3a - Méthode X13-ARIMA: décomposition avec X-11

Désaisonnalisation avec JDemetra+

Anna Smyk & Tanguy Barthélémy (Insee)

Insee
Mesurer pour comprendre

Time Series Software
for Official Statistics

Sommaire I

Section 1

Introduction

X13-ARIMA

X pour eXperience...

Deux modules :

- X11 : phase de décomposition

Décomposition de la série en tendance-cycle, saisonnalité et irrégulier, à l'aide de **moyennes mobiles**.

- REG-ARIMA : phase de pré-ajustement pour obtenir une série linéarisée (séquence suivante)
  Correction **préalable** par régression linéaire des points aberrants, ruptures de tendance, effets de calendrier.

Objectif de cette séquence: comprendre la phase de décomposition X11

Section 2

Phase de décomposition (X11)

Subsection 1

Les moyennes mobiles

Moyennes mobiles : définition (1/2)

Dans X-13-Arima, la série est décomposée à l'aide de moyennes mobiles. Le module de décomposition est souvent appelé X-11 (module historique). Bien qu'on en soit à X-13, la décomposition a peu varié.

Il est nécessaire de connaître quelques concepts sur les moyennes mobiles pour comprendre la phase de décomposition.

La moyenne mobile d'*ordre* $p+f+1$ de coefficients (θ_i) est 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.

Notée usuellement MX_t, la moyenne mobile est bien une fonction, on pourrait écrire $M(X_t)$

Moyennes mobiles : définition (2/2)

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

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

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

Les moyennes mobiles centrées symétriques sont celles qui ont les propriétés les plus intéressantes pour la décomposition (car elles conservent les droites).

Exemples de moyenne mobile simple d'ordre 3

Deux exemples de moyennes mobiles simples (tous les coefficients égaux) d'ordre 3 :

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

$\rightarrow$ cette moyenne mobile n'est pas centrée (donc pas symétrique non plus)

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

$\rightarrow$ celle-là est centrée et symétrique.

Moyennes mobiles : linéarité et composition

Une MM est un opérateur linéaire :
Linéarité : $M(X_t + Y_t) = M(X_t) + M(Y_t)$

$$X_t = T_t + S_t + I_t$$
$$\rightarrow MX_t = M(T_t) + M(S_t) + M(I_t)$$

Composition de moyennes mobiles

Moyenne arithmétique de p Moyennes Mobiles de même ordre (longueur) : $M_{p \times ordre}$

Moyennes mobiles : exemple de composition à l'ordre 12 (1/2)

Pour une MM d'ordre 12, deux écritures (naturelles) sont possibles :

$M_{1\times 12}$

$$M1X_t = \frac{1}{12}(X_{t-6} + X_{t-5} + X_{t-4} + X_{t-3} + X_{t-2} + X_{t-1} + X_t + X_{t+1} + X_{t+2} + X_{t+3} + X_{t+4} + X_{t+5})$$

Ou bien :

$M_{1\times 12}$ bis

$$M2X_t = \frac{1}{12}(X_{t-5} + X_{t-4} + X_{t-3} + X_{t-2} + X_{t-1} + X_t + X_{t+1} + X_{t+2} + X_{t+3} + X_{t+4} + X_{t+5} + X_{t+6})$$

Cette deuxième version a un point de moins dans le passé et un point de plus dans le futur. L'ordre 12 étant PAIR, on ne peut pas obtenir une moyenne mobile simple centrée symétrique.

Moyennes mobiles : exemple de composition à l'ordre 12 (2/2)

La **composition** permet d'obtenir une moyenne mobile centrée symétrique pour un **ordre pair**.

$$M_{2\times 12} = \frac{1}{2}(M1X_t + M2X_t)$$

ce qui donne, lorsque l'on développe et regroupe :

$$\begin{gathered}M_{2\times 12} = \frac{1}{24}(X_{t-6}) + \frac{1}{12}(X_{t-5} + X_{t-4} \\ + X_{t-3} + X_{t-2} + X_{t-1} + X_t + X_{t+1} + X_{t+2} \\ + X_{t+3} + X_{t+4} + X_{t+5}) + \frac{1}{24}(X_{t+6})\end{gathered}$$

On obtient une moyenne mobile centrée symétrique à $1 + (5 + 1 + 5) + 1 = 13$ termes (demi-poids aux extremités)

Moyennes mobiles : élimination de la saisonnalité

Si l'on se place dans l'hypothèse vue d'une saisonnalité constante : [{i=1}^{12}S{t+i} = 0]

L'effet d'une moyenne mobile d'ordre 12 sera de supprimer une saisonnalité mensuelle localement stable $M_{1\times 12}(S) = 0$

La moyenne $M_{2\times 12}$ aura aussi cet effet

$$M_{2\times 12}(S) = \frac{1}{2}(M1X_t(S) + M2X_t(S)) = \frac{1}{2}(0+0) = 0$$

L'avantage de la $M_{2\times 12}$ sur la $M_{1\times 12}$ est d'être centrée symétrique.

PROPRIETE ESSENTIELLE : une moyenne mobile dont l'ordre est égal à la périodicité élimine une saisonnalité localement stable.

Moyennes mobiles : extraction de la saisonnalité (1/4)

La saisonnalité est donc éliminée avec une moyenne mobile où ordre = périodicité

$$M_{2\times12}(X_t) = M_{2\times12}(T+S+I) = M_{2\times12}(T) + M_{2\times12}(S) + M_{2\times12}(I)$$

Comme $M_{2\times12}(S) = 0$, en négligeant* I à ce stade, on obtient une approximation de T. Puis de $S+I$ par soustraction (car $S+I = X-T$).

On va calculer S en négligeant* I.
Le calcul se fait période par période : type de mois par type de mois, type de trimestre par type de trimestre (on considère : la sous-série des janvier, des février...)
Pas de mélange de types mois/trimestres à ce stade, car on cherche à estimer ce qui est commun à chaque type de période.

Si on cherchait à estimer une saisonnalité strictement constante, le facteur S d'une période donnée serait égal à la moyenne empirique des $\widehat{S+I}$ de l'ensemble des valeurs correspondant à ce type de période.

Moyennes mobiles : extraction de la saisonnalité (2/4)

Dans le cas d'une saisonnalité strictement constante :
Pour calculer le coefficient S, commun à tous les mois d'avril de la série (par hypothèse), en notant $T =$ nombre d'avrils dans la série :

$$S_{avril} = \frac{1}{T}(\widehat{S+I}_{avril,1} + \ldots + \widehat{S+I}_{avril,T})$$

Toutefois, on considère que l'hypothèse d'une saisonnalité strictement constante est trop restrictive.

On va laisser la saisonnalité évoluer lentement au fil des ans en utilisant des moyennes mobiles 3×3 ou 3×5, le plus souvent. En effet, les MM permettent de faire contribuer un nombre limité de voisins à l'estimation du S d'une période. De plus, les poids des voisins décroissent lorsqu'ils sont plus lointains $\rightarrow$ importance moindre quand éloignement temporel.

Négliger I est une approximation justifiée dans ce calcul car les moyennes mobiles utilisées dans les deux cas réduisent I. (pas détaillé ici)

Moyennes mobiles : extraction de la saisonnalité (3/4)

La moyenne mobile 3×3 est une composition des moyennes mobiles simples d'ordre 3 vues en début de séquence.

$$M1X_t = \frac{1}{3}(X_{t-2} + X_{t-1} + X_t)$$

$$M2X_t = \frac{1}{3}(X_{t-1} + X_t + X_{t+1})$$

$$M3X_t = \frac{1}{3}(X_t + X_{t+1} + X_{t+2})$$

$$M_{3\times 3}X = \frac{1}{3}(M1X_t + M2X_t + M3X_t)$$

On obtient, après avoir développé et regroupé, une moyenne mobile centrée symétrique à 5 termes :

$$M_{3\times 3}X = \frac{1}{9}(X_{t-2}) + \frac{2}{9}(X_{t-1}) + \frac{3}{9}(X_t) + \frac{2}{9}(X_{t+1}) + \frac{1}{9}(X_{t+2})$$

Les fractions ont été laissées non simplifiées à dessein.

Moyennes mobiles : extraction de la saisonnalité (4/4)

La moyenne mobile 3×5, aussi utilisée par X-11, fonctionne sur le même principe : moyenne arithmétique de 3 moyennes simples d'ordre 5, qui est une moyenne mobile centrée symétrique à 7 termes.

Intérêt d'une moyenne mobile composée vs une moyenne mobile simple :

- pour l'élimination de la saisonnalité : obtenir une moyenne mobile symétrique d'ordre égal à la périodicité, alors que la périodicité est paire.
- pour l'extraction de la saisonnalité : attribuer des poids décroissants aux valeurs éloignées et réduire I.

Subsection 2

Le principe itératif de X11

Principe itératif de X11 (1/2)

Une première estimation de la CVS :

1. Estimation de la **tendance-cycle** par moyenne mobile 2×12 :

$$T_t^{(1)} = M_{2\times 12}(X_t)$$

2. Estimation de la composante **saisonnier-irrégulier** :

$$(S_t + I_t)^{(1)} = X_t - T_t^{(1)}$$

3. Estimation de la composante **saisonnière** par moyenne mobile 3×3 sur **chaque mois** :

$$S_t^{(1)} = M_{3\times 3}\left[(S_t + I_t)^{(1)}\right] \text{ et normalisation } Snorm_t^{(1)} = S_t^{(1)} - M_{2\times 12}\left(S_t^{(1)}\right)$$

4. Première estimation de la série corrigée des variations saisonnières :

$$Xsa_t^{(1)} = (T_t + I_t)^{(1)} = X_t - Snorm_t^{(1)}$$

Principe itératif de X11 (2/2)

Une seconde estimation de la CVS :

1. Estimation de la **tendance-cycle** par moyenne de Henderson (généralement 13 termes, cf infra) :

$$T_t^{(2)} = H_{13}(Xsa_t^{(1)})$$

2. Estimation de la composante **saisonnier-irrégulier** :

$$(S_t + I_t)^{(2)} = X_t - T_t^{(2)}$$

3. Estimation de la composante **saisonnière** par moyenne mobile 3×5 (généralement) pour **chaque mois/trimestre** :

$$S_t^{(2)} = M_{3\times5}\left[(S_t + I_t)^{(2)}\right] \text{ et normalisation } Snorm_t^{(2)} = S_t^{(2)} - M_{2\times12}\left(S_t^{(2)}\right)$$

4. Estimation de la série corrigée des variations saisonnières :

$$Xsa_t^{(2)} = X_t - Snorm_t^{(2)}$$

Bilan : les différentes moyennes mobiles utilisées par X11 (1/2)

3 types de MM utilisés par X11 :

1. Moyennes mobiles d'ordre = la périodicité (ex. $M_{2\times 12}$) pour éliminer une saisonnalité localement stable : $M(S_t) = 0$

On utilise la $M_{2\times 12}$ et pas simplement $M_{1\times 12}$, car les propriétés de symétrie sont importantes.

2. Moyennes mobiles $M_{3\times k}$ avec k impair, pour extraire la saisonnalité
3. Moyennes mobiles de Henderson (pour extraire la tendance d'une série NON saisonnière) $\rightarrow H_{13}$
   - conservent la tendance polynômiale (ordre 3) : $M(at^3 + bt^2 + ct + d) = at^3 + bt^2 + ct + d$
   - réduisent le bruit au maximum
   - n'éliminent pas la saisonnalité

Bilan : les différentes moyennes mobiles utilisées par X11 (2/2)

NB. Les Moyennes Mobiles d'extraction de la saisonnalité sont des compositions de MM d'ordre impair.

Elles peuvent être des 3×3 ou 3×5 ou 3×9… La longueur n'est pas la même à toutes les étapes de l'algorithme et elle est en partie paramétrable par l'utilisateur.

Les étapes de X11

3 grandes étapes

Étapes B et C : lissage de la série (enlève les points aberrants)

Étape D : désaisonnalisation finale (avec l'algorithme de désaisonnalisation décrit précédemment)

On retrouve les séries intermédiaires et finales dans JDemetra+.

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.)

Section 3

Choix des paramètres

Choix du filtre de tendance (Henderson)(1/2){allowfram}

L'algorithme choisit entre différentes longueur de filtres sur la base du ratio I/C (C désigne ici T, notation d'origine conservée)

Les calculs des premières étapes sont faits avec H_{13})

L'utilisateur peut modifier ce choix pour l'étape finale (étape 2 de la partie D)

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

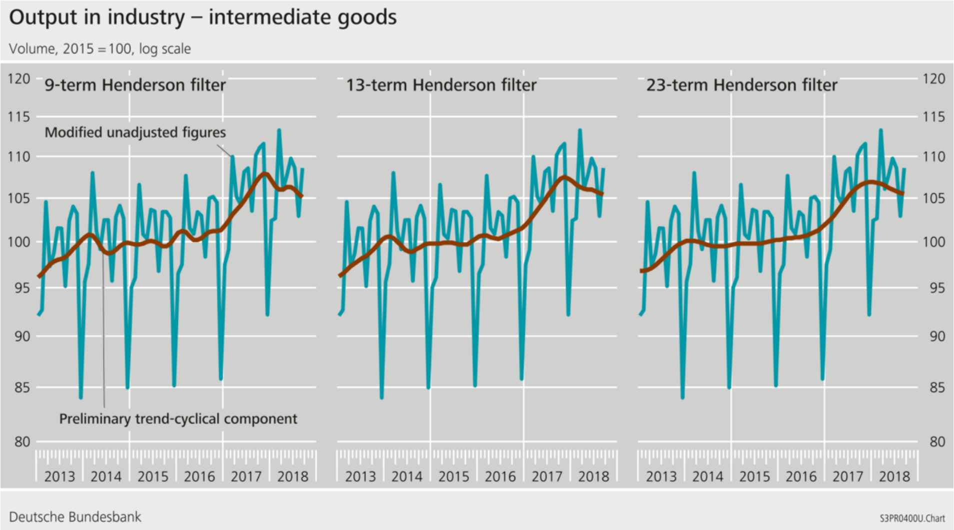
with $\tilde{\imath}_t$ = temporary irregular
$\tilde{t}_t$ = temporary trend-cycled

	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

Choix du filtre de tendance (Henderson) (2/2)



Choix du filtre d'extraction de la saisonnalité (1/2)

L'algorithme choisit entre différentes longueurs de filtre sur la base du ratio I/S. Les calculs des premières étapes sont faits avec $M_{3\times 3}$.

L'utilisateur peut modifier ce choix pour l'étape finale (étape 2 de la partie D).

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

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

$\tilde{s}_t =$ temporary seasonal

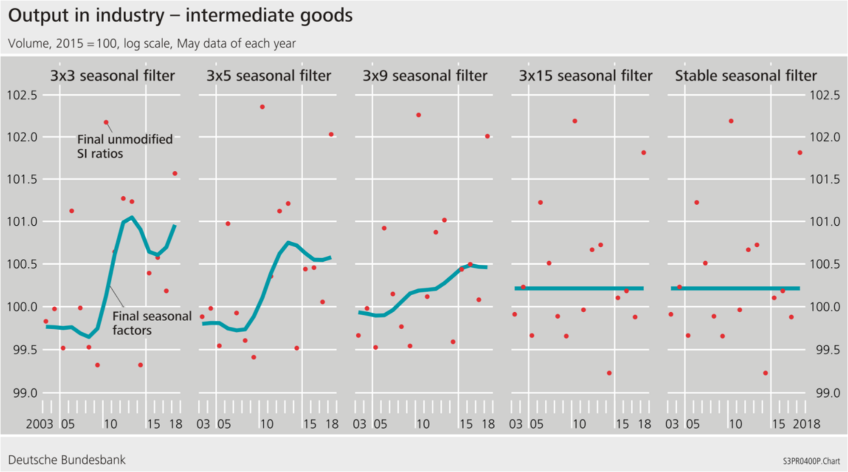
	Decision rule				
I/S	$[0, 2.5)$	$[2.5, 3.5]$	$(3.5, 5.5)$	$[5.5, 6.5]$	$[6.5, \infty)$
Seasonal filter	3×3	???	3×5	???	3×9

??? 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

Choix du filtre d'extraction de la saisonnalité (2/2)



(correction des valeurs extrêmes non détaillée ici)

Les statistiques M (1/2)

11 statistiques sur la qualité de la décomposition (M1 à M11) et deux statistiques moyennes (Q et Q-M2) (seuils de 1 calculés empiriquement)

M1 et M2 : contribution de l'irrégulier à la variance de la série stationnarisée

M3 et M5 : comparent les variations de I sur T (noté C)

$\rightarrow$ si tendance plate à ignorer

M4 teste I bruit blanc versus hyp AR(1). Si échec des CJO, outliers...

Éventuellement améliorer la linéarisation

M6, valable si filtre S est $M_{3\times5}$, vérifie si ce choix est adapté.

$\rightarrow$ **Si M6 échoue** et MSR global grand (6,5), choisir filtre long, filtre court si petit (2,5)

$\rightarrow$ regarder aussi les MSR par mois Decomposition > Quality Measures > Details

Les statistiques M (2/2)

M7 indique si saisonnalité identifiable

→ **M7 est important.** Rien à faire dans X11, actions en amont : rupture de S à corriger, série non saisonnière, série trop courte (modèle) ou trop longue (supprimer le début), schéma multiplicatif

Statistiques sur la fin de la série :

M8 et M9 mesurent respectivement variations de S à court et à long terme (linéairement)

M10 et M11 mêmes indicateurs sur la fin de série (4 années, N-2 à N-5)

Les statistiques Q et les priorités La stat Q est une moyenne pondérée des 11 stat M

La stat Q2 exclut la M2

Par ordre d'importance : - M7 - Q2 - M6 si filtre $M_{3\times5}$ - ...

Idée : agir au maximum dans la phase de pré-traitement (span, calendrier, outliers..)

Parametrès ajustables à l'interface

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*

Section 4

Conclusion

Les essentiels

- L'algorithme X13-ARIMA travaille en deux phases : pré-ajustement et décomposition
- Le pré-ajustement linéarise (par régression) et prolonge les séries en faisant des prévisions (par modèle ARIMA)
- La décomposition X11 estime les composantes T, S, I et calcule la série CVS (T+I ou T*I)
- X11 décompose la série linéarisée
- X11 utilise successivement plusieurs moyennes mobiles ayant des propriétés complémentaires
- Les deux indicateurs de qualité de la décomposition les plus importants sont M7 (essentiel) et Q2 (dans une moindre mesure).