Initiation à JDemetra+ et à la désaisonnalisation

3 - Méthode X13-ARIMA

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

Objectifs de cette séquence

Cette séquence a pour objectif de vous présenter la méthode X13-ARIMA.

Après cette séquence vous connaîtrez :

- les concepts relatifs à la méthode X13-ARIMA
- la structure de la méthode X13-ARIMA en deux étapes

  `- pré-ajustement`

  `- décomposition`

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 (modélisation)

Correction **préalable** par régression linéaire des points aberrants, ruptures de tendance, effets de calendrier.

Modélisation ARIMA : pour prolonger la série brute afin de résoudre partiellement le problème des fins de série lié aux moyennes mobiles symétriques.

Sommaire

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*

Nous verrons dans la suite que les moyennes mobiles centrées symétriques sont celles qui ont les propriétés les plus intéressantes pour la décomposition.

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 : $M(X_t + Y_t) = M(X_t) + M(Y_t)$

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

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 :

$$\sum_{i=1}^{12} S_{t+i} = 0$$

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. On verra en détail pour quoi c'est important dans le module perfectionnement.

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

$$M_{2\times 12}(X_t) = M_{2\times 12}(T + S + I) = M_{2\times 12}(T) + M_{2\times 12}(S) + M_{2\times 12}(I)$$

Comme $M_{2\times 12}(S) = 0$, 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 :

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

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, on l'expliquera lors du module perfectionnement.

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.

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

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

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

$$M_{3\times3}X = \frac{1}{3}(M1X_t + M2X_t + M3X_t)$$

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

$$M_{3\times3}X = \frac{1}{9}(X_{t-2}) + \frac{2}{9}(X_{t-1}) + \frac{3}{9}(X_t) + \frac{2}{9}(X_{t+1}) + \frac{1}{9}(X_{t+2})$$

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 (cf. Module perfectionnement).

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** :
$$S_t^{(2)} = M_{3\times 5}\left[(S_t + I_t)^{(2)}\right] \text{ et normalisation } Snorm_t^{(2)} = S_t^{(2)} - M_{2\times 12}\left(S_t^{(2)}\right)$$
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 (cf. module perfectionnement).

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 $3 \times 9\ldots$ La longueur n'est pas la même à toutes les étapes de l'algorithme et elle est en partie paramétrable par l'utilisateur (cf. module perfectionnement).

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 JD+.

Sommaire

Linéariser la série

On décrit ici ce que fait l'algorithme AVANT la décomposition

Enlève par regression les

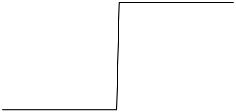
- outliers (points aberrants et ruptures)
- effets de calendrier $\rightarrow$ séquence de demain matin

Série linéarisée $= (Y_t - \sum \hat{\alpha}_i X_{it})$ où les X_i modélisent les effets deterministes

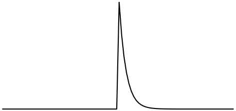
Les principaux types d'outliers

Choc ponctuel
Additive outlier (AO)
Alloué in fine, après la décomposition, à l'Irrégulier

Changement de niveau
Level Shift (LS)
Alloué in fine, après la décomposition, à la Tendance

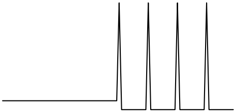


Changement de niveau transitoire
Transitory Change (TC)
Alloué in fine, après la décomposition, à l'Irrégulier



Seasonal Outlier

Rupture de profil saisonnier
Seasonal Outlier (SO)



Très rarement utilisé, non detecté automatiquement par défaut dans JD+
Affecte directement la Composante Saisonnière

La modélisation Reg-ARIMA

Le modèle Reg-ARIMA s'écrit comme suit :

$$\left(Y_t - \sum \alpha_i X_{it}\right) \sim ARIMA(p, d, q)(P, D, Q)$$

où les variables (regresseurs) X_i representent les effets de calendrier et les outliers

Le modèle ARIMA décrit la structure d'autocorrélation du modèle de régression : il capture toute l'information temporelle. Le résidu du modèle Reg-ARIMA est un bruit blanc.

Le modèle ARIMA est utilisé pour :

- permettre l'estimation du modèle de régression (on sort des MCO car autocorrélations)
- prévoir la série linéarisée, afin que des moyennes mobiles symétriques puissent être utilisées par X-11 jusqu'au dernier point brut disponible.

La structure du modèle Reg-ARIMA et sa détermination seront vues en détail lors du module perfectionnement.

Sommaire

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 utilise successivement plusieurs moyennes mobiles ayant des propriétés complémentaires