$  des séries brute et désaisonnalisée, et est décrit dans Soukup et Findley (1999). Pour les tableaux 5 à 10, chaque entrée de cellule énumère les fréquences saisonnières pour lesquelles un pic significatif a été découvert, avec  $j$  correspondant à  $\pi j/6$  pour  $j = 1, 2, 3, 4, 5$ ; une entrée  $\emptyset$  indique qu'aucun pic n'a été détecté. Pour les diagnostics M7 et M8, seule la valeur est consignée, puisqu'il n'existe aucune valeur  $p$  connexe (et qu'ils ne s'appliquent qu'à la série brute).

Les résultats de cette étude empirique sont présentés dans les tableaux 5 à 10. Toutes les valeurs de la statistique de Kruskal-Wallis étaient significatives avec  $p = 0,000$ , si bien qu'elles ne sont pas présentées dans les tableaux. L'ensemble de colonnes qui correspond à l'en-tête « Données originales » peut être considéré comme donnant la puissance empirique (pour chaque sous-ensemble de séries), en supposant que chaque série est effectivement saisonnière et qu'elle présente des pics spectraux saisonniers. Autrement dit, le total de résultats « corrects » donne la proportion de fois que chaque méthode détecte correctement la saisonnalité et cette proportion est donc une approximation grossière de la puissance empirique. Nous présentons aussi le nombre moyen de pics qui ont été détectés, ce qui constitue une mesure empirique de l'efficacité des méthodes (la méthode est d'autant meilleure que le nombre de pics détectés correctement est grand). L'ensemble de colonnes sous l'en-tête « Données désaisonnalisées » donne une taille empirique (pour chaque sous-ensemble de séries), en supposant que la désaisonnalisation a effectivement éliminé les pics spectraux. Il s'agit de considérations grossières, puisque nous ne savons pas vraiment a priori si les données désaisonnalisées ont été corrigées adéquatement.

Le test SV indique que toutes les séries brutes sont saisonnières et que la plupart des séries désaisonnalisées n'ont aucun pic spectral saisonnier ; les diagnostics M7 et M8 donnent des résultats comparables, bien qu'évidemment, ils n'indiquent pas quels pics saisonniers sont présents dans les données brutes. Notre procédure révèle quelques cas (quand  $\alpha = 0,10$  pour les tests de convexité) où la correction peut être inadéquate, mais ces cas sont compris dans la fourchette de la proportion attendue d'erreurs de type I. Pour les séries brutes, la puissance empirique (c'est-à-dire la proportion totale de résultats corrects) pour notre méthode varie de 0,66 à 0,89, avec une puissance plus élevée pour le niveau  $\alpha = 0,10$ , comme il fallait s'y attendre. Dans de nombreux cas, les pics indiqués sont les mêmes que ceux détectés par le test SV, mais parfois ils sont assez différents. Notons que le nombre moyen de pics détectés pour les séries brutes est habituellement beaucoup plus élevé dans le cas de notre procédure que dans celui de la méthode SV, qui produit souvent un nombre moyen de l'ordre de 3,2. Lorsque le seuil  $\alpha$  passe de 0,05 à 0,10, notre méthode produit naturellement un nombre moyen plus élevé de pics détectés ; le test SV ne peut pas être ajusté de cette façon. Inversement, pour les données désaisonnalisées, le nombre moyen de pics détectés a tendance à être inférieur à 1 dans le cas de notre méthode (sauf pour les séries allemandes).

Les résultats sont assez semblables pour les noyaux quartique et TH. Bien que les diagnostics M7, M8 et SV donnent des résultats un peu meilleurs que notre procédure de détection des pics spectraux pour  $\alpha = 0,10$ , il est important de souligner que notre méthode fournit un niveau de détail que les tests M7 et M8 ne peuvent pas reproduire, tandis que le diagnostic SV ne fournit de valeur  $p$  pour aucun des pics (M7 ou M8 ne le font pas non plus). Dans l'ensemble, nous pensons que les résultats sont très encourageants et informatifs.

## 5. Conclusion

Le présent article décrit une approche novatrice de détection statistique des pics spectraux. Le diagnostic de convexité correspond au calcul d'une moyenne du périodogramme pondérée par la dérivée seconde d'un noyau typique, tel que la fenêtre des décalages de Tukey-Hanning. Implicitement, ce type de statistique comprend une comparaison d'une moyenne du périodogramme près d'une fréquence donnée à sa moyenne un peu plus en dehors ; cela découle de la forme générale de  $\hat{A}_{\beta, \mu}$ . Le diagnostic de pente aide à dépister les cas où la convexité est négative, mais où il existe aussi une augmentation/diminution importante dans le spectre. Le fait que la méthode fonctionne effectivement comme prévu est prouvé par les simulations et les résultats d'analyse présentés dans les tableaux 1 à 10.

Pour le scénario de test de détection de pics multiples, nous employons des résultats connus extraits de la littérature sur les tests multiples (c'est-à-dire les applications pour contrôler le taux global d'erreur de type I) pour combiner les valeurs  $p$  issues des cinq fréquences saisonnières de manière à accroître considérablement la puissance statistique, comme il est démontré au tableau 4. Malgré un certain écart dans la taille (tableau 3) pour le test de détection de pics multiples, les résultats demeurent relativement utilisables. Sur un lot type de séries saisonnières, le nombre d'erreurs de type I est celui auquel on s'attendrait et la puissance est assez décente (tableaux 5 à 10). Bien que notre méthode se compare assez favorablement aux diagnostics SV, M7 et M8, aucun des diagnostics ne fournit une valeur  $p$  et seul le premier permet de discerner quels pics spectraux contribuent au comportement saisonnier. Il s'agit d'un aspect important pour le désaisonnalisateur qui veut savoir non seulement s'il existe une saisonnalité résiduelle, mais aussi à quelles fréquences saisonnières, afin de pouvoir prendre les mesures appropriées pour modifier les filtres de désaisonnalisation (ce qui peut se faire en lissant sur des années supplémentaires, en changeant les filtres saisonniers dans X-11-ARIMA ; ou bien, on pourrait envisager de raccourcir la série. Pour une description des travaux de recherche courants sur une approche fondée sur un modèle pour concevoir des filtres de désaisonnalisation ciblés sur des fréquences saisonnières particulières, voir Aston, Findley, McElroy, Wills et Martin (2007)).

Le choix du noyau a certainement une incidence sur les résultats, quoiqu'en pratique, nous ne constatons que peu de différence entre les noyaux quartique et TH ; ce dernier pourrait être marginalement plus puissant. Évidemment, de nombreux autres noyaux d'usage répandu pourraient aussi être utilisés par un praticien et nous en avons seulement choisi deux qui semblaient intuitifs et simples à mettre en œuvre. Le choix de l'emplacement  $\mu$  est clairement dicté par la caractérisation de la saisonnalité. Puisque la puissance statistique diminue généralement avec  $\beta$ , nous recommandons systématiquement de prendre la valeur maximale de  $\beta$  de sorte que les supports de noyaux soient disjoints, ce qui garantit la propriété d'indépendance asymptotique des divers diagnostics qui est un élément crucial de notre méthode de tests multiples.

Enfin, les résultats asymptotiques requièrent que les données soient différenciées jusqu'à la stationnarité. Comme les séries chronologiques économiques sont habituellement non stationnaires, il est souhaitable de différencier la tendance des données désaisonnalisées avant d'appliquer notre diagnostic. Cette différenciation pouvant atténuer la détection du premier pic saisonnier, les praticiens peuvent « recolorer » les données comme il est décrit à la section 3.4.

Tableau 5

Analyses des données pour 35 séries sur la fabrication (U.S. Census Bureau) comparant notre diagnostic de pics multiples avec les diagnostics SV, M7 et M8. Notre diagnostic de pics multiples s'appuie sur la méthode H-FWER pour contrôler le taux d'erreur de type I global à  $\alpha = 0,05$  et  $\alpha = 0,10$  et examine la pente à  $\delta = 0,10$  (voir la section 4.2)

Analyses des données – Séries sur la fabrication												
Série	Données originales					Données désaisonnalisées						
	H-FWER 0,10/0,05		H-FWER 0,10/0,10		SV	M7	M8	H-FWER 0,10/0,05		H-FWER 0,10/0,10		SV
	Quartique	TH	Quartique	TH				Quartique	TH	Quartique	TH	
$M_1$	2345	1235	2345	1235	12	0,24	0,39	∅	∅	∅	∅	∅
$M_2$	12345	12345	12345	12345	1235	0,20	0,32	∅	∅	∅	∅	∅
$M_3$	12345	12345	12345	12345	1235	0,28	0,46	∅	∅	∅	∅	∅
$M_4$	∅	∅	12345	∅	12	0,28	0,44	∅	∅	∅	∅	∅
$M_5$	12345	12345	12345	12345	12345	0,27	0,47	∅	∅	12345	12345	∅
$M_6$	∅	∅	123	123	12	0,28	0,49	∅	∅	∅	∅	∅
$M_7$	∅	∅	123	123	24	0,50	0,79	∅	∅	∅	∅	∅
$M_8$	12345	12345	12345	12345	12345	0,18	0,37	∅	∅	∅	∅	∅
$M_9$	12345	12345	12345	12345	124	0,42	0,73	∅	∅	∅	∅	∅
$M_{10}$	∅	∅	∅	∅	1	0,38	0,72	∅	∅	∅	∅	∅
$M_{11}$	∅	∅	12345	1234	123	0,15	0,27	∅	∅	∅	∅	∅
$M_{12}$	1234	1234	12345	12345	1234	0,30	0,54	∅	∅	∅	∅	∅
$M_{14}$	∅	∅	1234	1234	1234	0,24	0,39	∅	∅	∅	∅	∅
$M_{15}$	12345	12345	12345	12345	12345	0,23	0,43	∅	∅	∅	∅	∅
$M_{16}$	1234	1234	1234	1234	1234	0,23	0,40	∅	∅	∅	∅	∅
$M_{17}$	∅	∅	1234	12345	12	0,64	0,66	∅	∅	∅	∅	∅
$M_{18}$	12345	12345	12345	12345	245	0,20	0,37	∅	∅	∅	∅	∅
$M_{19}$	∅	∅	∅	∅	4	0,86	1,00	∅	∅	∅	∅	∅
$M_{20}$	∅	∅	∅	12345	4	0,56	0,84	∅	∅	∅	∅	∅
$M_{21}$	12345	12345	12345	12345	1234	0,37	0,58	∅	∅	∅	∅	∅
$M_{22}$	12345	12345	12345	12345	1234	0,26	0,45	∅	∅	∅	∅	∅
$M_{23}$	12345	12345	12345	12345	1234	0,20	0,47	∅	∅	∅	∅	∅
$M_{24}$	12345	12345	12345	12345	2345	0,26	0,43	∅	∅	∅	∅	∅
$M_{25}$	12345	12345	12345	12345	12345	0,27	0,42	∅	∅	∅	∅	∅
$M_{26}$	12345	12345	12345	12345	1235	0,37	0,62	∅	∅	∅	∅	∅
$M_{27}$	1345	1234	1345	1234	2345	0,25	0,22	∅	∅	∅	∅	∅
$M_{28}$	∅	∅	∅	∅	24	0,57	0,44	∅	∅	∅	∅	∅
$M_{29}$	∅	12345	12345	12345	24	0,78	1,13	∅	∅	∅	∅	∅
$M_{30}$	123	1234	12345	12345	245	0,45	0,65	∅	∅	∅	∅	∅
$M_{31}$	∅	∅	123	123	4	0,64	0,46	∅	∅	1234	1234	∅
$M_{32}$	1235	12345	1235	12345	12345	0,21	0,37	∅	∅	∅	∅	∅
$M_{33}$	12345	12345	12345	1234	1234	0,24	0,38	∅	∅	∅	∅	∅
$M_{34}$	12345	12345	12345	12345	234	0,46	0,85	∅	∅	∅	∅	∅
$M_{35}$	12345	12345	12345	12345	2345	0,25	0,66	∅	∅	∅	∅	∅
$M_{36}$	12345	12345	12345	12345	123	1,32	1,56	∅	∅	∅	∅	∅
Total de résultats	23/35	24/35	31/35	31/35	35/35	34/35	33/35	35/35	35/35	33/35	33/35	35/35
« corrects »												
Nombre moyen	3,09	3,29	4,09	4,09	3,23			0	0	0,26	0,26	0

Tableau 6

Analyses des données pour 30 séries du U.S. Census Bureau (10 sur le logement, 10 sur les importations/exportations et 10 sur les ventes au détail) comparant notre diagnostic de pics multiples avec les diagnostics SV, M7 et M8. Notre diagnostic de pics multiples s'appuie sur la méthode H-FWER pour contrôler le taux d'erreur de type I global à  $\alpha = 0,05$  et  $\alpha = 0,10$  et examine la pente à  $\delta = 0,10$  (voir la section 4.2)

Analyses des données – Séries sur la fabrication												
Série	Données originales					Données désaisonnalisées						
	H-FWER 0,10/0,05		H-FWER 0,10/0,10		SV	M7	M8	H-FWER 0,10/0,05		H-FWER 0,10/0,10		SV
	Quartique	TH	Quartique	TH				Quartique	TH	Quartique	TH	
MW1Fam	12345	12345	12345	12345	12	0,13	0,25	∅	∅	∅	∅	∅
NW1Tot	12345	12345	12345	12345	12	0,18	0,31	∅	∅	∅	∅	∅
NE1Fam	12345	12345	12345	12345	12	0,16	0,33	∅	∅	∅	∅	∅
NE1Tot	12345	12345	12345	12345	123	0,25	0,27	∅	∅	∅	∅	∅
S1Fam	12345	12345	12345	12345	125	0,22	0,47	∅	∅	∅	∅	∅
STot	124	124	1245	1245	125	0,29	0,57	∅	∅	∅	∅	∅
US1Fam	12345	12345	12345	12345	125	0,17	0,39	∅	∅	∅	∅	∅
USTot	12345	12345	12345	12345	125	0,20	0,42	∅	∅	∅	∅	∅
W1Fam	1234	1234	12345	12345	125	0,21	0,44	∅	∅	∅	∅	∅
WTot	1234	1234	12345	1234	12	0,27	0,56	∅	∅	∅	∅	∅
Total de résultats « corrects »	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10
Séries sur les importations/exportations												
Série	Données originales					Données désaisonnalisées						
	H-FWER 0,10/0,05		H-FWER 0,10/0,10		SV	M7	M8	H-FWER 0,10/0,05		H-FWER 0,10/0,10		SV
	Quartique	TH	Quartique	TH				Quartique	TH	Quartique	TH	
M00120	12345	∅	12345	12345	125	0,23	0,48	∅	∅	∅	∅	∅
M00190	12345	12345	12345	12345	1235	0,38	0,59	∅	∅	∅	∅	∅
M3000	12345	12345	12345	12345	234	0,48	0,95	∅	∅	∅	∅	∅
M3010	1234	1234	1234	12345	2345	0,52	0,88	∅	∅	∅	∅	∅
M12060	12345	12345	12345	12345	123	0,53	0,77	∅	∅	∅	∅	∅
X3	12345	12345	12345	12345	2345	0,57	0,94	∅	∅	∅	∅	∅
X00300	134	134	134	134	2	0,56	0,97	∅	∅	∅	∅	∅
X3020	12345	12345	12345	12345	12345	0,39	0,70	∅	∅	∅	∅	∅
X3022	12345	12345	12345	12345	23	0,69	1,04	∅	∅	∅	∅	∅
X10140	1234	1234	1234	1234	15	0,29	0,47	∅	∅	∅	∅	∅
Total de résultats « corrects »	10/10	9/10	10/10	10/10	10/10	10/10	9/10	10/10	10/10	10/10	10/10	10/10
Séries sur les ventes de détail												
Série	Données originales					Données désaisonnalisées						
	H-FWER 0,10/0,05		H-FWER 0,10/0,10		SV	M7	M8	H-FWER 0,10/0,05		H-FWER 0,10/0,10		SV
	Quartique	TH	Quartique	TH				Quartique	TH	Quartique	TH	
s0b441x0	12345	12345	12345	12345	135	0,22	0,41	∅	∅	∅	∅	∅
s0b 44000	12345	12345	12345	12345	2345	0,12	0,26	∅	∅	∅	∅	∅
s0b 44100	12345	12345	12345	12345	135	0,21	0,40	∅	∅	∅	∅	∅
s0b 44130	12345	12345	12345	12345	1235	0,21	0,42	∅	∅	∅	∅	∅
s0b 44200	12345	12345	12345	12345	12345	0,13	0,27	∅	∅	∅	∅	∅
s0b 44300	1234	12345	1234	12345	12345	0,12	0,18	∅	∅	∅	∅	∅
s0b 44312	1234	1234	1234	1234	12345	0,31	0,48	∅	∅	∅	∅	∅
s0b 44400	12345	12345	12345	12345	1235	0,16	0,32	∅	∅	∅	∅	∅
s0b 44410	12345	12345	12345	12345	1235	0,14	0,32	∅	∅	∅	∅	∅
s0b 44500	12345	12345	12345	12345	235	0,14	0,23	∅	∅	∅	∅	∅
Total de résultats « corrects »	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10	10/10
Total général des résultats « corrects »	30/30	29/30	30/30	30/30	30/30	30/30	29/30	30/30	30/30	30/30	30/30	30/30
Nombre moyen	4,67	4,5	4,77	4,8	3,23			0	0	0	0	0



Tableau 7

Analyses des données pour 22 séries allemandes de l'OCDE comparant notre diagnostic de pics multiples avec les diagnostics SV, M7 et M8. Notre diagnostic de pics multiples s'appuie sur la méthode H-FWER pour contrôler le taux d'erreur de type I global à  $\alpha = 0,05$  et  $\alpha = 0,10$  et examine la pente à  $\delta = 0,10$  (voir la section 4.2)

Analyses des données – OCDE, Allemagne												
Série ALLEMAGNE	Données originales				Données désaisonnalisées							
	H-FWER 0,10/0,05		H-FWER 0,10/0,10		SV	M7	M8	H-FWER 0,10/0,05		H-FWER 0,10/0,10		SV
	Quartique	TH	Quartique	TH				Quartique	TH	Quartique	TH	
PRMNCG03	12345	12345	12345	12345	12345	0,17	0,33	∅	∅	∅	∅	∅
PRMNCS01	1234	1234	1234	1234	235	0,25	0,79	∅	∅	∅	∅	∅
PRMNIG01	12345	12345	12345	12345	12345	0,28	0,48	∅	∅	∅	∅	∅
PRMNTO01	134	134	134	134	12345	0,26	0,46	∅	∅	∅	∅	∅
PRMNVG01	1235	1235	1235	1235	2345	0,29	0,46	∅	∅	∅	∅	∅
SLMNCD01	1245	1235	1245	1235	23	0,22	0,40	∅	∅	∅	∅	∅
SLMN CN01	1345	2345	1345	2345	123	0,37	0,72	∅	∅	∅	∅	∅
SLMN DM01	2345	2345	2345	2345	123	0,32	0,63	∅	∅	∅	∅	∅
SLMN EX01	12345	12345	12345	12345	12	0,32	0,51	1:5	∅	12345	12345	∅
SLMN IG01	2345	2345	2345	2345	123	0,21	0,65	∅	∅	∅	∅	∅
SLMN TO01	245	345	245	345	23	0,20	0,66	∅	∅	234	∅	∅
SLR TCR01	1345	1345	1345	1345	1234	0,19	0,51	∅	∅	∅	∅	∅
SLR TTO01	12345	12345	12345	12345	12345	0,12	0,25	∅	∅	∅	∅	∅
SLR TTO02	12345	12345	12345	12345	12345	0,13	0,29	∅	∅	∅	∅	∅
SLW HTO01	2345	2345	2345	2345	123	0,20	0,62	∅	∅	134	123	∅
SLW HTO02	2345	2345	2345	2345	123	0,20	0,62	∅	∅	134	123	∅
UNL VRG01	23	23	23	23	124	0,23	0,48	∅	∅	1	∅	5
UNL VSUMA	345	345	345	345	12	0,30	0,53	12	∅	1245	12345	∅
UNL VSUTT	234	234	234	234	12	0,24	0,53	∅	∅	45	45	25
UNR TRG01	235	1245	235	1245	123	0,19	0,59	∅	∅	2345	2345	∅
XTEXVA01	1234	1234	1234	1234	23	0,28	0,77	∅	∅	∅	∅	∅
XTIMVA01	234	1234	2345	12345	23	0,31	0,95	∅	∅	∅	∅	∅
Total de résultats « corrects »	22/22	22/22	22/22	22/22	22/22	22/22	22/22	20/22	22/22	14/22	16/22	20/22
Nombre moyen	3,86	3,95	3,91	4,00	3,23			0	0	1,14	1,00	0,14

Tableau 8

Analyses des données pour 15 séries sur la zone euro de l'OCDE comparant notre diagnostic de pics multiples avec les diagnostics SV, M7 et M8. Notre diagnostic de pics multiples s'appuie sur la méthode H-FWER pour contrôler le taux d'erreur de type I global à  $\alpha = 0,05$  et  $\alpha = 0,10$  et examine la pente à  $\delta = 0,10$  (voir la section 4.2)

Analyses des données – OCDE, Euro												
Série EURO	Données originales				Données désaisonnalisées							
	H-FWER 0,10/0,05		H-FWER 0,10/0,10		SV	M7	M8	H-FWER 0,10/0,05		H-FWER 0,10/0,10		SV
	Quartique	TH	Quartique	TH				Quartique	TH	Quartique	TH	
PRCNT001	345	1345	345	1345	12345	0,14	0,44	∅	∅	∅	∅	∅
PRINT001	12345	12345	12345	12345	12345	0,10	0,23	∅	∅	∅	∅	∅
PRMNCG03	1245	1245	1245	1245	12345	0,12	0,32	1	1	1234	1234	∅
PRMNCS01	1234	12345	1234	12345	1234	0,22	0,47	∅	∅	∅	∅	∅
PRMNIG01	12345	12345	12345	12345	2345	0,15	0,27	∅	∅	∅	∅	∅
PRMNTO01	2345	2345	2345	2345	2345	0,14	0,23	∅	∅	1	∅	∅
PRMNVG01	1234	12345	1234	12345	2345	0,13	0,23	∅	∅	∅	∅	∅
SLMN CN02	12345	12345	12345	12345	123	0,31	0,57	∅	∅	∅	∅	∅

Tableau 8 (suite)

Analyses des données pour 15 séries sur la zone euro de l'OCDE comparant notre diagnostic de pics multiples avec les diagnostics SV, M7 et M8. Notre diagnostic de pics multiples s'appuie sur la méthode H-FWER pour contrôler le taux d'erreur de type I global à  $\alpha = 0,05$  et  $\alpha = 0,10$  et examine la pente à  $\delta = 0,10$  (voir la section 4.2)

Analyses des données – OCDE, Euro												
Série	Données originales						Données désaisonnalisées					
	H-FWER 0,10/0,05		H-FWER 0,10/0,10		SV	M7	M8	H-FWER 0,10/0,05		H-FWER 0,10/0,10		SV
	EURO	Quartique	TH	Quartique				TH	Quartique	TH	Quartique	
SLMNIG02	12345	12345	12345	12345	23	0,21	0,45	∅	∅	∅	∅	∅
SLMNT002	1345	2345	1345	2345	23	0,20	0,41	24	∅	24	34	∅
SLMNVG02	12345	12345	12345	12345	2345	0,17	0,30	1	1	1245	1345	∅
SLRTO01	12345	12345	12345	12345	12345	0,05	0,12	∅	∅	∅	∅	∅
SLRTO02	12345	12345	12345	12345	12345	0,05	0,11	∅	∅	∅	∅	∅
XTEXVA01	1345	2345	1345	2345	23	0,31	0,57	∅	∅	12	12	∅
XTIMVA01	2345	2345	2345	2345	23	0,40	0,72	∅	∅	∅	∅	∅
Total de résultats « corrects »	15/15	15/15	15/15	15/15	15/15	15/15	15/15	12/15	13/15	10/15	11/15	15/15
Nombre moyen	4,40	4,60	4,40	4,60	3,73			0,20	0,13	0,87	0,80	0

Tableau 9

Analyses des données pour 11 séries françaises de l'OCDE comparant notre diagnostic de pics multiples avec les diagnostics SV, M7 et M8. Notre diagnostic de pics multiples s'appuie sur la méthode H-FWER pour contrôler le taux d'erreur de type I global à  $\alpha = 0,05$  et  $\alpha = 0,10$  et examine la pente à  $\delta = 0,10$  (voir la section 4.2)

Analyses des données – OCDE, France												
Série	Données originales						Données désaisonnalisées					
	H-FWER 0,10/0,05		H-FWER 0,10/0,10		SV	M7	M8	H-FWER 0,10/0,05		H-FWER 0,10/0,10		SV
	FRANCE	Quartique	TH	Quartique				TH	Quartique	TH	Quartique	
PRAFAG01	12345	12345	12345	12345	123	0,13	0,29	∅	∅	∅	∅	∅
PRNCTO01	2345	2345	2345	2345	235	0,14	0,44	∅	∅	∅	∅	∅
PRMNCG01	12345	12345	12345	12345	234	0,15	0,38	1	1	1	1	∅
PRMNCS01	12345	12345	12345	12345	234	0,25	0,58	∅	∅	∅	∅	∅
PRMNIG01	12345	12345	12345	12345	12345	0,11	0,26	∅	∅	∅	∅	∅
PRMNTO01	12345	12345	12345	12345	12345	0,16	0,29	123	123	123	1234	∅
PRMNVE01	12345	12345	12345	12345	12345	0,24	0,34	∅	∅	1245	1245	∅
SLRTO01	1345	2345	1345	2345	123	0,27	0,71	∅	∅	∅	∅	∅
SLRTO02	12345	12345	12345	12345	12345	0,16	0,36	∅	∅	∅	∅	∅
XTEXVA01	1345	1345	1345	1345	23	0,14	0,44	∅	∅	∅	∅	∅
XTIMVA01	1245	1245	1245	1245	23	0,18	0,54	∅	∅	∅	∅	∅
Total de résultats « corrects »	11/11	11/11	11/11	11/11	11/11	11/11	11/11	9/11	9/11	8/11	8/11	11/11
Nombre moyen	4,64	4,64	4,64	4,64	3,55			0,36	0,36	0,73	0,81	0

Tableau 10

Analyses des données pour 17 séries britanniques de l'OCDE comparant notre diagnostic de pics multiples avec les diagnostics SV, M7 et M8. Notre diagnostic de pics multiples s'appuie sur la méthode H-FWER pour contrôler le taux d'erreur de type I global à  $\alpha = 0,05$  et  $\alpha = 0,10$  et examine la pente à  $\delta = 0,10$  (voir la section 4.2)

Série	Analyses des données – OCDE, Grande-Bretagne										SV	
	Données originales				Données désaisonnalisées							
	H-FWER 0,10/0,05		H-FWER 0,10/0,10		H-FWER 0,10/0,05		H-FWER 0,10/0,10		H-FWER 0,10/0,10			
GRANDE-BRETAGNE	Quartique	TH	Quartique	TH	Quartique	TH	Quartique	TH	Quartique	TH		
PPIAMP01	1234	1234	1234	1234	24	0,56	1,58	∅	∅	∅	∅	∅
PPIAMP02	1	1	123	123	2	0,53	0,92	∅	∅	∅	∅	∅
PPIPFU01	1345	12345	1345	12345	12	0,64	0,59	∅	∅	∅	∅	∅
PRINTO01	1345	1345	1345	1345	23	0,16	0,40	∅	∅	∅	∅	∅
PRMNCG02	2345	2345	2345	2345	123	0,23	0,56	∅	∅	∅	∅	∅
PRMNCG03	12345	12345	12345	12345	123	0,20	0,49	∅	∅	∅	∅	∅
PRMNCS01	123	12	123	1234	12	0,68	1,31	∅	∅	∅	∅	∅
PRMNIG01	2345	2345	2345	2345	123	0,15	0,47	1	∅	∅	∅	∅
PRMNTO01	1345	1345	1345	1345	23	0,17	0,45	∅	∅	∅	∅	∅
PRMNVE02	12345	12345	12345	12345	12345	0,25	0,76	∅	∅	∅	∅	∅
PRMNVE03	1234	1234	1234	1234	234	0,29	0,91	∅	∅	∅	∅	∅
PRMNVG01	124	134	124	134	234	0,18	0,58	∅	∅	∅	∅	∅
SLRTRC03	1345	1345	1345	1345	124	0,42	0,74	124	123	124	123	∅
SLRTOO02	12345	12345	12345	12345	12345	0,05	0,15	1234	∅	1234	12345	∅
UNLVRG01	12345	12345	12345	12345	1245	0,63	0,61	∅	∅	∅	∅	∅
XTEXVA01	134	134	1345	1345	23	0,34	1,02	∅	∅	∅	∅	∅
XTIMVA01	23	23	23	23	23	0,31	0,90	∅	∅	∅	∅	∅
Total de résultats « corrects »	17/17	17/17	17/17	17/17	17/17	17/17	14/17	14/17	16/17	13/17	15/17	17/17
Nombre moyen	3,76	3,76	3,94	4,06	2,76			0,47	0,18	0,53	0,47	0

## Remerciements

Le présent article est diffusé en vue de tenir les parties intéressées au courant de la recherche courante et de favoriser la discussion des travaux en cours. Les opinions exprimées sont celles des auteurs et ne représentent pas forcément celles du U.S. Census Bureau. Les travaux de Holan ont été financés par une bourse de recherche de l'ASA/NSF/BLS.

## Annexe

Nous établissons ici les formules asymptotiques pour les mesures statistiques  $\psi_A$  de la pente et de la convexité. Ces résultats peuvent alors être appliqués dans le paradigme de test pour obtenir des valeurs critiques asymptotiques. Certaines conditions faibles sont requises sur les données

pour la théorie asymptotique; nous nous inspirons de Taniguchi et Kakizawa (2000, section 3.1.1). La condition (B), due à Brillinger (1981), énonce que le processus est strictement stationnaire et la condition (B1) de Taniguchi et Kakizawa (2000, page 55) est vérifiée. La condition (HT), due à Hosoya et Taniguchi (1982), énonce que le processus a une représentation linéaire et les conditions (H1) à (H6) de Taniguchi et Kakizawa (2000, pages 55-56) sont vérifiées. L'hypothèse 1 (8) de Chiu (1988) est une condition de sommabilité sur divers cumulants d'ordre élevé, qui est satisfaite, par exemple, par un processus gaussien dont la densité spectrale est dans  $C^2$ . Aucune de ces conditions n'est stricte; par exemple, un processus causal linéaire avec quatre moments satisfait (HT). Le résultat principal est une convergence conjointe de toute paire de mesures  $\psi_A(I)$ ; par exemple, il peut s'agir d'une mesure de pente et de convexité avec le même noyau  $A$ . Nous présentons le théorème général qui couvre ces deux cas.

*Théorème 1* Supposons que les cumulants de quatrième ordre de  $\{X_t\}$  disparaissent ; que la condition (B) ou (HT) est vérifiée ; et que l'hypothèse 1 (8) de Chiu (1988) est vérifiée. Soit les noyaux  $A$  et  $B$  satisfaisant les conditions (i) à (iv) de la section 2.1. Alors

$$\left\{ \begin{array}{l} \sqrt{n} \frac{(\theta_A(I) - \theta_A(f))}{\sqrt{\theta_{A^2}(I^2)/2}}, \\ \sqrt{n} \frac{(\theta_B(I) - \theta_B(f))}{\sqrt{\theta_{B^2}(I^2)/2}} \end{array} \right\} \xrightarrow{\mathcal{L}} N(0, V)$$

quand  $n \rightarrow \infty$ . Ici,  $0$  dénote le vecteur nul  $(0, 0)'$ , et  $V$  est une matrice de dimension  $2 \times 2$  dont les entrées sont

$$V_{11} = V_{22} = 1 \quad V_{12} = V_{21} = \frac{\theta_{AB}(f^2)}{\sqrt{\theta_{A^2}(f^2)\theta_{B^2}(f^2)}}.$$

*Preuve.* Pour commencer, établissons que  $\theta_{A^2}(I^2) \xrightarrow{p.s.} 2\theta_{A^2}(f^2)$ . Puisque le noyau  $A$  est continu dans un intervalle (tel que  $[\mu - \beta/2, \mu + \beta/2]$ ), ce résultat découle directement du corollaire 1 de Chiu (1988), en notant qu'il traite l'approximation de la fonction intégrale par les sommes de Riemann (Chiu (1988) définit aussi le périodogramme avec un facteur  $2\pi$ ). Évidemment, les mêmes résultats tiennent avec  $B$  à la place de  $A$ . Deuxièmement, considérons la convergence conjointe de  $\theta_A(I)$  et  $\theta_B(I)$ . Nous utilisons la technique de Cramér-Wold et appliquons le lemme 3.1.1 de Taniguchi et Kakizawa (2000), généralisé comme il convient pour inclure les fonctions impaires (voir le théorème 3 de Chiu (1988)). Donc, pour tout  $x, y$  réel,

$$\sqrt{n} \left( x \frac{(\theta_A(I) - \theta_A(f))}{\sqrt{\theta_{A^2}(f^2)}} + y \frac{(\theta_B(I) - \theta_B(f))}{\sqrt{\theta_{B^2}(f^2)}} \right) \xrightarrow{\mathcal{L}} N \left( 0, \frac{1}{2\pi} \int_{-\pi}^{\pi} (C(\lambda)C(-\lambda) + C^2(\lambda)) f^2(\lambda) d\lambda \right)$$

en utilisant le théorème de Slutsky (Bickel et Doksum 1977), où le noyau  $C$  est défini par

$$C(\lambda) = \frac{x}{\sqrt{\theta_{A^2}(f^2)}} A(\lambda) + \frac{y}{\sqrt{\theta_{B^2}(f^2)}} B(\lambda).$$

De toute évidence,  $C(\lambda)C(-\lambda) = 0$  et

$$\begin{aligned} C^2(\lambda) &= \frac{x^2}{\theta_{A^2}(f^2)} A^2(\lambda) \\ &+ 2 \frac{xy}{\sqrt{\theta_{A^2}(f^2)}\sqrt{\theta_{B^2}(f^2)}} A(\lambda)B(\lambda) \\ &+ \frac{y^2}{\theta_{B^2}(f^2)} B^2(\lambda). \end{aligned}$$

En prenant  $x$  et  $y$  égaux à zéro et un dans diverses combinaisons, nous déduisons la matrice de variance énoncée  $V$ .

Ensuite, nous discutons du scénario de test de détection de pics multiples. Supposons donc que nous avons une série finie de noyaux  $A_i$  pour  $i = 1, 2, \dots, d$ , qui satisfont chacun les hypothèses de la section 2. Alors, nous pouvons facilement généraliser le théorème 1 pour passer de deux à  $d$  noyaux comme il suit. La matrice de covariance asymptotique  $V$  aura la  $ij^e$  entrée

$$\frac{\theta_{A_i A_j}(f^2)}{\sqrt{\theta_{A_i^2}(f^2)\theta_{A_j^2}(f^2)}}.$$

Donc, si le support pour n'importe quelle paire de noyaux est disjoint, nous obtenons l'indépendance asymptotique et nous pouvons par conséquent appeler la procédure de tests multiples H-FWER.

## Bibliographie

- Aston, J., Findley, D., McElroy, T., Wills, K. et Martin, D. (2007). New ARIMA Models for Seasonal Time Series and Their Application to Seasonal Adjustment and Forecasting. *SRD Research Report No. RRS 2007-14, U.S. Census Bureau*.
- Bell, W., et Hillmer, S. (1984). Issues involved with the seasonal adjustment of economic time series. *Journal of Business and Economic Statistics*, 2, 291-320.
- Bickel, P., et Doksum, K. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*. Englewood Cliffs, New Jersey : Prentice Hall.
- Benjamini, Y., et Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57, 289-300.
- Brillinger, D. (1981). *Time Series Data Analysis and Theory*. San Francisco : Holden-Day.
- Chiu, S. (1988). Weighted least squares estimators on the frequency domain for the parameters of a time series. *The Annals of Statistics*, 16, 1315-1326.
- Evans, T., Holan, S. et McElroy, T. (2006). Evaluating measures for assessing spectral peaks. *2006 Proceedings American Statistical Association*, [CD-ROM] : Alexandria, VA}.
- Findley, D. (2006). Communication personnelle.
- Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C. et Chen, B.C. (1998). New capabilities and methods of the X-12-ARIMA seasonal adjustment program. *Journal of Business and Economic Statistics*, 16, 127-177 (avec discussion).
- Hosoya, Y., et Taniguchi, M. (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes. *The Annals of Statistics*, 10, 132-153.
- Lothian, J., et Morry, M. (1978). A test for identifiable seasonality when using the X-11-ARIMA program. Document de travail, Division de la recherche et analyse en séries chronologiques, Statistique Canada.

- Maravall, A., et Caporello, G. (2004). Program TSW: Revised Reference Manual. Document de travail 2004, *Research Department, Bank of Spain*. <http://www.bde.es>.
- Nerlove, M. (1964). Spectral analysis of seasonal adjustment procedures. *Econometrica*, 32, 241-286.
- Newton, H., et Pagano, M. (1983). A method for determining periods in time series. *Journal of the American Statistical Association*, 78, 152-157.
- Parzen, E. (1983). Autoregressive spectral estimation. *Handbook of Statistics III*, (Éds., D. Brillinger et P. Krishnaiah), Amsterdam : North Holland, 221-247.
- Priestley, M. (1981). *Spectral Analysis and Time Series*. London : Academic Press.
- Soukup, R.J., et Findley, D.F. (1999). On the spectrum diagnostics used by X-12-ARIMA to indicate the presence of trading day effects after modeling or adjustment. Aussi au [www.census.gov/pub/ts/papers/tr9903s.pdf](http://www.census.gov/pub/ts/papers/tr9903s.pdf). *Proceedings of the Business and Economic Statistics Section*, American Statistical Association, 144-149.
- Taniguchi, M., et Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. New York City, New York : Springer-Verlag.
- U.S. Census Bureau (2002). X-12 ARIMA Reference Manual (Version 0.2.10), Washington, DC.