

4 - La correction des effets de calendrier

Désaisonnalisation avec JDemetra+

Anna Smyk & Tanguy Barthélémy (Insee)



Sommaire II

Tests de significativité des coefficients

Test de l'effet graduel de Pâques

⑤ Conclusion

Section 1

Introduction

Objectifs de cette séquence

Cette séquence a pour but de vous présenter les différents effets de calendrier et la manière de corriger une série de ces effets.

Après cette séquence vous saurez :

- distinguer les différents types d'effet de calendrier
- identifier les raisons pour lesquelles il est utile de corriger une série de ces effets
- modéliser la correction d'une série
- lire les diagnostics disponibles sous JDemetra+

Section 2

Pourquoi corriger des effets de calendrier ?

Subsection 1

Un calendrier hétérogène

Un calendrier hétérogène

Le calendrier est hétérogène :

- Les jours ouvrables :
 - jours normalement travaillés compte tenu du calendrier français, le plus souvent, il s'agit des lundis, mardis,..., vendredis non fériés
- Les week-ends
- Les jours fériés (fêtes)

Tous les mois n'ont pas la même composition :

⇒ Tous les mois ne sont pas « égaux » entre eux, même pour un type de mois donné on parle d'**effets de calendrier**

Objectif de la correction : rendre les périodes de même nature (les janvier, les T1..) comparables entre elles avant de désaisonnaliser (estimer un effet commun à chaque type de période).

Les différents effets de calendrier

- Effet longueur du mois/trimestre

La production est en principe plus élevée au cours d'un mois comportant davantage de jours ouvrables (Se traduira par l'effet « année bissextile » (*leap year*))

- Effet type de jour

Les ventes du commerce de détail sont plus importantes le samedi que les autres jours de la semaine

- Effets des jours fériés

Date fixe : ces jours fériés tombent toujours le même mois mais pas forcément le même jour de la semaine (en France 01/01, 01/05, 08/05, 14/07, 15/08, 01/11, 11/11, 25/12) Fêtes Mobiles en France: Lundi de Pâques, Jeudi de l'Ascension, Lundi de Pentecôte

L'effet graduel de Pâques

Pour certaines séries les variations liées à Pâques peuvent s'observer pendant les jours ou les semaines qui précèdent la fête : c'est ce qu'on appelle *effet graduel de Pâques*

Exemple : les ventes de fleurs et de chocolats augmentent sensiblement à l'approche de Pâques

Cet effet est rarement pertinent, attention à ne pas le corriger par défaut.

Subsection 2

Corriger pour comparer

Corriger pour comparer, comme pour la correction des variations saisonnières

Il est nécessaire de corriger des effets de calendrier pour pouvoir faire :

- Des comparaisons temporelles
 - Exemple : mois de longueurs différentes
- Des comparaisons sectorielles (i.e. entre différents secteurs d'activité)
 - Exemple : le commerce et l'industrie (effet « type de jours »)
- Des comparaisons spatiales (exemple : entre différents pays)
 - Exemple : la France et l'Allemagne n'ont pas le même nombre de jours ouvrables par mois
 - Jours fériés différents
 - Religions différentes (calendrier orthodoxe vs calendrier chrétien)

Section 3

Comment corriger des effets de calendrier ?

Subsection 1

De la saisonnalité dans le calendrier

Les effets de calendrier sont en partie saisonniers

Une part des effets de calendrier est saisonnière :

- Le nombre de jours ouvrables du mois de février est presque toujours inférieur à celui du mois de mars
- Certains mois comptent plus de jours fériés (et donc moins de jours ouvrables) que les autres mois
 - Exemple : le mois de mai en France

⇒ Une part des effets de calendrier est de toute façon déjà prise en compte dans la correction des variations saisonnières

Subsection 2

Approche économétrique pour corriger des effets de calendrier

Approche économétrique

Hypothèse : l'effet de la longueur du mois (ou du trimestre) ou d'un certain type de jours de la semaine est constant sur toute la période d'étude. On privilégie une approche économétrique : on va estimer un effet « moyen » par régression lineaire.

Une méthode plus simple et longtemps utilisée en l'absence de logiciels économétriques performants, consistait à appliquer un coefficient à la série brute, avec pour chaque mois :

$$\text{coefficient} = \frac{\text{nombre de jours moyen de ce type de mois sur « longue période »}}{\text{nombre de jours ouvrables du mois}}$$

Important : dans les méthodes présentées ici, la saisonnalité peut évoluer (moyennes mobiles) mais PAS les effets de calendrier. Cela a des conséquences si la série est très longue : les effets de calendrier peuvent être mal estimés en fin de série, qui est en général la période d'intérêt.

Hypothèses sur les types de jours

Construction d'un jeu de regressseurs « effets de calendrier » adapté à chaque série

Pour capter des effets « significatifs » et réduire le nombre de paramètres à estimer dans la régression, il est nécessaire de formuler des hypothèses, c'est à dire de fixer les similitudes et les différences entre les types de jours.

- Lundis non fériés
- Mardis non fériés
- ...
- Dimanches non fériés
- Lundis fériés
- Mardis fériés
- ...
- Dimanches fériés

Usuellement, on traite tous les jours fériés comme des dimanches, et on distingue plus ou moins les autres jours entre eux, selon le sens des grandeurs mesurées.

Les caractéristiques des regressseurs :

- désaisonnalisés

On cherche à désaisonnaliser les regressseurs au préalable, afin de ne pas affaiblir le signal saisonnier avant la décomposition.

- pertinents d'un point de vue économique : on formule des hypothèses sur les effets des différents types de jours

Ce qui amène à construire différents jeux de regressseurs (variables explicatives), puis à choisir le jeu qui corrige le mieux une série donnée.

La diapo suivante montre les différents jeux proposés par le DMS. On voit que les variables sont exprimées en contraste → expliqué plus loin en détaillant le modèle.

Subsection 3

Les différents jeux de regressseurs

Exemples de jeux de regressseurs

Jeu de re- gresseurs	Hypothèses	Référence (contraste)	Nombre de regresseurs
REG1	(lundi non férié = ... = vendredi non férié) et (samedi = dimanche = jours fériés)	Samedis, dimanches et jours fériés	1 + LPY
REG2	(lundi non férié = ... = vendredi non férié), (samedi) et (dimanche = jours fériés)	Dimanches et jours fériés	2 + LPY
REG3	(lundi non férié), (mardi non férié = ... = vendredi non férié) et (samedi = dimanche = jours fériés)	Samedis, dimanches et jours fériés	3 + LPY
REG5	(lundi non férié), ..., (vendredi non férié) et (samedi = dimanche = jours fériés)	Samedis, dimanches et jours fériés	5 + LPY
REG6	(lundi non férié), ..., (vendredi non férié), (samedi) et (dimanche = jours fériés)	Dimanches et jours fériés	6 + LPY
Leap Year (LPY)	Tous les jours ont le même effet	Tous les jours	1

Modèle économétrique (1/4)

On se place dans le cadre d'un modèle linéaire multivarié et on introduit le nombre de jours d'un type donné comme variable explicative, soit 7 variables.

$$X_t = \sum_{i=1}^7 \alpha_i N_{it} + \varepsilon_t$$

- N_{it} est le nombre de jours de lundis ($i = 1$), ..., dimanches ($i = 7$)
- Exemple: $N_{3,t=jan2007}$ est le nombre de mercredis en janvier 2007 (soit 4 ou 5)
- α_i est l'effet d'un jour de type i (coefficient associé à la variable)

Comme $N_t = \sum_{i=1}^7 N_{it}$, il y a un problème de colinéarité entre régresseurs, qui empêcherait d'estimer les coefficients α_i

Modèle économétrique (2/4)

On réécrit le modèle en décomposant α_i , l'effet d'un jour de type i , en deux :

- effet spécifique d'un jour de type i : β_i
- effet moyen d'un jour quelconque $\bar{\alpha}$, avec $\bar{\alpha} = \frac{1}{7} \sum_{i=1}^7 \alpha_i$

Ainsi on pose : $\alpha_i = \beta_i + \bar{\alpha}$

En remplaçant α_i le modèle s'écrit :

$$X_t = \sum_{i=1}^7 \beta_i N_{it} + \bar{\alpha} \sum_{i=1}^7 N_{it} + \varepsilon_t$$

Or $\sum_{i=1}^7 N_{it}$ est le nombre de jours du mois t que l'on notera N_t

Soit :

$$X_t = \sum_{i=1}^7 \beta_i N_{it} + \bar{\alpha} N_t + \varepsilon_t$$

Modèle économétrique (3/4)

On constate de plus que $\sum_{i=1}^7 \beta_i = 0$

L'effet spécifique s'annule sur une semaine, comme la saisonnalité s'annule sur un an. Cela permet d'exprimer un β_i en fonction des autres et de se débarrasser du problème de colinéarité.

Usuellement on associe β_7 au dimanche et on écrit

$$\beta_7 = -\sum_{i=1}^6 \beta_i$$

En remplaçant β_7 par $-\sum_{i=1}^6 \beta_i$, le modèle devient :

$$X_t = \sum_{i=1}^6 \beta_i N_{it} - \sum_{i=1}^6 \beta_i N_{7t} + \bar{\alpha} N_t + \varepsilon_t$$

Modèle économétrique (4/4)

Le terme $\bar{\alpha}N_t$ est purement saisonnier sauf pour les mois de février. On va désaisonnaliser ce terme en lui retirant sa moyenne de long terme : on remplace N_t par $(N_t - \bar{N}_t)$

On a $\bar{N}_t = N_t$ sauf pour les mois de février où $\bar{N}_t = 28.25$, donc $(N_t - \bar{N}_t) = 29 - 28.25 = 0.75$ pour les années bissextiles et $(N_t - \bar{N}_t) = 28 - 28.25 = -0.25$ pour les années non bissextiles

Et finalement le modèle s'écrit

$$X_t = \sum_{i=1}^6 \beta_i (N_{it} - N_{7t}) + \bar{\alpha} (N_t - \bar{N}_t) + \varepsilon_t$$

Il met en évidence deux termes :

- l'effet spécifique d'un type de jour i sous forme de contraste par rapport au nombre de dimanches et jours fériés
- un effet de longueur de mois qui, désaisonnalisé, se réduit à un effet d'année bissextile (Leap year ou LY dans les logiciels)

Subsection 4

Comment choisir et valider le choix du jeu de regressseurs ?

Quelles questions se poser

Avant de choisir un jeu de régresseurs, il faut se poser les questions suivantes :

- est-ce que la série peut présenter des effets de calendrier (sens économique) ?
- quels jours ont a priori un effet sur les valeurs de la série ?
 - Les jours non fériés du lundi au samedi ?
 - Les jours non fériés du lundi au vendredi ?
 - Tous les jours non fériés ont-ils a priori le même effet sur les valeurs de la série ?

Pour plus de détail sur la construction des régresseurs et l'estimation des modèles, il est conseillé de lire l'article correspondant qui figure dans le répertoire Biblio.

Section 4

Les outils disponibles sous JDemetra+

Subsection 1

Les regresseurs de JDemetra+

Les outils disponibles sous JDemetra+ en version 2

En version 2 JDemetra+ propose 2 jeux de regresseurs :

- Trading Days (6 regresseurs)
 - On distingue tous les jours de la semaine du lundi au samedi
- Working Days (1 regresseur)
 - On distingue les jours de la semaine (lundi = mardi = ... = vendredi) et les week-ends
- Leap year (1 regresseur ou corrigé a priori sur la série brute)

Attention : JDemetra+ ne prend pas en compte par défaut l'ensemble des jours fériés d'un calendrier national, mais permet d'importer ses propres jeux de regresseurs ou de définir un calendrier.

Les outils disponibles sous JDemetra+ en version 3

En version 3 JDemetra+ propose

- 5 jeux de regresseurs et une possibilité de les selectionner automatiquement selon 3 critères statistiques différents (demo)
- Leap year (1 regresseur ou corrigé a priori sur la série brute)

Attention : JDemetra+ ne prend pas en compte par défaut l'ensemble des jours fériés d'un calendrier national, mais permet d'importer ses propres jeux de regresseurs ou de définir un calendrier.

Subsection 2

Tests de significativité des coefficients

Deux tests disponibles

Dans les diagnostics (« Pre-processing »), JD+ fournit des tests de significativité portant sur le jeu de régresseurs choisi par l'utilisateur :

- Test de Fisher de nullité conjointe des coefficients (H_0 tous les coefficients sont nuls et H_1 au moins un des coefficients n'est pas nul)
 - Peut conduire à enlever le jeu de régresseurs du modèle Reg-ARIMA
- Pour chaque régresseur, test de Student de nullité du coefficient (H_0 le coefficient est nul et H_1 le coefficient n'est pas nul)
 - Peut conduire à changer de jeu de régresseurs

Subsection 3

Test de l'effet graduel de Pâques

Présence d'un effet graduel de Pâques ?

Est-ce que la série peut présenter un effet graduel de Pâques ?

- Si oui, on introduit un régresseur permettant de corriger la série de l'effet graduel
 - Variable « Easter » dans JDemetra+
- Si non, on n'en introduit pas
- JDemetra+ peut réaliser un « pré-test » afin de savoir si la série présente un effet graduel de Pâques. Si c'est le cas, JD+ introduit le régresseur « Easter [n] » ($n=1, \dots, 20$) dans le modèle

Dans les diagnostics (« Pre-processing »), JDemetra+ fournit un test de la significativité du coefficient associé au régresseur « Easter [n] » (test de Student : H_0 coeff(Easter [n]) = 0 et H_1 coeff(Easter [n]) \neq 0)

Section 5

Conclusion

Les essentiels

- La correction des effets de calendrier est nécessaire pour faire des comparaisons temporelles et spatiales (mécanisme identique à la correction des variations saisonnières, mais effets estimés fixes sur la période)
- On distingue 3 différents effets : effet « longueur du mois » , effet « type de jours » et effet Pâques, toutefois très rare
- On peut élaborer ses propres jeux de regressseurs en faisant des hypothèses sur la similitude des types de jours
- JDemetra+ teste la présence d'effets de calendrier résiduels