2 - Décompostion avec X-11

Anna Smyk et Tanguy Barthélémy
Division Recueil et Traitement de l'Information
Département des Méthodes Statistiques

Objectifs de cette séquence

Objectifs : définitions et propriétés des moyennes mobiles (MM) de X11.

Après cette séquence, vous :

- connaîtrez mieux les principales propriétés des moyennes mobiles
- pourrez identifier les moyennes mobiles utilisées par X11
- pourrez modifier les choix de X-11

Réactivation

Le **module initiation** permet de répondre aux questions (importantes) suivantes :

Qu'est ce que moyenne mobile centrée et symétrique ?

Quel lien entre l'ordre de la moyenne mobile et périodicité de la série ?

Quel type de problème affecte le traitement de la fin de série avec des moyennes mobiles ?

Nous allons revoir et approfondir les réponses.

Les moyennes mobiles (1/2)

Moyenne mobile d'*ordre* $p + f + 1$ de coefficients (θ_i), l'opérateur M défini par :

$$MX_t = \sum_{i=-p}^{f} \theta_i X_{t+i}$$

Valeur en t remplacée par une moyenne pondérée de p valeurs passées, de la valeur courante et de f valeurs futures.

Exemple de moyenne mobile simple (tous les coeffs égaux) d'ordre 3 :

$$MX_t = \frac{1}{3}(X_{t-2} + X_{t-1} + X_t)$$

Les moyennes mobiles (2/2)

Si $p = f$, la moyenne mobile est dite *centrée*.

Si, de plus $\theta_{-i} = \theta_i$, elle est dite *symétrique*.

Remarque : une moyenne d'ordre pair ne peut être centrée et symétrique.

Objectif : Construire une moyenne centrée et symétrique en combinant 2 moyennes d'ordre pair.

$$\left.\begin{aligned} M_1X_t &= \tfrac{1}{4}(X_{t-2} + X_{t-1} + X_t + X_{t+1}) \\ M_2X_t &= \tfrac{1}{4}(X_{t-1} + X_t + X_{t+1} + X_{t+2}) \end{aligned}\right\} \rightarrow M_{2\times 4} = \frac{M_1 + M_2}{2}$$

La moyenne $M_{2\times 12}$ a été vue dans le module initiation :

$$M_{2\times 12} = \frac{1}{24}(X_{t-6}) + \frac{1}{12}(X_{t-5} + X_{t-4}$$

$$+X_{t-3} + X_{t-2} + X_{t-1} + X_t + X_{t+1} + X_{t+2} + X_{t+3} + X_{t+4} + X_{t+5}) + \frac{1}{24}(X_{t+6})$$

Propriétes des moyennes mobiles

Propriété évidente, la *linéarité* :

$$X_t = TC_t + S_t + I_t$$
$$MX_t = M(TC_t) + M(S_t) + M(I_t)$$

Trois propriétés utiles pour la désaisonnalisation :

- P1 : élimination de la saisonnalité $M(S_t) = 0$
- P2 : conservation de la tendance $M(TC_t) = TC_t$
- P3 : réduction de l'irrégulier ie. $\mathbb{V}\left[M(I_t)\right]$ minimale

P1 : Élimination de la saisonnalité

Élimination de saisonnalité constante = élimination de mouvements périodiques.

À retenir : Il suffit de choisir une MM simple d'ordre de la périodicité.

Dans le module initiation, on a montré qu'une MM simple d'ordre 12 élimine les saisonnalités (mensuelles) stables et qu'il en était donc de même pour la moyenne composée $M_{2\times 12}$, qui est centrée symétrique.

P2 : Conservation des tendances linéaires

Toute MM **symétrique** qui conserve les constantes (resp. les polynômes de degré d) conserve les droites (resp. les polynômes de degré $d+1$).La série résultante est linéaire par morceaux.

Posons $X_t = C$, la conservation de la constante s'écrit :

$$MX_t = \sum_{i=-p}^{p} \theta_i X_{t+i} = C \sum_{i=-p}^{p} \theta_i => \sum_{i=-p}^{p} \theta_i = 1$$

Dans ce cas, posons $X_t = at + b$,

$$MX_t = \sum_{i=-p}^{p} \theta_i X_{t+i} = \sum_{i=-p}^{p} \theta_i(a(t+i)) + b \sum_{i=-p}^{p} \theta_i$$

$$MX_t = \sum_{i=-p}^{p} \theta_i X_{t+i} = at \sum_{i=-p}^{p} \theta_i + a \sum_{i=-p}^{p} i\theta_i + b \sum_{i=-p}^{p} \theta_i$$

On utilise la condition précédente $\sum_{i=-p}^{p} \theta_i = 1$ et $\sum_{i=-p}^{p} i\theta_i = 0$ qui découle du caractère symétrique de la moyenne mobile pour conclure : $M(at+b) = at + b$.

Les MM sont **sensibles aux points extrêmes** :

P2 : Conservation d'une tendance linéaire (Questions)

Exercice : les moyennes mobiles M suivantes conservent-elles les droites ?

$$M_1[X_t] = \frac{1}{3}(X_{t-1} + X_t + X_{t+1})$$
$$M_2[X_t] = \frac{1}{4}(X_{t-2} + X_{t-1} + X_t + X_{t+1})$$
$$M_3[X_t] = \frac{1}{4}(X_{t-2} + 2X_{t-1} + X_t + 2X_{t+1} + X_{t+2})$$

Qu'en est-il des moyennes $M_{2\times 12}$ et $M_{3\times 3}$?

P2 : Conservation d'une tendance linéaire (Réponses)

Il faut regarder si elles sont symétriques et si elles conservent les constantes.

$M_1[X_t]$: oui

$M_2[X_t]$: non (pas symétrique)

$M_3[X_t]$: non (symétrique mais ne conserve pas les constantes)

La $M_{2\times 12}$ est symétrique avec les coefficients : $[\frac{1}{24}, 11 \text{ fois } \frac{1}{12}, \frac{1}{24}]$.

La $M_{3\times 3}$ a pour coefficients $[\frac{1}{9}, \frac{2}{9}, \frac{3}{9}, \frac{2}{9}, \frac{1}{9}]$.

Dans les deux cas, on a bien $\sum_{i=-p}^{p} \theta_i = 1$, donc elles conservent les droites.

P3 : Réduction du bruit (1/2)

On ne peut éliminer la composante irrégulière I mais on peut l'atténuer.

On veut $\mathbb{V}\left[M(I_t)\right]$ minimale.

Une moyenne mobile transforme un bruit blanc de variance σ^2 en un bruit de moyenne nulle, autocorrélé et de variance :

$$\sigma^{*2} = \sigma^2 \sum_{i=-p}^{p} \theta_i^2$$

car

$$V(M(\varepsilon_t)) = V(\sum_{i=-p}^{p} \theta_i \varepsilon_{t+i}) = \sum_{i=-p}^{p} \theta_i^2 V(\varepsilon_{t+i}) = \sigma^2 \sum_{i=-p}^{p} \theta_i^2$$

car $cov(\varepsilon_t, \varepsilon_{t'}) = 0$

La quantité $\sum_{i=-p}^{p} \theta_i^2$ est le *pouvoir réducteur* de la moyenne mobile.

P3 : Réduction du bruit (2/2)

Réduire la variance du bruit $\equiv$ **minimiser la somme des carrés des coefficients**.

Pouvoir réducteur de la moyenne mobile $M_{2\times 12}$:

La $M_{2\times 12}$ est symétrique avec les coefficients : $[\frac{1}{24}, 11 \text{ fois } \frac{1}{12}, \frac{1}{24}]$

Donc $\sum_{i=-p}^{p} \theta_i^2 = 2 \times (\frac{1}{24})^2 + 11 \times (\frac{1}{12})^2 = 0.07986111$.

Pouvoir réducteur de la moyenne mobile $M_{3\times 3}$:

La $M_{3\times 3}$ a pour coefficients $[\frac{1}{9}, \frac{2}{9}, \frac{3}{9}, \frac{2}{9}, \frac{1}{9}]$

Donc $\sum_{i=-p}^{p} \theta_i^2 = 2 \times (\frac{1}{9})^2 + 2 \times (\frac{2}{9})^2 + 1 \times (\frac{3}{9})^2 = 0.23457$.

Le problème des fins de série

Une moyenne mobile centrée d'ordre 2p+1 ne peut être appliquée aux « p » premiers ni aux « p » derniers points

Solution 1 : utiliser des moyennes mobiles asymétriques

Les MM asymétriques de MUSGRAVE permettent de minimiser les révisions (associées à celles d'Henderson)

Méthode historique...en voie de réapparition ?

Solution 2 : prolonger la série par prévision et appliquer une moyenne mobile symétrique (par défaut 12 mois prévus)

(Les prévisions sont une combinaison linéaire du passé, ça reste asymétrique, mais « mieux » que MUSGRAVE.)

Les moyennes mobiles dans X11

- MM centrée d'ordre la périodicité de la série (2x12, 13 termes, pour les séries mensuelles) : conserve des droites, supprime une saisonnalité stable, relativement faible pouvoir de lissage
- MM de *Macurves* (asymétrique sur la fin de série), pour extraire la composante saisonnière : 3x3 (5 termes), 3x5 (7 termes), 3x9 et 3x15
- MM de *Henderson* (5, 7, 9, 13 ou 23 termes) : conserve des polynômes, bon pouvoir de lissage
- MM de *Musgrave* tournent sur les prévisions

Le choix de la longueur des filtres dépend des variations de l'irrégulier par rapport à celles de la tendance T (Henderson) et de la saisonnalité S (Macurves)

cf. Principe Itératif de X-11 vu lors de la session "Initiation"

Choice of trend (Henderson) filter length (1/2)

The algorithm chooses between different filter lengths automatically according to the I/C ratio, the user can modify this choice (first step is computed with H_{13})

$$\frac{I}{C} = \frac{\sum_t \left| \frac{\tilde{\imath}_t}{\tilde{\imath}_{t-1}} - 1 \right|}{\sum_t \left| \frac{\tilde{t}_t}{\tilde{t}_{t-1}} - 1 \right|},$$

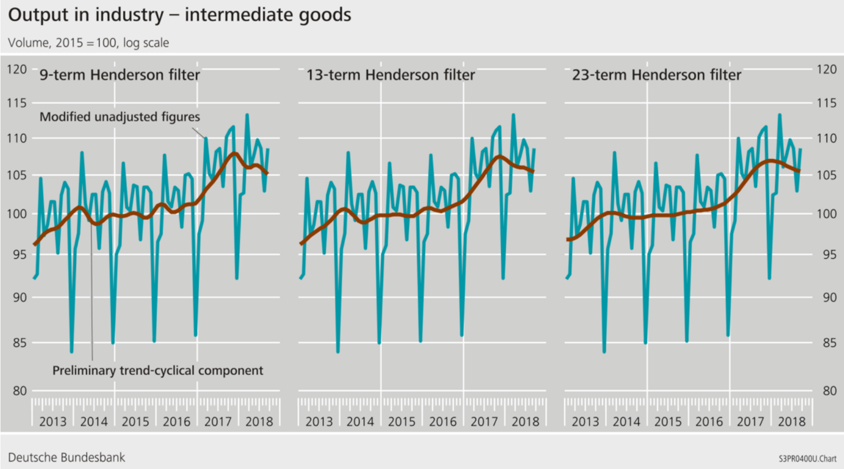
with $\tilde{\imath}_t$ = temporary irregular
$\tilde{t}_t$ = temporary trend-cycled

	Decision rule		
I/C	$[0,1)$	$[1,3.5)$	$[3.5,\infty)$
Henderson filter (m)	9-term	13-term	23-term

Aim:

- dominance of irregular (I/C ratio large) → choose long filter
- dominance of trend-cycle (I/C ratio small) → choose short filter

Choice of trend (Henderson) filter length (2/2)



Choice of seasonal filter length (1/2)

The algorithm chooses between different filter lengths automatically according to the I/S ratio, the user can modify this choice (first step is computed with $M_{3\times 3}$)

$$\frac{I}{S} = \frac{\sum_t \left| \frac{\tilde{\imath}_t}{\tilde{\imath}_{t-12}} - 1 \right|}{\sum_t \left| \frac{\tilde{s}_t}{\tilde{s}_{t-12}} - 1 \right|},$$

with $\tilde{\imath}_t$ = temporary irregular
$\tilde{s}_t$ = temporary seasonal

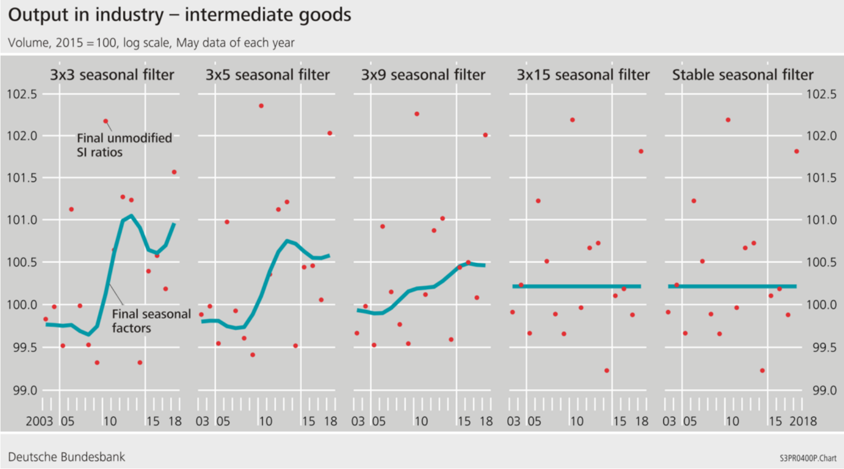
	Decision rule				
I/S	$[0, 2.5)$	$[2.5, 3.5]$	$(3.5, 5.5)$	$[5.5, 6.5]$	$[6.5, \infty)$
Seasonal filter	3×3	???	3×5	???	3×9

??? Maximum of five I/S recalculations under omission of the respective last year, application of 3×5 in case still no decision could be taken.

Aim:

- dominance of irregular (I/S ratio large) → choose long filter
- dominance of trend-cycle (I/S ratio small) → choose short filter

Choice of seasonal filter length (2/2)



(correction des valeurs extremes non détaillée ici)

Decomposition quality in X11 : the M-stats

- M statistics :$0 < M_x < 3$, acceptance region $M_x \leq 3$
- 11 statistics of the decomposition quality (M1 to M11) and 2 summary indicators (Q et Q-M2)
- Warning : not to be used exclusively as basis for decision making (very old benchmarking)

M1 to M6 : focus on the size of the irregular component (1/2)

A large influence of the irregular component makes identifying and separating the components more difficult

- **M1** The relative contribution of the irregular over three months span
- **M2** The relative contribution of the irregular component to the stationary portion of the variance

→ reduce I : outliers or *sigma limit*, filters' length

- **M3** The amount of month to month change in the irregular component as compared to the amount of month to month change in the trend-cycle (I/C-ratio)
- **M5** MCD (Months for Cyclical Dominance) : The number of months it takes the change in the trend-cycle to surpass the amount of change in the irregular

→ **M3** and **M5** can be ignored when the trend is flat, important only for economic cycles analysis

M1 to M6 : focus on the size of the irregular component (2/2)

M4 test whether I is a white noise vs an AR(1). Fails often (because of calendar effects, data collection...)

$\rightarrow$ independent from the other M-stats, can be ignored

M6 : only valid if the seasonal filter is $M_{3\times 5}$, because the test verifies that this MA is appropriate.

$\rightarrow$ **If M6 fails** and the global MSR is large (6.5), choose a long filter. If the MSR is small (2.5), choose a short filter.

$\rightarrow$ MSR by month : Decomposition > Quality Measures > Details

M7 to M11 : focus on seasonal factors

M7 The amount of moving seasonality present relative to the amount of stable seasonality M8 The size of the fluctuations in the seasonal component throughout the whole series M9 The average linear movement in the seasonal component throughout the whole series M10 Same as 8, calculated for recent years only (4 years, N-2 to N-5) M11 Same as 9, calculated for recent years only

M7 indicates an identifiable seasonality $\rightarrow$ **M7 is important.** Cannot be improved in the X11 part, only during the pre-adjustment phase : correction of seasonal breaks, non-seasonal series, too short (model adaptation) or too long (remove the start from the modelling) series, multiplicative decomposition...

Parameters in GUI

Parameter options for x11 in JDemetra+	
Parameter	Options (*default*)
Mode	*Undefined*, Additive, Multiplicative, LogAdditive, PseudoAdditive
Seasonal component	*yes*/no
Forecasts horizon	no. of periods (positive values) or years (negative values) (*-1*)
Backcasts horizon	no. of periods (positive values) or years (negative values) (*0*)
LSigma	> 0.5 (*1.5*)
USigma	> LSigma (*2.5*)
Seasonal filter	3x1, 3x3, 3x5, 3x9, 3x15, stable, X11Default, *Msr*
Details on seasonal filters	period specific filters
Automatic henderson filter	*yes*/no
Henderson filter	odd number [3,101] (*13*)
Calendarsigma	*None*, Signif, All, Select
Excludeforecast	yes/*no*

Les essentiels

X11 est une application successive de différentes moyennes mobiles ayant des fonctions spécifiques.

X11 decompose la série linéarisée. (les effets des outliers sont re-allouées à la fin)

Les 3 principales propriétés des MM : suppression des fonctions périodiques, conservation des droites ou polynômes et pouvoir réducteur du bruit.

Les deux indicateurs de qualité de la décomposition les plus importants sont M7 (essentiel) et Q2 (dans une moindre mesure).