

BANQUE NATIONALE DE BELGIQUE

# Désaisonnalisation par modélisation avec JDemetra+

---

*Draft*

Jean Palate

3/27/2015

Même s'il fournit une méthode non-paramétrique (X11), le programme JDemetra+ (JD+) est fortement imprégné de l'approche d'ajustement saisonnier par modélisation stochastique. L'évolution des bibliothèques statistiques sous-jacentes au programme reflète d'ailleurs en grande partie les explorations successives dans ce domaine. Les limites de la modélisation sont bien connues : instabilité de certains modèles, complexité de la matière... sont autant de freins à son utilisation en masse dans un environnement de production. Bien utilisée, la modélisation est toutefois un outil remarquablement riche et puissant. Ce papier a pour objectif principal de décrire les méthodes de désaisonnalisation par modélisation implémentées dans JD+ ainsi que certaines caractéristiques qui les relient.

## 0.Introduction

Les modèles de désaisonnalisation peuvent prendre de multiples formes. Les solutions les plus utilisées reposent sur l'approche de type « Arima model-based » (AMB) ou font partie de la famille des modèles structurels (STS). La première solution a été popularisée par le programme SEATS (Gomez-Maravall) tandis que la seconde, décrite notamment par Harvey [1989], Durbin et Koopman [2001], Proietti [2000]... , est implémentée – en autres – dans le programme STAMP.

Dans la mesure où les différentes composantes de la série peuvent s'exprimer comme des modèles ARIMA, les approches de type AMB et STSappartiennent à la même famille de modèles, appelée « Unobserved ComponentsARIMA » (UCARIMA).

Suivant le critère habituel de « MMSE », les composantes des modèles UCARIMA peuvent être estimées et analysées suivant deux types principaux d'algorithmes. Le premier se base sur les filtres de Wiener-Kolmogorov (WK) et le second sur les filtresde Kalman (KF). SEATS s'appuie sur la première approche et STAMP sur la seconde. JD+ fournit pour sa part les deux solutions.Pour mémoire, on signalera également la possibilité d'estimer les modèles UCARIMA au moyen de formules matricielles directes (voir Mc Elroy[2008]). Cette dernière solution, généralement inefficace, est également implémentée dans JD+.

Les filtres de WK et KF reposent sur des logiques différentes : les premiers partent de l'hypothèse que l'on dispose de séries (semi-)infinies ; ils s'inscrivent donc dans une approche essentiellement théorique. Les seconds fournissent par contre une estimation exacte, en fonction des observations effectivement disponibles.

Le choix de fournir les deux approches dans JD+ repose d'abord sur la volonté de comprendre et de mesurer les différences entre les solutions exactes de KF et les solutions approchées ou théoriques de WK. Il s'agit également d'un moyen de contrôle des différents algorithmes. Plus fondamentalement, KF se révèle un outil beaucoup plus flexible que WK. Il devient ainsi un algorithme souvent indispensable lorsque l'on aborde des modèles plus complexes, et en particulier des modèles qui ne sont pas invariants dans le temps.

Dans un premier paragraphe, on décrit brièvement les modèles UCARIMA qui regroupent à la fois les modèles issus de l'approche AMB et les modèles STS de JD+. On présente ensuite la décomposition canonique sous-jacente aux modèles AMB, ainsi que la forme « espace-état » utilisée dans JD+ pour les modèles structurels, avec une attention plus particulière pour la composante saisonnière. Après avoir rappelé les méthodes d'estimation des modèles UCARIMA, on compare brièvement les deux approches à partir d'exemples.

Les modèles discutés dans le premier paragraphe n'offrent pas toute la flexibilité d'une méthode telle que le X11. Il est toutefois possible de les enrichir pour mieux correspondre à des séries au profil particulier. On présente brièvement dans le second paragraphe quelques solutions, disponibles dans JD+.

Une des raisons d'être de la désaisonnalisation par modélisation est de fournir un cadre cohérent, qui permette de comprendre et d'analyser les propriétés statistiques des estimations. On termine en illustrant ce point par l'analyse des révisions des ajustements saisonniers par modélisation.

## 1. Modèles de référence

Une hypothèse commune à l'ensemble des modèles considérés dans ce document concerne la distribution de leurs innovations : on supposera qu'elles suivent des distributions normales. Cette hypothèse ne sera plus rappelée par la suite.

### 1.1. Modèles UCARIMA

Pour simplifier les dérivations, on considère le modèle de décomposition « signal-bruit » suivant:

$$y_t = s_t + n_t$$

$$\Phi(B)y_t = \Theta(B)\varepsilon_t, \quad \varepsilon_t \sim N(0, v)$$

$$\Phi_s(B)s_t = \Theta_s(B)\gamma_t, \quad \gamma_t \sim N(0, v_s)$$

$$\Phi_n(B)n_t = \Theta_n(B)\delta_t, \quad \delta_t \sim N(0, v_n)$$

avec les hypothèses habituelles:

- $\gamma_t$  et  $\delta_t$  sont indépendants
- $\Phi_s(B)$  et  $\Phi_n(B)$  ne contiennent pas de racines communes.
- $\Theta_s(B)$  et  $\Theta_n(B)$  ne contiennent pas de racines unitaires communes.

Il en résulte que :

$$\Phi_s(B)\Phi_n(B) = \Phi(B)$$

et

$$\Theta(B)\varepsilon_t = \Phi_n\Theta_s(B)\gamma_t + \Phi_s(B)\Theta_n(B)\delta_t$$

Par l'égalisation des auto-covariances, cette dernière relation implique que

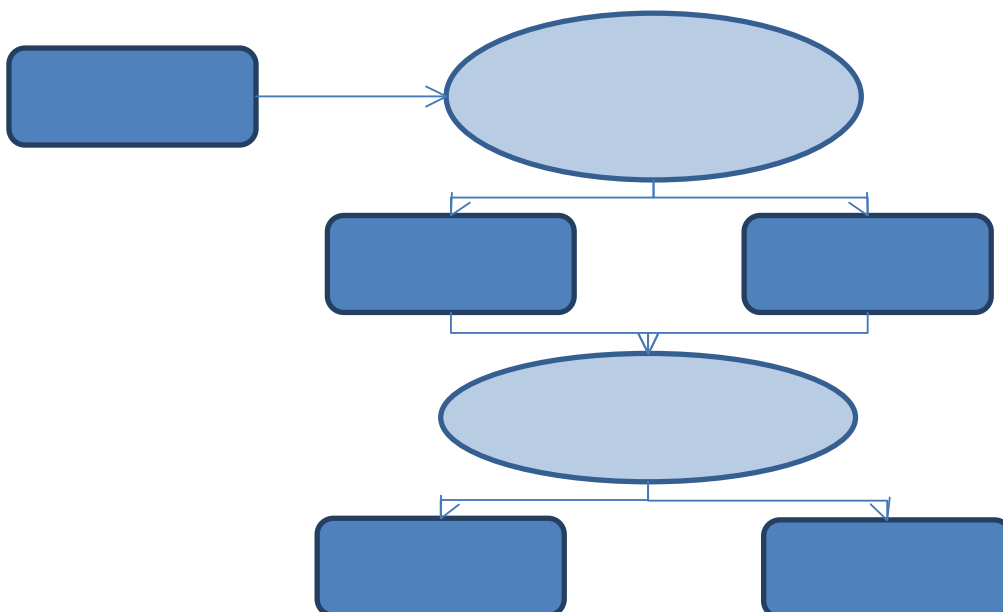
$$v_s\Phi_n(B)\Theta_s(B)\Phi_n(F)\Theta_s(F) + v_n\Phi_s(B)\Theta_n(B)\Phi_s(F)\Theta_n(F) = v\Theta(B)\Theta(F)$$

Les deux dernières contraintes peuvent être lues de deux manières différentes : soit elles expriment la manière dont il faut décomposer le modèle de la série en deux composantes, soit elles expriment la manière d'agréger les modèles du signal et du bruit, qui sont définis de manière préalable. Les paragraphes suivants illustrent ces deux manières d'aborder les modèles de décomposition.

### 1.2. Approche « ARIMA model-based » (AMB) et décomposition canonique

L'approche AMB part d'une modélisation ARIMA de la série agrégée. Le modèle initial est factorisé en répartissant les racines de son polynôme autorégressif entre le signal et le bruit. En imposant une contrainte sur l'ordre du polynôme moyenne mobile du signal (strictement inférieur à l'ordre du polynôme autorégressif), la factorisation est unique. L'annexe B.1 décrit la procédure suivie dans SEATS (et dans JD+) à cette fin.

La décomposition ainsi obtenue est rendue « canonique » en enlevant du signal le maximum de bruit. Ce bruit correspond au minimum du (pseudo-)spectre du modèle ARIMA du signal. Le modèle ainsi obtenu est alors non inversible, le (pseudo-)spectre s'annulant au moins en un point.



Une fois le modèle initial donné, la décomposition canonique est donc définie de manière univoque. Elle vise à obtenir des composantes ayant des caractéristiques a priori raisonnables : capture de certaines

fréquences et minimisation du bruit dans le signal. Elle ne cherche donc pas à trouver un sens dans les modèles de la décomposition mais à produire un modèle correspond à un objectif mathématique.

### 1.3. Modèles structurels et forme « espace-état » (STS)

On présente dans un premier temps la forme « espace-état » générale utilisée dans JD+ et on décrit ensuite le modèle signal-bruit dans cette forme. Les modèles structurels, et plus particulièrement la composante saisonnière sont ensuite discutés.

#### 1.3.1. Forme générale dans JD+(2.0)

Les modèles espace-état linéaires, avec résidus de type gaussien, peuvent s'écrire de différentes manières. Dans JD+, on utilise la formulation suivante :

L'équation de mesure s'écrit

$$y_t = Z_t \alpha_t$$

et l'équation de transition

$$\alpha_{t+1} = T_t \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2 V_t)$$

Les  $y_t$  sont les observations en t,  $\alpha_t$  est le vecteur d'état. Les innovations  $\varepsilon_t$  sont indépendantes en tous points.

Les résidus de l'équation de transition sont modélisés par

$$\varepsilon_t = S_t \xi_t, \quad \xi_t \sim N(0, \sigma^2 Q_t)$$

où  $Q_t$  est une matrice non singulière. En d'autres mots,  $V_t = S_t Q_t S_t'$

Les conditions initiales du filtre prennent la forme habituelle:

$$\alpha_0 = a_0 + B_0 \delta + \mu_0$$

$$\delta \sim N(0, \kappa I)$$

$$\mu_0 \sim N(0, \sigma^2 P_{\cdot 0})$$

avec  $\kappa$  arbitrairement large.

$P_{\cdot 0}$  est la variance de la partie stationnaire du vecteur initial et  $B_0 B_0' = P_{\infty 0}$  modélise la partie diffuse.

Contrairement à l'usage, l'équation de mesure ne contient donc pas de résidu. Il est toujours possible d'obtenir ce type de représentation en intégrant les résidus éventuels de l'équation de mesure dans le vecteur d'état.

### 1.3.2. *Modèle de décomposition*

Pour autant que l'on possède une représentation espace-état des composantes, le modèle « signal-bruit » s'écrit de manière triviale dans la forme de JD+ :

$$y_t = [Z_s \quad Z_n] \begin{bmatrix} \alpha_{st} \\ \alpha_{nt} \end{bmatrix}$$

$$\begin{bmatrix} \alpha_{st+1} \\ \alpha_{nt+1} \end{bmatrix} = \begin{bmatrix} T_s & \mathbf{0} \\ \mathbf{0} & T_n \end{bmatrix} \begin{bmatrix} \alpha_{st} \\ \alpha_{nt} \end{bmatrix} + \begin{bmatrix} \mu_{st} \\ \mu_{nt} \end{bmatrix}$$

$$\begin{bmatrix} \mu_{st} \\ \mu_{nt} \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \Omega_s & \mathbf{0} \\ \mathbf{0} & \Omega_n \end{bmatrix}\right)$$

### 1.3.3. *Modèles structurels*

Les modèles structurels de JD+ correspondent aux modèles décrits par exemple dans Harvey [1989]. Les séries se décomposent en une composante tendancielle, une composante saisonnière et un résidu. La composante tendancielle prend la forme d'un « Local Level » ou d'un « Local Linear Trend » ; suivant la terminologie de Proietti [2000], la composante saisonnière suit un modèle de type « Trigonometric », « Harrison-Stevens », « Crude » ou « Dummy » ; on revient plus en détail sur sa définition dans le point 1.3.4; le résidu est un simple bruit blanc. La forme espace-état de ces différentes composantes est reprise dans l'annexe A.3.

### 1.3.4. *Composante saisonnière*

La modélisation dans une forme structurelle de la composante saisonnière peut prendre différentes formes. On présente ci-dessous le modèle d'Harrison-Stevens, qui est une généralisation naturelle d'un modèle de régression avec dummies saisonnières et la représentation plus générale de West-Harrison, utilisée dans JD+.

#### 1.3.4.1. *Modèle Harrison-Stevens*

Une manière très simplifiée de modéliser l'effet saisonnier repose sur l'utilisation de dummies saisonnières, prenant la valeur 1 pour une saison (mois, trimestre...) spécifique et 0 ailleurs.

$$y_t = X_t \delta, \quad X_{it} = \begin{cases} 1, & t \bmod s = i \\ 0, & t \bmod s \neq i \end{cases}$$

Une telle spécification est en pratique trop rigide, et on peut introduire plus de souplesse en considérant que les coefficients de chaque saison suivent par exemple un randomwalk

$$y_t = X_t \delta_t$$

$$\delta_t = \delta_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Omega), \quad \delta_1 = \delta_0 + \varepsilon_1, \quad \varepsilon_1 \sim N(0, \Omega)$$

Ce système peut facilement se transformer en:

$$\tilde{y}_t = y_t - \delta_t = \sum_{i=0}^{t-1} (X_{it} - X_{i,t-1}) (\delta_{it} - \delta_t) = \tilde{X}_t \tilde{\delta}_t$$

Par définition,

$$\sum_{i=0}^{t-1} \tilde{\delta}_{it} = 0$$

et

$$\tilde{\delta}_{t+1} = \tilde{\delta}_t + \mu_t, \quad \mu_t \sim N(0, D\Omega D') = N(0, \tilde{\Omega}), \quad D = I_s - \frac{\mathbf{1}_s \mathbf{1}_s'}{s}$$

On notera que cette approche est strictement identique à celle que l'on suit pour la définition de l'effet des jours ouvrables (avec coefficients variant dans le temps).

Les modèles traditionnels présuppose que  $\tilde{\delta}_t = 0$  (ou encore que  $\mathbf{1}_s \Omega = 0$ ). Des modèles plus généraux sont discutés dans Proietti [2004]. Ces derniers ne sont actuellement pas disponibles dans JD+. Les modèles Harrison-Stevens sans hétéroscédasticité saisonnière peuvent s'écrire selon la forme reprise dans le paragraphe suivant. Les modèles avec hétéroscédasticité saisonnière sont brièvement discutés dans le point relatif aux extensions des modèles traditionnels. Ils sont tous deux disponibles dans les algorithmes de base de JD+ (version 2.1.0, en développement), sans interface graphique.

#### 1.3.4.2. Représentation de type West-Harrison

Les modèles structurels présentés dans JD+ utilisent, pour leur composante saisonnière, la forme espace-état de West-Harrison.

$$\alpha_{\gamma, \tau+1} = \begin{bmatrix} -1 & -1 & \dots & -1 & -1 \\ 1 & 0 & \dots & \dots & 0 \\ 0 & \backslash & \backslash & \backslash & \vdots \\ \vdots & \backslash & 1 & 0 & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix} \alpha_{\gamma, \tau} + \mu_{\gamma, \tau}, \quad \mu_{\gamma, \tau} \sim N(0, \Omega_{\tau})$$

$$\gamma_{\tau} = [1 \ 0 \ \dots \ 0] \alpha_{\gamma, \tau}$$

La matrice  $\Omega_{\tau}$  peut prendre des formes différentes, suivant le modèle considéré. JD+ fournit une implémentation pour les modèles repris dans le tableau ci-dessous (voir Proietti[2000]).

Modèle	Variance
Trigonometric (s pair)	$\begin{pmatrix} s/2 & -1 & 0 & \dots & -1 & 0 \\ -1 & s/2 & -1 & \backslash & \vdots & \vdots \\ 0 & -1 & s/2 & \backslash & 0 & -1 \\ -1 & 0 & -1 & \backslash & -1 & 0 \\ \vdots & \vdots & \vdots & \backslash & s/2 & -1 \\ 0 & -1 & 0 & \dots & -1 & s/2 \end{pmatrix}$
Crude	$\frac{1}{s} \begin{pmatrix} 1 & \dots & \dots & 1 \\ \vdots & \backslash & \dots & \vdots \\ \vdots & \vdots & \backslash & \vdots \\ 1 & \dots & \dots & 1 \end{pmatrix}$
Harrison-Stevens	$\frac{1}{s} \begin{pmatrix} s-1 & -1 & \dots & -1 \\ -1 & s-1 & \dots & \vdots \\ \vdots & \vdots & \backslash & -1 \\ -1 & \dots & -1 & s-1 \end{pmatrix}$
Dummy	$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & \vdots \\ \vdots & \vdots & \backslash & \vdots \\ 0 & \dots & \dots & 0 \end{pmatrix}$

On notera encore que ces modèles impliquent que la somme des composantes saisonnières sur une année suit un modèle « moyenne mobile »

$$S(B)\gamma_{\tau} = MA(q), \quad q \leq s - 2$$

On reprend ci-dessous la partie MA des modèles saisonniers pour le cas mensuel. A noter que les polynômes MA des modèles structurels sont indépendants de la série.



$$\theta_{cri} = 1 + .74B + .63B^2 + .43B^3 + .36B^4 + .22B^5 + .18B^6 + .09B^7 + .07B^8 + 0.02B^9 + 0.02B^{10}$$

$$\theta_{hs} = 1 + .79B + .62B^2 + .47B^3 + .35B^4 + .25B^5 + .17B^6 + .10B^7 + .06B^8 + 0.03B^9 + 0.01B^{10}$$

$$\theta_{crude} = 1 + .91B + .82B^2 + .73B^3 + .64B^4 + .55B^5 + .45B^6 + .36B^7 + .27B^8 + 0.18B^9 + 0.09B^{10}$$

$$\theta_{dummy} = 1$$

Ils peuvent donc tous être mis dans une forme ARIMA et le modèle complet de la série est bien du type UCARIMA.

#### 1.4. Estimation (MMSE)

L'estimation des modèles UCARIMA peut se faire soit via les filtres de Wiener-Kolmogorov (WK), soit via les filtres de Kalman (KF). L'approche WK repose sur la représentation en modèles ARIMA; elle consiste en effet essentiellement en des dérivations simples à partir des polynômes qui définissent le modèle et ses composantes. Les filtres KF sont pour leur part étroitement liés à une représentation en modèle espace-état. Si les modèles résultant d'une décomposition canonique s'écrivent naturellement dans la forme ARIMA, ils peuvent se mettre aisément en modèles espace-état. On se référera à l'annexe A.1, A.2 pour la représentation espace-état d'un modèle AR(I)MA utilisée dans JD+. Les modèles structurels sont pour leur part d'abord exprimés dans une forme espace-état. Comme discuté plus haut, les formes réduites des composantes de ces modèles dans leur définition classique peuvent également s'écrire sous une forme ARIMA.

Modèles AMB et modèles STS sont donc deux formes de modèles UCARIMA et peuvent ainsi être estimés et analysés via les filtres WK ou KF.

##### 1.4.1. Approche de Wiener-Kolmogorov (SEATS)

Les filtres de WK fournissent les estimateurs MMSE du signal et du bruit:

$$\hat{s}_t = \frac{v_s \Phi_n(B) \Theta_s(B) \Phi_n(F) \Theta_s(F)}{v \Theta(B) \Theta(F)} y_t$$

$$\hat{n}_t = \frac{v_n \Phi_s(B) \Theta_n(B) \Phi_s(F) \Theta_n(F)}{v \Theta(B) \Theta(F)} y_t$$

Ces formules sont également valides pour les modèles non stationnaires

On peut également les exprimer en fonction des innovations du modèle agrégé :

$$\hat{s}_t = \frac{k_s \Theta_s(B) \Phi_n(F) \Theta_s(F)}{\Phi_s(B) \Theta(F)} \varepsilon_t$$

$$\hat{n}_t = \frac{k_n \Theta_n(B) \Phi_s(F) \Theta_n(F)}{\Phi_n(B) \Theta(F)} \varepsilon_t$$

$$k_s = v_s / v, k_n = v_n / v$$

Après factorisation (polynômes en B et polynômes en F), les estimateurs des différentes composantes peuvent ainsi s'écrire comme une moyenne mobile infinie des innovations.

$$\hat{s}_t = \sum_{i=0}^{\infty} \xi_i^- \varepsilon_{t-i} + \sum_{i=1}^{\infty} \xi_i^+ \varepsilon_{t+i}$$

Cette représentation - appelée «PsiE-weights» dans SEATS – est particulièrement adaptée à l'analyse des révisions. On y reviendra plus loin.

Ces différentes formules permettent de dériver de manière simple de nombreuses propriétés des modèles UCARIMA. On se référera aux multiples articles de Gomez-Maravall pour plus de détails sur le sujet. La grille d'analyse développée dans SEATS en illustre les principaux. Elle a été largement reprise dans JD+, à la fois pour les modèles AMB et STS.

L'estimation proprement dite des composantes pour une série réelle (finie) repose sur l'algorithme de Burman, qui est une application simple des filtres sur la série prolongée au moyen d'un nombre limité de « forecasts/backcasts ». Contrairement à l'analyse des propriétés du modèle, qui repose sur une représentation infinie de la série, les estimations fournies par l'algorithme de Burman sont « exactes » (projection sur l'espace généré par les observations réelles).

### 1.4.2. Estimation par les filtres de Kalman

Les filtres/lisseurs de Kalman fournissent l'espérance et la variance des vecteurs d'états, et par conséquent, celles des composantes.

Pour autant que le filtre soit correctement initialisé – par exemple en utilisant la procédure d'initialisation diffuse de Koopman pour les séries non stationnaires – les estimations des composantes par KF sont identiques à celles par WK. JD+ s'inspire largement de Durbin-Koopman [2001], que l'on pourra consulter pour plus de détail.

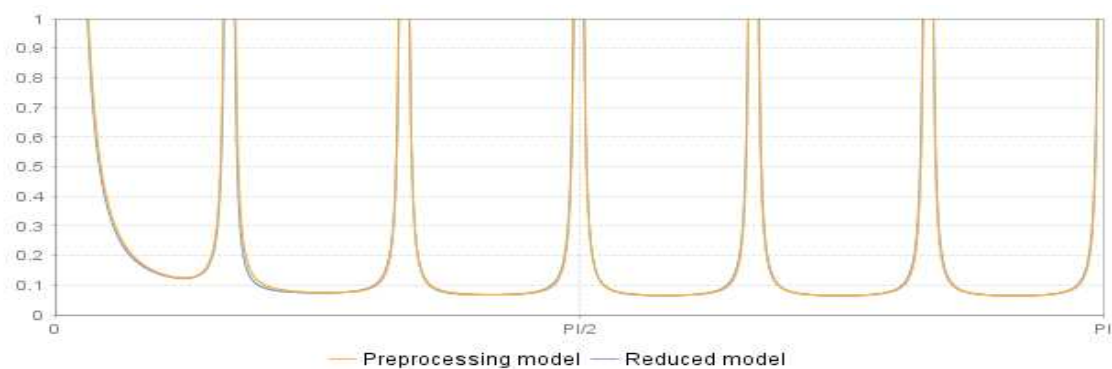
La dérivation des propriétés des estimateurs est généralement plus complexe que dans le cas des filtres WK. L'analyse des révisions discutée plus loin en est un exemple. Les résultats obtenus sont en revanche conditionnels aux observations effectives et donc « exacts ».

## 1.5. Comparaison des approches AMB et STS

### 1.5.1. Comparaison d'un modèle STS et d'un modèle airline

Les estimations d'une série par un modèle structurel tel que défini plus haut et par un modèle airline sont généralement très proches. La forme réduite du modèle structurel diffère en effet souvent de manière marginale de l'airline et les deux modèles ont des caractéristiques extrêmement similaires. Ce phénomène est illustré par l'exemple suivant.

Série INSEE 001654507



Preprocessing model (0,1,1)(0,1,1)

#### Polynomials

regular MA: 1.00000 - 0.709555 B  
seasonal MA: 1.00000 - 0.487508 S

#### Reduced model

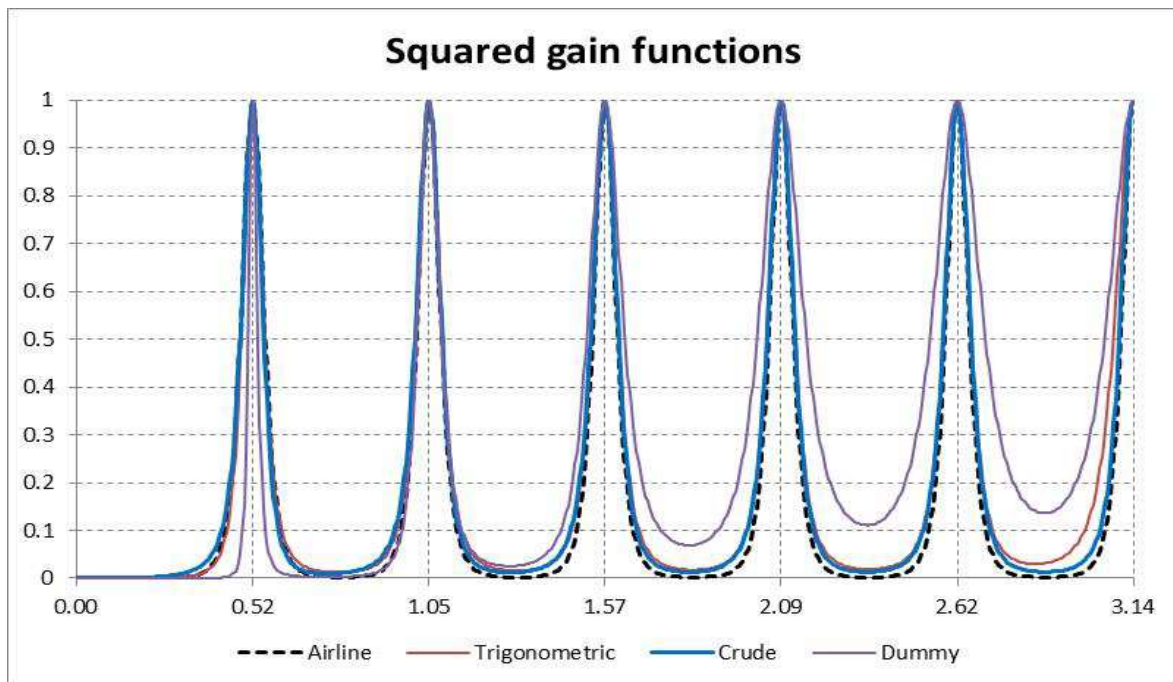
D: 1.00000 - B - B<sup>12</sup> + B<sup>13</sup>  
MA: 1.00000 - 0.731773 B + 0.00617887 B<sup>2</sup> + 0.00561459 B<sup>3</sup> + 0.00482616 B<sup>4</sup> + 0.00377561 B<sup>5</sup> + 0.00240978 B<sup>6</sup> + 0.000656604 B<sup>7</sup> - 0.00157990 B<sup>8</sup> - 0.00442638 B<sup>9</sup> - 0.00804923 B<sup>10</sup> - 0.0126666 B<sup>11</sup> - 0.481117 B<sup>12</sup> + 0.345014 B<sup>13</sup>

Des différences plus marquées apparaîtront généralement lorsque l'estimation du modèle structurel conduit à un modèle dégénéré, l'une des variances des composantes tombant à 0. On notera qu'une telle contrainte conduit à un modèle réduit ayant une racine unitaire dans sa partie MA, ce qui le rend alors impropre à une analyse au moyen des filtres WK.

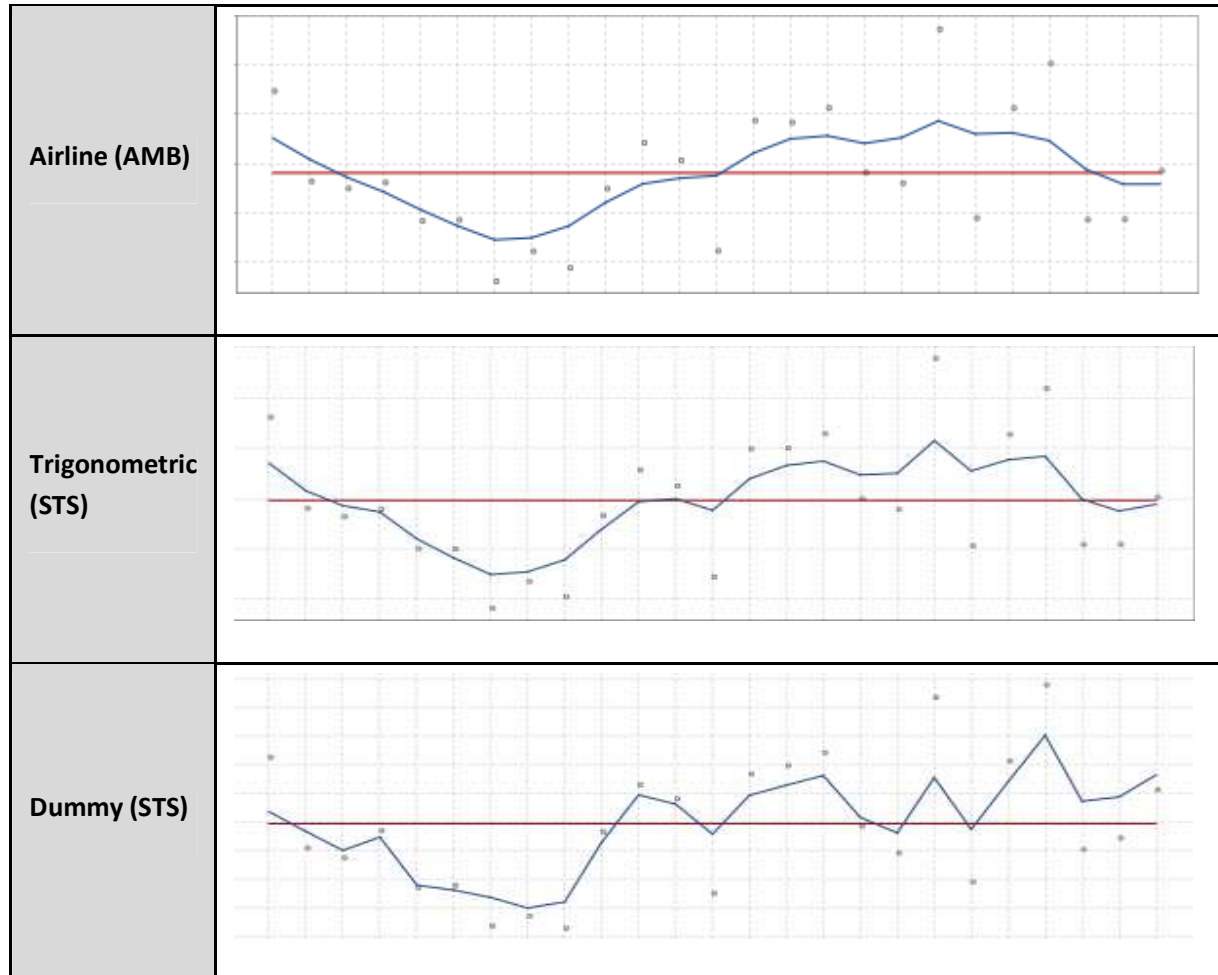
### 1.5.2. Comparaison des composantes saisonnières

La comparaison des composantes saisonnières met en évidence la notion de décomposition canonique : elle conduit, dans l'approche AMB, à enlever un maximum de bruit des composantes tendancielle et saisonnière. Dans les modèles structurels, en revanche, les composantes saisonnières (et tendancielle) ne sont pas exemptes de bruit. Les deux graphiques suivant illustrent le phénomène. On notera les relations suivantes en termes de bruit dans la composante saisonnière:

AMB < Crude < Harrison-Stevens  $\approx$  Trigonometric < Dummy.



Série INSEE 001654507, facteurs saisonniers pour le mois de septembre



### 1.5.3. Conclusion

Les modèles airline et structurels conduisent à des modélisations souvent proches et donc à des prévisions, des coefficients de régression... qui ne diffèrent que marginalement.

En revanche, des écarts plus marqués apparaissent au niveau des composantes saisonnières. La décomposition canonique d'un modèle airline génère ainsi une tendance et une composante saisonnière plus stables que celles estimées à partir d'un modèle structurel (et automatiquement l'inverse pour la composante irrégulière et la série désaisonnalisée).

## 2. Modèles alternatifs

Les modèles classiques considérés jusqu'à présent ne permettent pas de prendre en compte certaines caractéristiques moins communes des séries chronologiques. On explore ci-dessous quelques-uns de ces problèmes et on propose des modèles permettant de les traiter, d'une manière parfois assez similaire à celle proposée dans des méthodes non paramétriques telles que le X11,

### 2.1. Hétéroscédasticité des séries en fonction de la saison et modèle « Mixed airline »

Dans certaines séries, on peut observer de plus grandes irrégularités en fonction des saisons. On peut modéliser ce type de phénomène en augmentant un modèle donné avec un bruit blanc, pour les saisons à problème.

La forme espace-état du nouveau modèle est décrite ci-dessous.

$$y_t = [Z_{arima} \quad Z_{nt}] \begin{bmatrix} \alpha_{arima,t} \\ \alpha_{nt} \end{bmatrix}$$

$$\begin{bmatrix} \alpha_{arima,t+1} \\ \alpha_{nt,t+1} \end{bmatrix} = \begin{bmatrix} T_{arima} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \alpha_{arima,t} \\ \alpha_{nt} \end{bmatrix} + \begin{bmatrix} \mu_{arima,t} \\ \varepsilon_t \end{bmatrix}$$

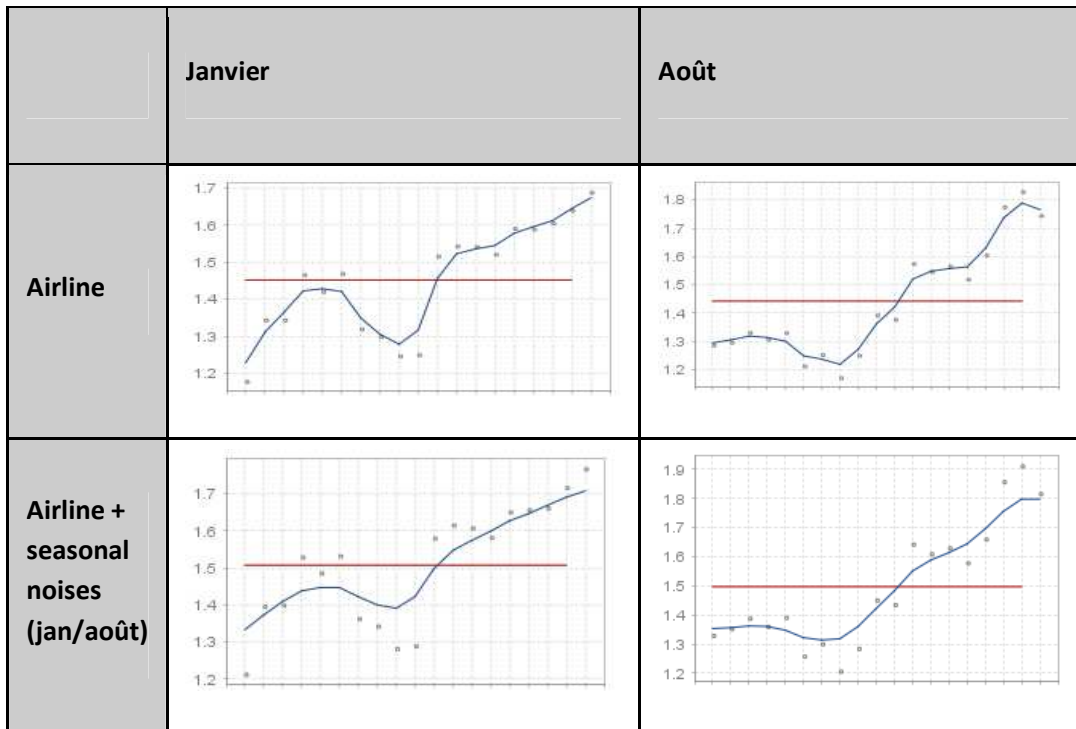
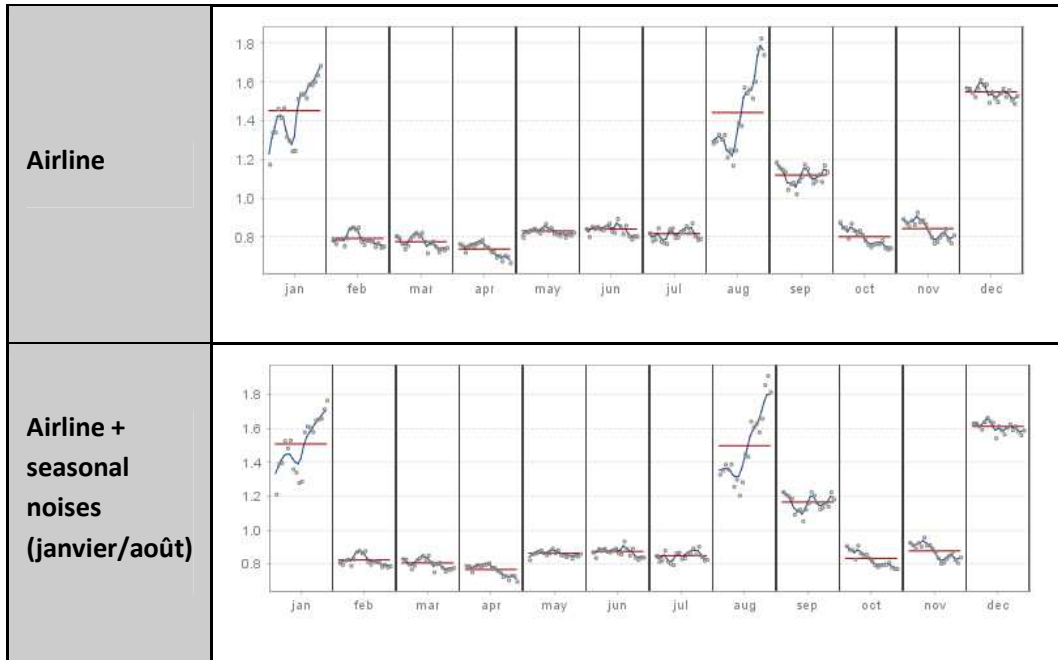
$$\begin{bmatrix} \mu_{arima,t} \\ \varepsilon_t \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \Omega_{arima} & \mathbf{0} \\ \mathbf{0} & v_n \end{bmatrix}\right)$$

$$Z_{nt} = 1 \text{ si } t \bmod s \in S, 0 \text{ sinon}$$

(S est l'ensemble des saisons « perturbées »)

JD+ fournit ce type de modèle en partant d'un airline. Il est appelé « mixed airline ». Voir Trimbur[2008]

Son impact est présenté dans l'exemple suivant :



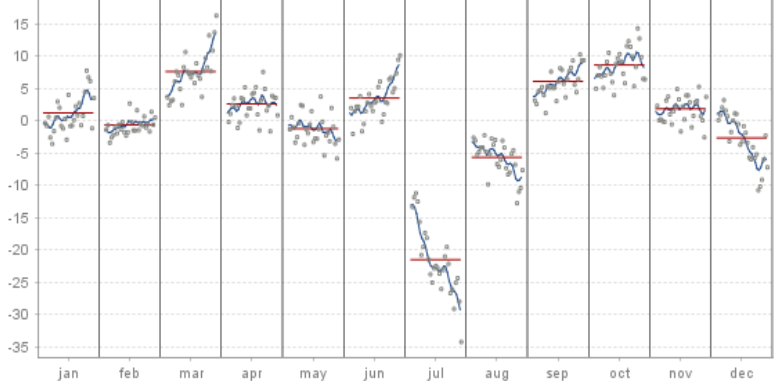
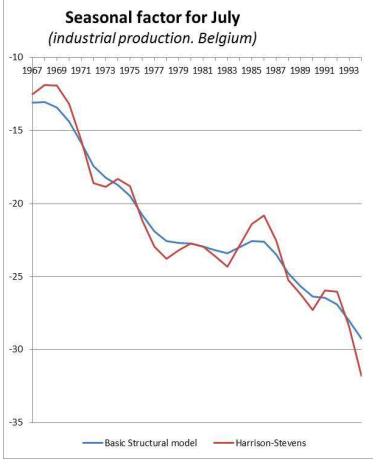
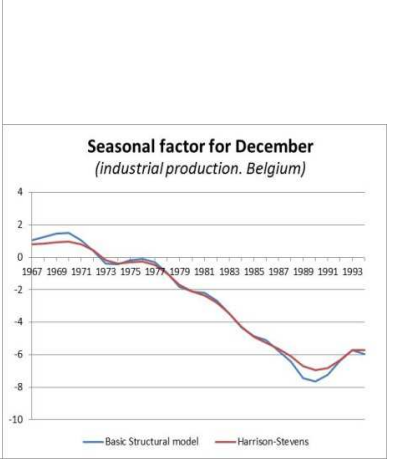
## 2.2. Modèle Harrison-Stevens avec variances saisonnières hétérogènes

Les facteurs saisonniers peuvent évoluer de manière très différente en fonction des saisons. Les modèles classiques sont homogènes dans le temps et ne peuvent pas prendre en compte cette disparité. Le modèle Harrison-Stevens présenté au point 2.3.4.1 peut traiter ce phénomène à partir du moment où on attribue aux innovations de chaque saison une variance spécifique. Dans la pratique, on utilisera un nombre réduit de variances différentes : par exemple

$$\gamma_t = X_t \delta_t, \quad X_t \text{ dummies saisonnières,}$$

$$\delta_{i,t+1} = \delta_{i,t} + \varepsilon_{i,t}, \quad \begin{cases} \varepsilon_{i,t} \sim N(0, v_0), & i \in S \\ \varepsilon_{i,t} \sim N(0, v_1), & i \notin S \end{cases}$$

Belgique, production industrielle.

Modèle	Estimation	Composantes saisonnière
STS (Harrison-Stevens)	$v_0 = 0.0214$ Loglikelihood = -856.66	
STS (Harrison-Stevens) avec 2 variances (juillet et autres mois)	$v_0 = 0.0105$ , $v_1 = 0.1094$ , $S = \{\text{July}\}$ Loglikelihood = -850.76	<div style="display: flex; justify-content: space-around;"> <div data-bbox="630 1344 1005 1803"> <p><b>Seasonal factor for July</b> (industrial production, Belgium)</p>  </div> <div data-bbox="1013 1344 1412 1803"> <p><b>Seasonal factor for December</b> (industrial production, Belgium)</p>  </div> </div>



### 2.3. Fréquences spécifiques et modèle airline généralisé

Dans certaines séries, on observe des phénomènes récurrents à certaines fréquences, qui ne sont pas nécessairement annuelles. Un exemple typique est celui des séries liées à l'habillement, qui ont des cycles de 6 mois, correspondant au renouvellement des collections.

La partie saisonnière du polynôme moyenne mobile d'un modèle airline peut s'écrire, dans le domaine des fréquences (cas mensuel):

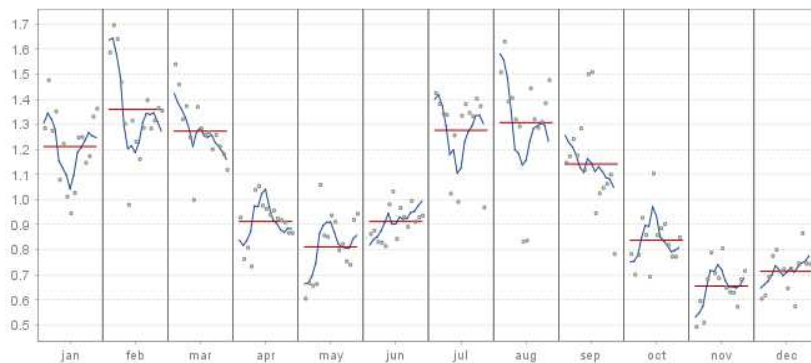
$$1 - \Theta B^{12} = (1 - \Theta^{1/12} B)(1 + \Theta^{1/12} B) \prod_{i=1}^5 \left(1 - 2\Theta^{1/12} \cos \frac{2\pi i}{12} B + \Theta^{2/12} B^2\right)$$

On peut capturer les mouvements relatifs à certaines fréquences en introduisant davantage de flexibilité dans ce polynôme :

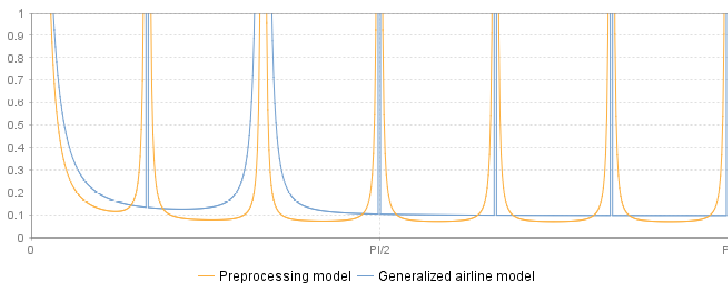
$$(1 - c_0 B)(1 + c_6 B) \prod_{i=1}^5 \left(1 - 2c_i \cos \frac{2\pi i}{12} B + c_i B^2\right)$$

Dans la pratique, on se limitera également à un nombre limité de paramètres. Le modèle ainsi généré est appelé dans JD+ modèle airline généralisé. Voir Aston...[2008]

Le graphique suivant illustre le problème dans le cas des importations belges de chaussures.



Les propriétés du modèle qui permet d'appréhender ce type de série apparaissent clairement dans son spectre, qui est comparé à celui du modèle airline initial.



## 2.4. Conclusion

L'approche par modélisation ne se limite pas aux modèles classiques. Il est ainsi possible de « mimer » en grande partie certaines solutions attractives fournies par des méthodes non-paramétriques comme le X11. JD+ fournit, essentiellement à titre d'exemples, un certain nombre de ces modèles.

Il doit toutefois être clair qu'ils ne se prêtent guère à une utilisation en masse et ne se révèlent vraiment utiles que dans un nombre limité de situations.

## 3. Analyse des révisions

La révision des séries désaisonnalisées peut dépendre de nombreux paramètres : révision des données de base, révision de la modélisation et des paramètres du modèle, révision inhérente à l'utilisation des filtres saisonniers proprement-dits. On s'intéresse dans ce point au dernier aspect, qui est généralement à l'origine de la plus grande partie des révisions. L'approche par modélisation en permet une explication simple de ce type de révision. Lorsqu'ils sont applicables, les filtres WK en fournissent une solution quasi-immédiate. Les filtres KF permettent également de dériver des résultats équivalents, applicables dans l'ensemble des cas.

### 3.1. Approche par les filtres WK

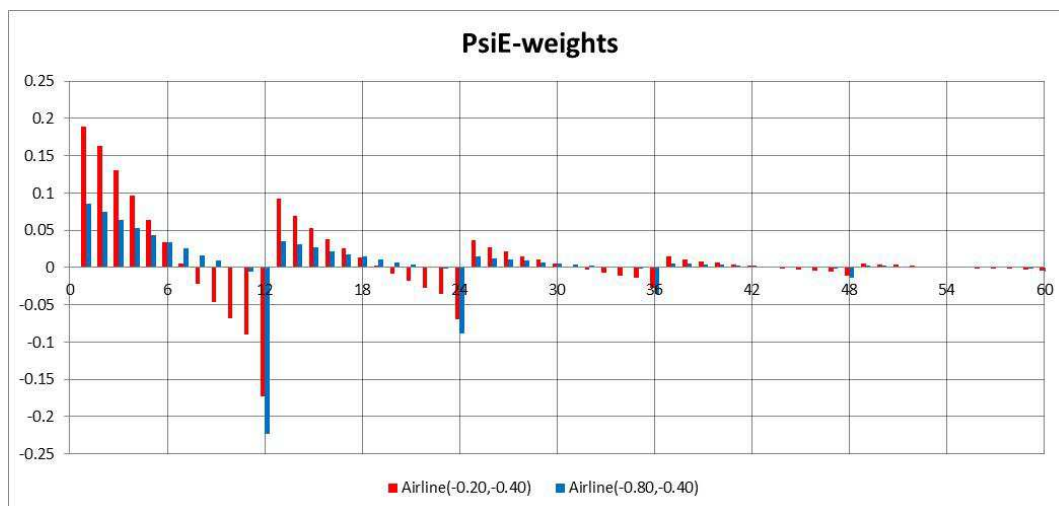
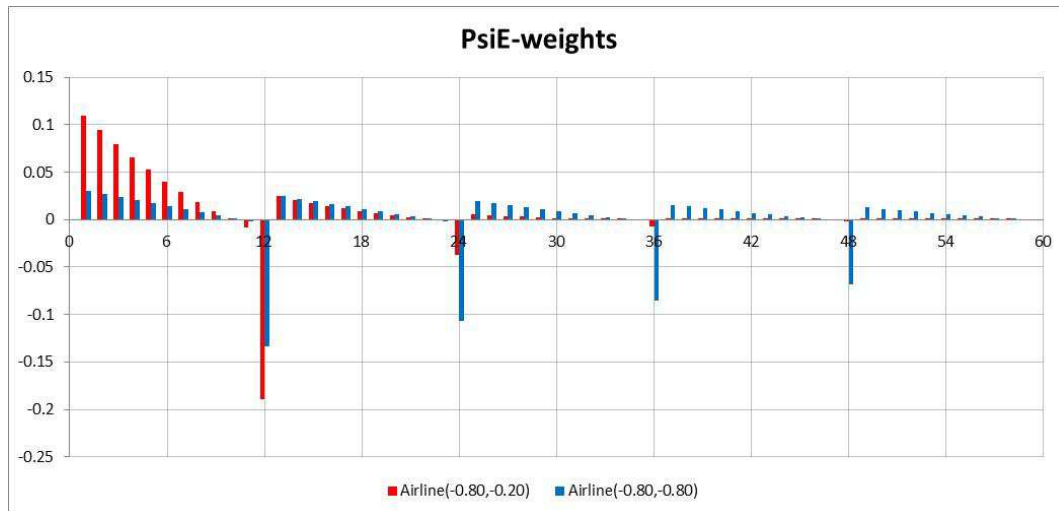
En utilisant la représentation suivant les PsiE-weights de la série, on obtient une expression immédiate des révisions de chaque composante du modèle. Plus précisément,

$$\hat{r}_{(t|t+k)} = \sum_{i=1}^k \xi_i^+ \varepsilon_{t+i}$$

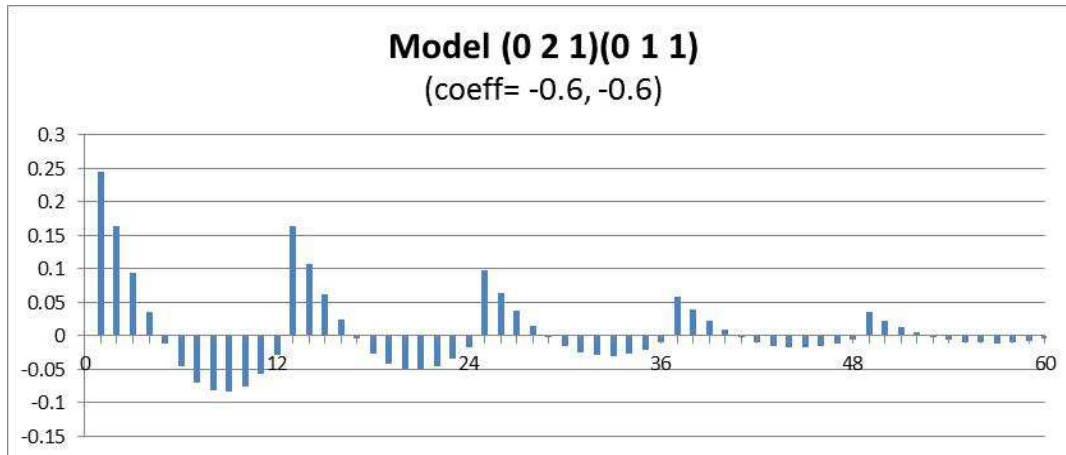
En d'autres termes, chaque innovation du modèle (différence entre la valeur observée et sa prévision une période à l'avance) se répercute sur l'estimation d'une composante avec un poids égal au PsiE-weight correspondant. Les erreurs de prévisions sont fournies directement par les filtres de Kalman.

On présente ci-dessous les poids pour quelques modèles :

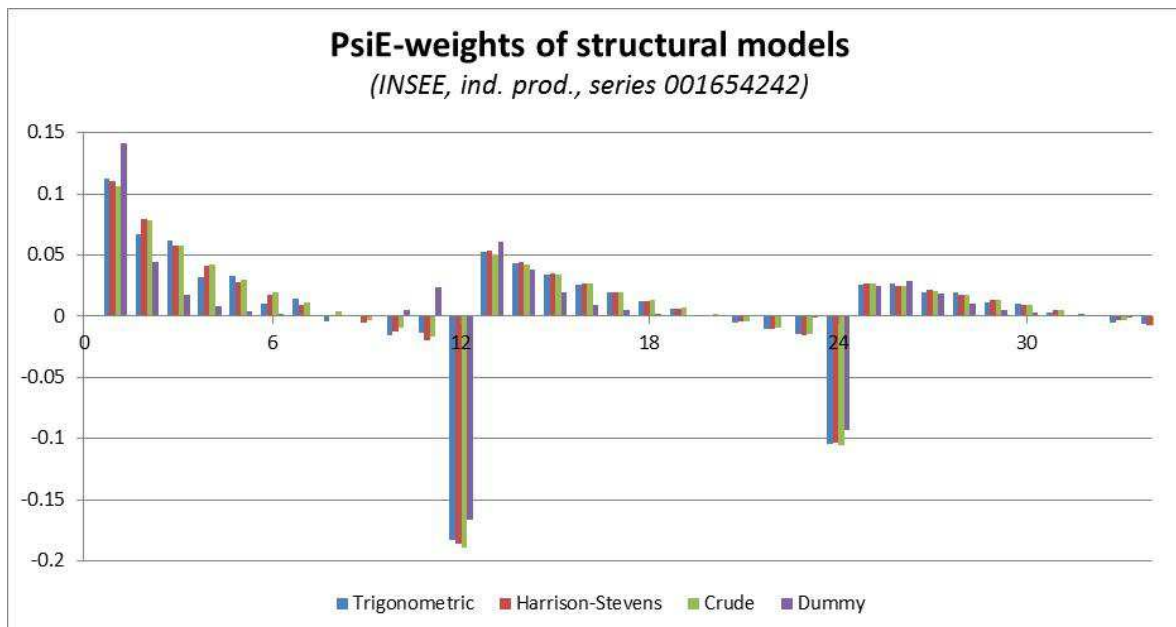
### Modèles Airline



Autre modèle ARIMA :



Modèles structurels :



On notera en particulier que, dans le cas d'une modélisation airline ou dans le cas de modèles structurels, les révisions ont principalement lieu soit juste après la première observation (corrélation positive avec les erreurs de prévisions), soit après 1 an, 2 ans... (corrélation négative avec les erreurs de prévisions). Ainsi qu'on le voit pour le modèle ARIMA (0 2 1)(0 1 1), ce profil, fortement lié à l'ordre des différences du modèle, n'est toutefois pas généralisable.

### 3.2. Approche par les filtres KF

L'approche par les filtres WK n'est pas toujours possible. C'est en particulier le cas des modèles variant dans le temps. Même dans le cas de modèles invariants, on peut s'interroger sur la marge d'erreur liée à l'approche semi-infinie des filtres WK.

Il est aisé de développer une formulation exacte des erreurs de prévisions, comparable à celle basée sur les PsiE-weights. Plus précisément, en utilisant un lissage à point fixe, on a :

$$r(t|t+k) = \sum_{i=1}^k \gamma_i \varepsilon_i$$

En effet, partant du modèle espace-état

$$y_t = [Z_s \quad Z_n] \begin{bmatrix} \alpha_{st} \\ \alpha_{nt} \end{bmatrix}$$

$$\begin{bmatrix} \alpha_{st+1} \\ \alpha_{nt+1} \end{bmatrix} = \begin{bmatrix} T_s & 0 \\ 0 & T_n \end{bmatrix} \begin{bmatrix} \alpha_{st} \\ \alpha_{nt} \end{bmatrix} + \begin{bmatrix} \mu_{st} \\ \mu_{nt} \end{bmatrix}$$

On considère le modèle augmenté

$$y_{t+k} = [Z_s \quad Z_n \quad 0] \begin{bmatrix} \alpha_{st+k} \\ \alpha_{nt+k} \\ \alpha_{st} \end{bmatrix}$$

$$\alpha_{t+k+1} = \begin{bmatrix} \alpha_{st+k+1} \\ \alpha_{nt+k+1} \\ \alpha_{st} \end{bmatrix} = \begin{bmatrix} T_s & 0 & 0 \\ 0 & T_n & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} \alpha_{st+k} \\ \alpha_{nt+k} \\ \alpha_{st} \end{bmatrix} + \begin{bmatrix} \mu_{st+k} \\ \mu_{nt+k} \\ 0 \end{bmatrix}$$

Le filtre de Kalman calcule de manière récursive  $E(\alpha)_{t+k+1|t+k}$  et sa matrice de covariance. Si on considère uniquement la partie  $\alpha_{st}$  du vecteur d'état, on voit facilement qu'à chaque étape, on a :

$$E(\alpha)_{st|t+k+1} = E(\alpha)_{st|t+k} + [0 \quad 0 \quad I] K_{t+k} \varepsilon_{t+k}$$

Ou encore, en appliquant la matrice  $Z_s$

$$E(s_t|t+k+1) - E(s_t|t+k) = \gamma_{t+k} \varepsilon_{t+k}$$

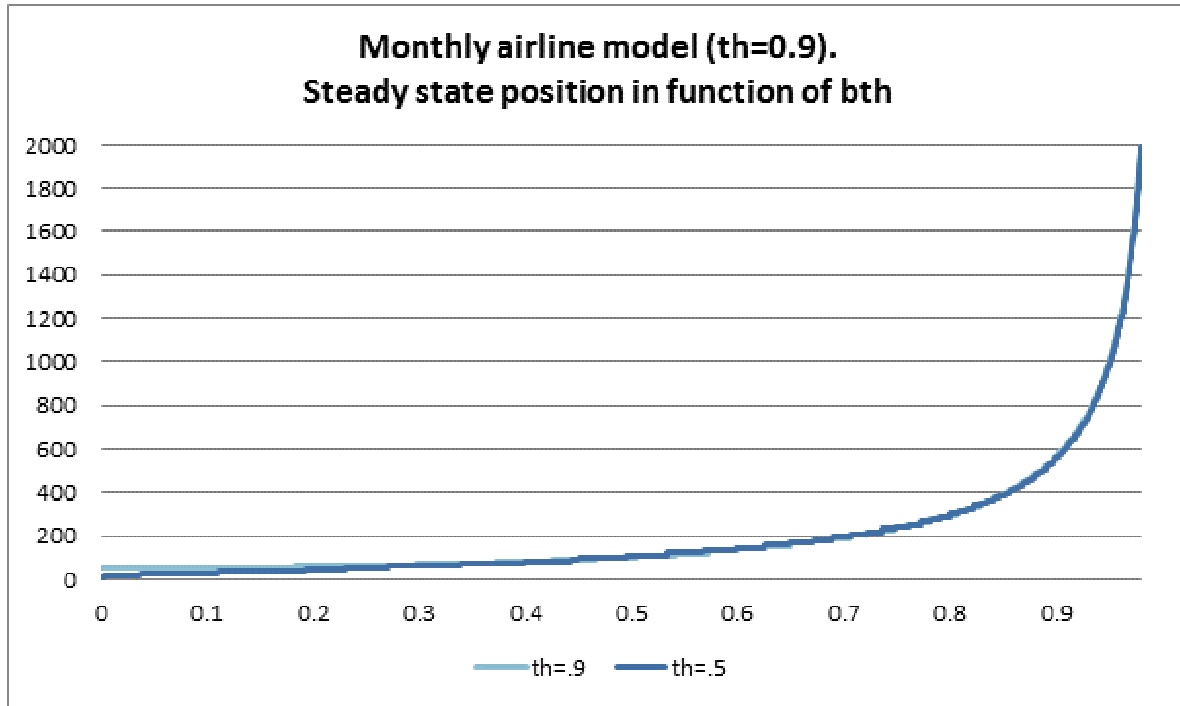
$\varepsilon_{t+k}$  est l'erreur de prévision au temps  $t+k$  et  $\gamma_{t+k} = [0 \quad 0 \quad Z_s] K_{t+k}$ ,  $K_{t+k}$  étant le gain du filtre de Kalman du modèle augmenté.

Enfin, il est également évident que

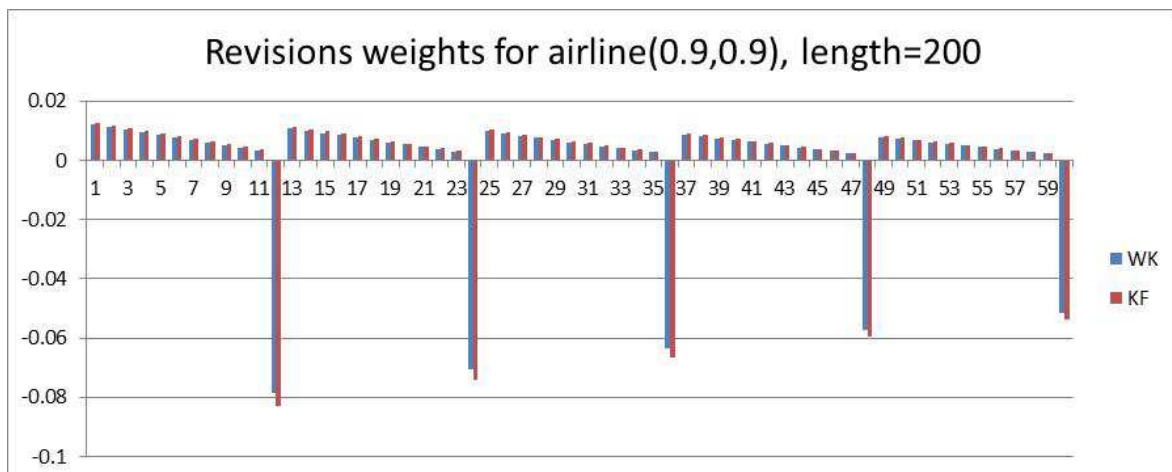
$$\forall i > 0 : \xi_i^+ \approx \gamma_{t+i}$$

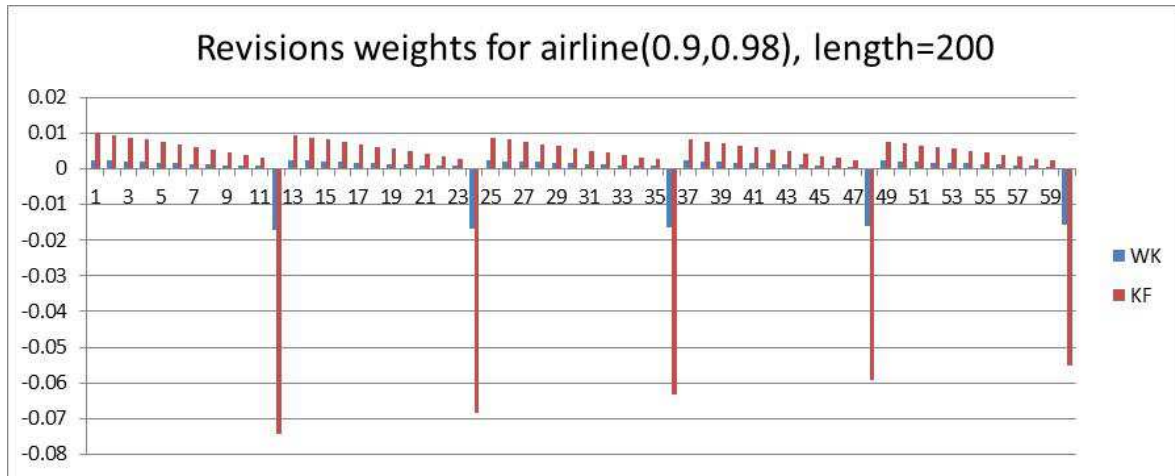
implique que le modèle (nécessairement invariant dans le temps) soit en état de (quasi-) équilibre à partir de la période  $t$ .

On montre dans le graphique suivant, dans le cas d'un modèle airline mensuel, la longueur minimum des séries pour parvenir à un modèle en (quasi-)équilibre, et donc tel que ses propriétés dérivées avec les filtres WK et les filtres KF – y compris les poids des révisions – soient équivalentes.



Les deux diagrammes suivants illustrent pour leur part les différences dans le poids des révisions, suivant les approches théoriques et exactes. On notera en particulier – comme attendu – la forte sous-estimation de ces poids dans l'approche par les filtres WK lorsque le modèle est quasi non-inversible.



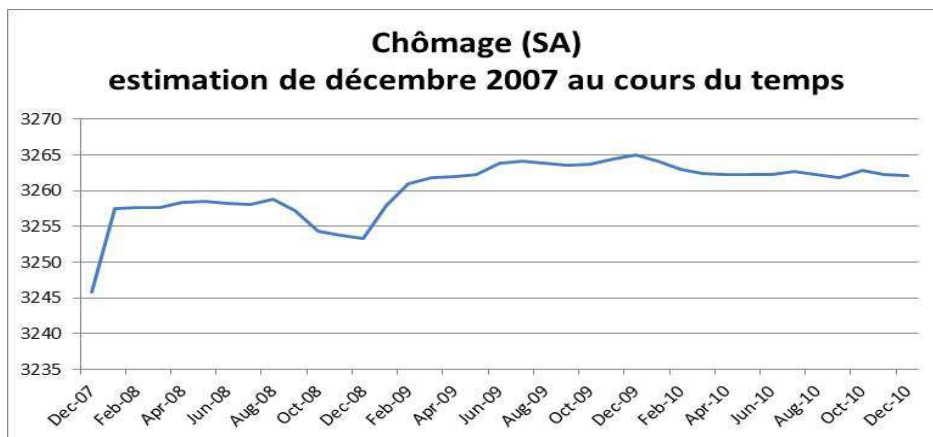


### 3.3. Quelques exemples réels

On présente ci-dessous l'application des formules de révision à deux séries réelles. Dans les deux cas, la modélisation s'est effectuée à partir de Tramo-Seats. Les modèles et leurs paramètres ont été fixés, si bien que les révisions ne prennent en compte que les effets liés aux erreurs de prévisions et à la forme des filtres sous-jacents aux modèles sélectionnés. Les modèles étant en équilibre, les filtres WK et KF fournissent les mêmes poids de révision. Pour chacun des exemples, on fournit un graphique reprenant l'évolution de l'estimation de la série désaisonnalisée au cours du temps (et donc ses révisions cumulées) et un tableau décomposant chacune des révisions suivant la formule liée aux PsiE-weights.

#### 3.3.1. Chômage (réf. INSEE :001572440) ; estimation du mois de décembre 2007.

La série du chômage, comme de nombreuses séries du marché du travail, est modélisée par un modèle ARIMA (0 2 1)(0 1 1).



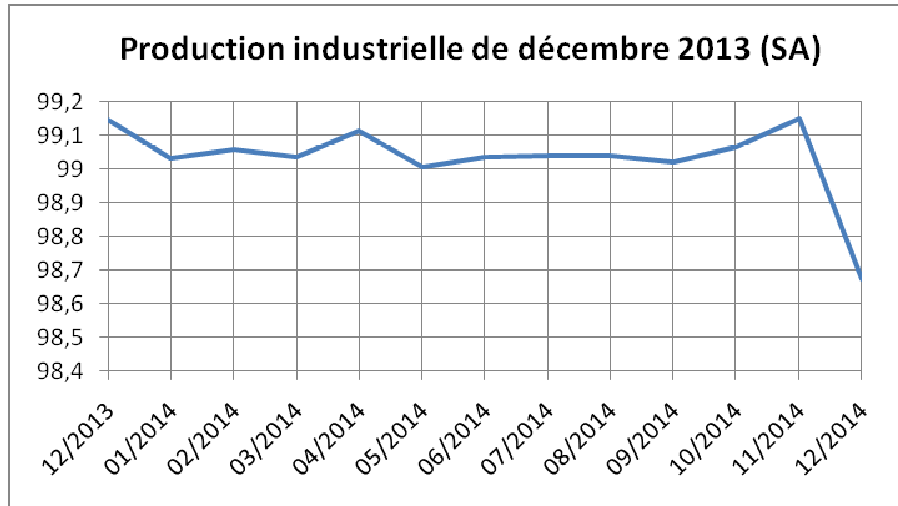
Période	Innovations	PsiE-weight	Révisions	SA (décembre 2007)
---------	-------------	-------------	-----------	--------------------

Dec-07				3245.8
Jan-08	61.8	0.188	11.6	3257.4
Feb-08	2.0	0.125	0.2	3257.7
Mar-08	-1.7	0.071	-0.1	3257.6
Apr-08	31.3	0.026	0.8	3258.4
May-08	-15.6	-0.011	0.2	3258.5
Jun-08	8.9	-0.038	-0.3	3258.2
Jul-08	3.3	-0.056	-0.2	3258.0
Aug-08	-12.6	-0.066	0.8	3258.8
Sep-08	25.5	-0.067	-1.7	3257.1
Oct-08	48.3	-0.059	-2.9	3254.3
Nov-08	10.5	-0.044	-0.5	3253.8
Dec-08	27.8	-0.020	-0.6	3253.2
Jan-09	33.7	0.141	4.7	3258.0
Feb-09	31.4	0.093	2.9	3260.9
Mar-09	17.8	0.053	0.9	3261.8
Apr-09	9.3	0.020	0.2	3262.0
May-09	-36.6	-0.006	0.2	3262.3
Jun-09	-58.3	-0.026	1.5	3263.8
Jul-09	-9.2	-0.039	0.4	3264.1
Aug-09	6.9	-0.046	-0.3	3263.8
Sep-09	6.3	-0.046	-0.3	3263.5
Oct-09	-4.6	-0.041	0.2	3263.7
Nov-09	-21.2	-0.031	0.6	3264.4
Dec-09	-42.2	-0.014	0.6	3265.0
Jan-10	-8.2	0.099	-0.8	3264.1
Feb-10	-18.8	0.065	-1.2	3262.9
Mar-10	-12.7	0.037	-0.5	3262.5
Apr-10	-13.7	0.014	-0.2	3262.3
May-10	-6.6	-0.004	0.0	3262.3
Jun-10	-0.6	-0.018	0.0	3262.3
Jul-10	-13.5	-0.027	0.4	3262.7
Aug-10	13.8	-0.032	-0.4	3262.2
Sep-10	13.2	-0.033	-0.4	3261.8
Oct-10	-37.0	-0.029	1.1	3262.9
Nov-10	31.9	-0.021	-0.7	3262.2
Dec-10	7.9	-0.010	-0.1	3262.1



### 3.3.2. Production industrielle (réf. INSEE : 001654235) ; estimation du mois de décembre 2013.

La production industrielle est modélisée, comme près de la moitié des séries, au moyen d'un airline.



	Innovation	PsiE-weight	Impact sur déc. 2013 (SA)
01/01/2014	-1.2%	0.092	-0.11%
01/02/2014	0.3%	0.078	0.02%
01/03/2014	-0.3%	0.062	-0.02%
01/04/2014	1.6%	0.047	0.08%
01/05/2014	-3.4%	0.031	-0.11%
01/06/2014	1.9%	0.016	0.03%
01/07/2014	0.7%	0.001	0.00%
01/08/2014	-0.2%	-0.013	0.00%
01/09/2014	0.7%	-0.027	-0.02%
01/10/2014	-1.1%	-0.040	0.04%
01/11/2014	-1.6%	-0.053	0.08%
01/12/2014	3.6%	-0.130	-0.47%

#### Explication « intuitive »

Le résultat anormalement bon de décembre 2014 amène à considérer que le facteur saisonnier pour ce mois doit être revu à la hausse. Cette hausse est à répercuter en partie sur les mois de décembres précédents, et en particulier sur décembre 2013. La donnée ajustée des variations saisonnières de décembre 2013 doit donc être revue à la baisse.

## 4. Conclusions

La modélisation fournit un cadre flexible et varié pour la désaisonnalisation. Elle permet également une compréhension précise des estimations qu'elle fournit et de leurs caractéristiques. Même si le sujet n'a pas été abordé dans ce document, ces avantages ne peuvent bien sûr pas faire oublier les limites de la modélisation, notamment en termes de robustesse.

JD+ a été conçu pour prendre en compte à la fois l'analyse à partir des filtres WK et la modélisation par les formes espace-état. De la sorte, il constitue un outil puissant pour comparer et mieux comprendre de multiples approches de la désaisonnalisation par modélisation.

Dans la plupart des cas, l'approche WK, plus facile et plus légère, fournit une information riche sur les caractéristiques du modèle, très proche de la solution exacte donnée par les filtres KF. Ces derniers permettent toutefois d'explorer des modèles plus complexes.

## 5. Bibliographie

Anderson, B. D. O. and Moore, J. B. (1979), "Optimal Filtering", Prentice Hall.

Ansley F. Craig (1979), "An Algorithm for the Exact Likelihood of a Mixed Autoregressive-Moving Average Process", *Biometrika*, 66, 1, 59-65.

Aston, J. A. D., Findley, D. F., Wills, K. C., and Martin, D. E. K. (2004), "Generalizations of the Box-Jenkins Airline Model With Frequency-Specific Seasonal Coefficients", *Proceeding of the NBER/NSF Time Series Conference*.

BELL, W.R. (2004): "ON REGCOMPONENT TIME SERIES MODELS AND THEIR APPLICATIONS", IN *STATE SPACE AND UNOBSERVED COMPONENTS MODELS: THEORY AND APPLICATIONS*, CAMBRIDGE UNIVERSITY PRESS.

Bell, W. R. and Martin, D.E.K. (2002), "Computation of asymmetric signal extraction filters and mean squared error for Arima component models", *Statistical Research Division, Research Report Series #2002-04*.

Box G.E.P, Jenkins G.M. and Reinsel G.C. (1994) "Time Series Analysis. Forecasting and control", third edition, Prentice Hall.

Burman, J.P. (1980), "Seasonal adjustment by signal extraction", *Journal of the Royal Statistical Society, Serie A*. 143, 321-337.

DE JONG P. (1991): "Stable Algorithms For the State Space Model", *Journal of Time Series Analysis*, 12, 2, 143-157.

DE JONG P. AND CHU-CHUN-LIN S. (2003): "Smoothing with an Unknown Initial Condition", *Journal of Time Series Analysis*, 24, 2, 141-148.

DURBIN J. AND KOOPMAN S.J. (2001): "Time Series Analysis by State Space Methods". *Oxford University Press*.

\_\_\_\_\_ (2003): "Filtering and smoothing of state vector for diffuse state space models", *Journal of Time Series Analysis*, vol. 24, n°1, 85 - 98.

Findley D.F. (2005), "Some Recent Developments and Directions in Seasonal Adjustment", *Journal of Official Statistics*, 21, 2, pp. 343–365.

GOMEZ V. AND MARAVALL A. (1993): "Initializing the Kalman Filter with Incompletely Specified Initial Conditions", *WORKING PAPER 93/7, European University Institute*.

\_\_\_\_\_ (1994): "Estimation, Prediction, and Interpolation for Nonstationary Series With the Kalman Filter", *Journal of the American Statistical Association*, vol. 89, n° 426, 611-624.

Gomez V. and Maravall A. (2001), "Seasonal Adjustment and Signal Extraction in Economic Time Series", Ch. 8 in D. Peña, G.C. Tiao and R.S. Tsay, eds., "A Course in Time Series Analysis", New York: J. Wiley and Sons, 202-246.

HARVEY, A.C. (1989): "Forecasting, Structural Time Series Models and the Kalman Filter", *Cambridge University Press*.

- Kailath T. Lev-Ari H. Sated A.H. (1994), "Generalized Chandrasekhar Recursions from the Generalized Schur Algorithm", *IEEE Transactions on Automatic Control*, 39, 11, 2265-2269.
- Kailath T. Lev-Ari H. Sayed A.H. (1999), "Fast Reliable Algorithms for Matrices with Structure", Siam.
- Kailath T., Sayed A.H. and Hassibi B. (1999), "State-Space Estimation", Prentice-Hall, Englewood Cliffs, NJ.
- KOHN R. AND ANSLEY C.F. (1985): "Efficient estimation and prediction in time series regression models", *Biometrika*, 72, 3, 694-697.
- KOOPMAN S.J. (1993): "Disturbance smoother for state space models", *Biometrika*, 80, 1, 117-126.
- KOOPMAN S.J. AND HARVEY A. (1999): "Computing Observation Weights for Signal Extraction and Filtering".
- Ladiray D. and Quenneville B. (2001), "Seasonal Adjustment With the X-11 Method", *Lecture Notes in Statistics*, 158, Springer.
- Ljung G. M., Box G.E.P. (1979), "The Likelihood Function of Stationary Autoregressive-Moving Average Models", *Biometrika*, 66, 2, 265-270.
- McElroy T. (2008), "Matrix formulas for nonstationary ARIMA signal Extraction", *Econometric Theory*, 24, pp 988-1009.
- Oppenheim A.V. and Schafer R.W. (1999), "Discrete-Time Signal Processing", second edition, Prentice Hall.
- Otto M. C., Bell W.R., Burman J.P. (1987), "An Iterative GLS Approach to Maximum Likelihood Estimation of Regression Models with Arima Errors", Bureau of The Census, SRD Research Report CENSUS/SRD/RR\_87/34.
- Palate Jean (2005), "Reusable components for benchmarking using Kalman filters", *Working Papers and Studies*, European Communities.
- Proietti T. (2000), "Comparing seasonal components for structural time series models", *International Journal of Forecasting*, 16, 2, 247-260.
- Proietti T. (2004), "Seasonal specific structural time series", *Studies in Nonlinear Dynamics & Econometrics*, 8, 2, article 16.
- Trimbur T.M., Bell W.R. (2008), "Seasonal Heteroskedasticity in Time Series Data: Modeling, Estimation, and Testing", US Census Bureau, Research Reports Series (Statistics #2008-11)

## A.1. State space form for Arma models

The ARMA process is defined by

$$\Gamma(B)y(t) = \Theta(B)\varepsilon(t)$$

where

$$\Gamma(B) = 1 + \varphi_1 B + \dots + \varphi_p B^p$$

$$\Theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$$

are the auto-regressive and the moving average polynomials.

The MA representation of the model is

$$y(t) = \sum_{i=0}^{\infty} \psi_i \varepsilon(t-i)$$

Let  $\gamma_t$  be the autocovariances of the model. We also define:

$$r = \max(p, q + 1)$$

$$s = r - 1$$

Using those notations, the state-space model can be written as follows<sup>1</sup>:

State vector:

$$\alpha(t) = \begin{pmatrix} y(t) \\ y(t+1|t) \\ \vdots \\ y(t+s|t) \end{pmatrix}$$

where  $y(t+i|t)$  is the orthogonal projection of  $y(t+i)$  on the subspace generated by  $\{y(s): s \leq t\}$ . Thus, it is the forecast function with respect to the semi-infinite sample.

We also have that

---

<sup>1</sup> See for example Gomez-Maravall (1994)

$$y(t+k|t) = \sum_{i=k}^{\infty} \psi_i \varepsilon(t+k-i)$$

System matrices:

The matrices of the model are

$$Z(t) = (\mathbf{1} \quad \mathbf{0} \quad \dots \quad \mathbf{0})$$

$$T(t) = \begin{pmatrix} \mathbf{0} & \mathbf{1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{1} \\ -\varphi_r & \dots & \dots & \dots & -\varphi_1 \end{pmatrix}$$

$$S = S(t) = \begin{pmatrix} \psi_0 \\ \psi_1 \\ \vdots \\ \vdots \\ \psi_s \end{pmatrix}$$

$$Q = Q(t) = \mathbf{1}$$

$$V = V(t) = SS'$$

and the initial conditions can be written:

$$a_0 = (\mathbf{0} \quad \dots \quad \mathbf{0})$$

$$P_{\cdot 0} = \Omega$$

$$P_{\infty 0} = \mathbf{0}$$

$\Omega$  is the unconditional covariance of the state array; it can be easily derived using the MA representation. We have:

$$\Omega(i, \mathbf{0}) = \gamma_{st,i}$$

$$\Omega(i, j) = \Omega(i-1, j-1) - \psi_i \psi_j$$

## A.2. State space form for Arima models

The state space form of an ARIMA process is a generalization of the ARMA representation. The process is defined by

$$\Delta(B)\Gamma(B)y(t) = \Theta(B)\varepsilon(t)$$

where

$$\Delta(B) = 1 + \delta_1 B + \dots + \delta_d B^d$$

$$\Gamma(B) = 1 + \varphi_1 B + \dots + \varphi_p B^p$$

$$\Theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$$

are the differencing, auto-regressive and moving average polynomials. We also write:

$$\Phi(B) = \Delta(B)\Gamma(B) = 1 + \phi_1 B + \dots + \phi_q B^{p+d}$$

Let  $\psi_i$   $\psi_i$  be the psi-weights of the ARIMA model  $\Psi_{sc,t}$ , and  $\gamma_{sc,t}$ , the psi-weights and the autocovariances of the differenced model. We also define:

$$r = \max(p + d, q + 1)$$

$$s = r - 1$$

Using those notations, the state-space model can now be written as follows:

State vector:

See the ARMA model

System matrices:

The matrices of the model are

$$Z(t) = (\mathbf{1} \quad \mathbf{0} \quad \dots \quad \mathbf{0})$$

$$T(t) = \begin{pmatrix} \mathbf{0} & \mathbf{1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{1} \\ -\phi_r & \dots & \dots & \dots & -\phi_1 \end{pmatrix}$$

$$S = S(t) = \begin{pmatrix} \psi_0 \\ \psi_1 \\ \vdots \\ \psi_s \end{pmatrix}$$

$$Q = Q(t) = \mathbf{1}$$

$$V = V(t) = SS'$$

and the initial conditions can be written:

$$a_0 = (\mathbf{0} \quad \dots \quad \mathbf{0})$$

$$P_{-0} = \Sigma \Omega \Sigma'$$

$$P_{\infty 0} = \Lambda \Lambda'$$

$\Omega$  is the variance/covariance of the stationary model. See above.

$\Sigma$  is a  $r \times r$  lower triangular matrix with ones on the main diagonal; other cells are defined by the recursive relationship:

$$\Sigma(i, j) = -\delta_1 \Sigma(i-1, j) - \dots - \delta_d \Sigma(i-d, j), i > j$$

with the convention  $\Sigma(i, j)$  if  $i < 0$ .

$\Lambda$  is a  $r \times d$  matrix; its first  $d$  rows form an identity matrix; other cells are defined as above:

$$\Lambda(i, j) = -\delta_1 \Lambda(i-1, j) - \dots - \delta_d \Lambda(i-d, j)$$

### A.3 State space form for the components of basic structural models

The components of the (basic) structural models are directly defined in their state space form. We present below the most important of them.

#### A.3.1 Local level

$$\eta_{t+1} = \eta_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, v_t)$$

$$\alpha_t = (\eta_t)$$



$$Z_t = (\mathbf{1})$$

$$T_t = (\mathbf{1})$$

$$V_t = (v_t)$$

$$P_{\infty 0} = (\mathbf{1})$$

### A.3.2 Local linear trend

$$\eta_{t+1} = \eta_t + u_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, v_t)$$

$$u_{t+1} = u_t + \omega_t, \quad \omega_t \sim N(0, w_t)$$

$$\alpha_t = \begin{pmatrix} \eta_t \\ u_t \end{pmatrix}$$

$$Z_t = (\mathbf{1} \quad \mathbf{0})$$

$$T_t = \begin{pmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{0} & \mathbf{1} \end{pmatrix}$$

$$V_t = \begin{pmatrix} v_t & \mathbf{0} \\ \mathbf{0} & w_t \end{pmatrix}$$

$$P_{\infty 0} = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix}$$

### A.3.3 Seasonal component (West-Harrison form)

$$\alpha_t = \begin{pmatrix} \gamma_t \\ \gamma_{t-1} \\ \vdots \\ \gamma_{t-s+2} \end{pmatrix}$$

$$Z_t = (\mathbf{1} \quad \mathbf{0} \quad \dots \quad \mathbf{0})$$

$$T_t = \begin{pmatrix} -\mathbf{1} & -\mathbf{1} & \dots & -\mathbf{1} & -\mathbf{1} \\ \mathbf{1} & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & \backslash & \backslash & \backslash & \vdots \\ \vdots & \backslash & \mathbf{1} & \mathbf{0} & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{1} & \mathbf{0} \end{pmatrix}$$

$$P_{\infty 0} = I_{s-1}$$

### A.3.4 Harrison-Stevens seasonal component (with variances varying with the seasons)

$$\alpha_t = \begin{pmatrix} \delta_{0,t} \\ \delta_{1,t} \\ \vdots \\ \delta_{s-2,t} \end{pmatrix}$$

$$T_t = I_{s-1}$$

$$V_t = D \begin{pmatrix} v_{0,t} & \mathbf{0} \\ \mathbf{0} & v_{s-1,t} \end{pmatrix} D'$$

$$D = (I_{s-1} \quad \mathbf{0}_{s-1 \times 1}) - \frac{1}{s} \mathbf{1}_{s-1 \times s}$$

$$Z_t = \begin{cases} (\dots \quad \mathbf{1}_{t \bmod s} \quad \dots), & t \bmod s \neq s-1 \\ (-\mathbf{1} \quad \dots \quad -\mathbf{1}), & t \bmod s = s-1 \end{cases}$$

$$F_{\text{co0}} = I_{s-1}$$

## B.1 Factorisation d'un modèle ARIMA.

Reprenant les notations du texte, on suppose que les polynômes  $\Phi(B), \Theta(B), \Phi_s(B), \Phi_n(B)$  sont de degrés respectifs  $p, q, p_s, p_n = p - p_s$

Dans le domaine des fréquences, par le remplacement des opérateurs  $B, F$  par  $e^{i\omega}, e^{-i\omega}$ , les fonctions génératrices  $\Theta(B)\Theta(F)$  deviennent  $g_\Theta(\omega)$

Plus spécifiquement, si

$$\Theta(B)\Theta(F) = \alpha_{-q}F^q + \dots + \alpha_{-1}F + \alpha_0 + \alpha_1B + \dots + \alpha_qB^q$$

on a que

$$g_\Theta(\omega) = \alpha_{00} + 2\alpha_{01} \cos \omega + \dots + 2\alpha_{0q} \cos q\omega$$

La relation (2) peut ainsi s'écrire

$$v_s g_{\Theta_s}(\omega) g_{\Phi_n}(\omega) + v_n g_{\Theta_n}(\omega) g_{\Phi_s}(\omega) = v g_\Theta(\omega)$$

Tenant compte de la relation trigonométrique

$$\cos j \cos k = \frac{1}{2} \cos(j+k) \cos(j-k) = \frac{1}{2} \cos(|j+k|) \cos(|j-k|)$$

on peut écrire que :

$$A \begin{pmatrix} 1/2 \alpha_{\Theta_s 0} \\ \alpha_{\Theta_s 1} \\ \vdots \\ \alpha_{\Theta_s q_s} \\ 1/2 \alpha_{\Theta_n 0} \\ \alpha_{\Theta_n 1} \\ \vdots \\ \alpha_{\Theta_n q_n} \end{pmatrix} = \begin{pmatrix} 1/2 \alpha_{\Theta 0} \\ \alpha_{\Theta 1} \\ \vdots \\ \alpha_{\Theta q} \\ [0] \end{pmatrix}$$

Les dimensions  $q_s, q_n$  sont définies comme suit :

$$q_s = \begin{cases} p_s - 1, & p_s > 0 \\ 0, & p_s = 0 \end{cases}$$

$$q_n = \begin{cases} p_n - 1, & p_n > 0 \text{ et } q < p \\ q - p_s, & p_n > 0 \text{ et } q \geq p \\ 0, & p_n = 0 \end{cases}$$

Les coefficients des polynômes « moyennes mobiles » s'obtiennent alors aisément en résolvant ce système linéaire, par exemple via l'algorithme QR (solution dans JD+).