

Régresseurs pour effets de calendrier : comment les construire, comment les choisir ?

*Journées de Méthodologie Statistique,
24-26 janvier 2012*

Ketty Attal-Toubert
Département des statistiques de court terme



Introduction

Les effets de calendrier sont présents dans de nombreuses séries économiques.

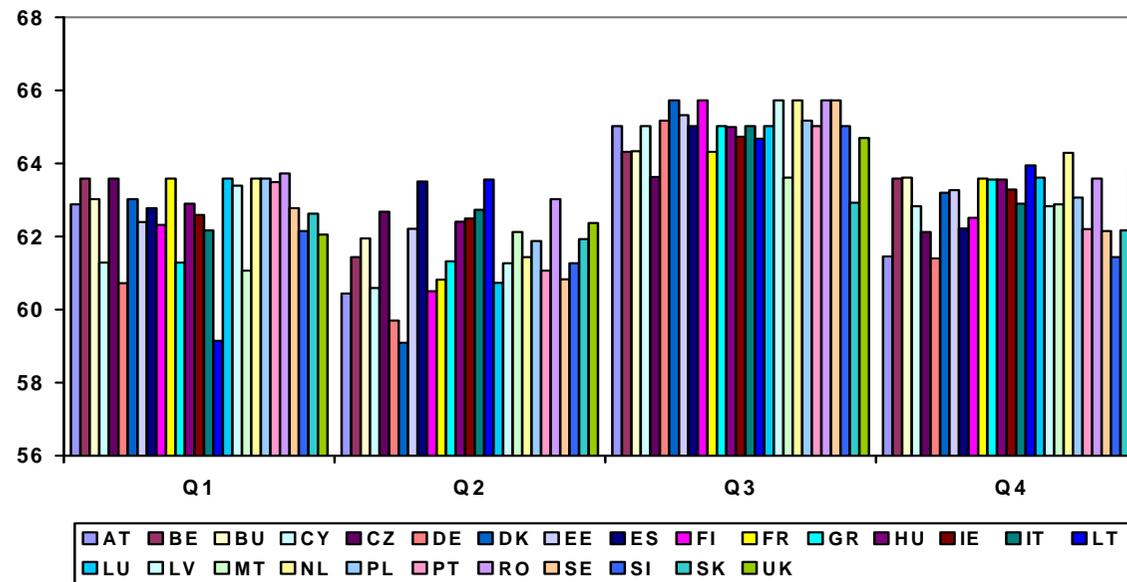
Dans les logiciels de désaisonnalisation X12-ARIMA et Tramo-Seats, ils sont détectés et estimés à l'aide de régresseurs dans un modèle **Reg-ARIMA**.

Les régresseurs sont soit **prédéfinis**, soit **spécifiés par l'utilisateur**.

Dans cette présentation, on cherche à construire des régresseurs adaptés au **calendrier national** (jours fériés) et tenant compte du **secteur d'activité** sous revue.

Pourquoi tenir compte des spécificités des calendriers nationaux? (1)

Nombre de jours ouvrables par trimestre dans les pays de l'UE



Impact potentiel important sur les séries économiques => Chaque pays doit effectuer ses propres corrections des effets de JO pour les séries économiques diffusées à Eurostat.

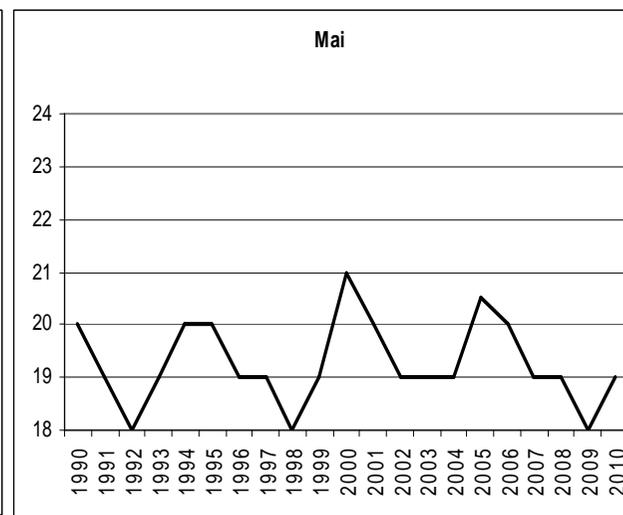
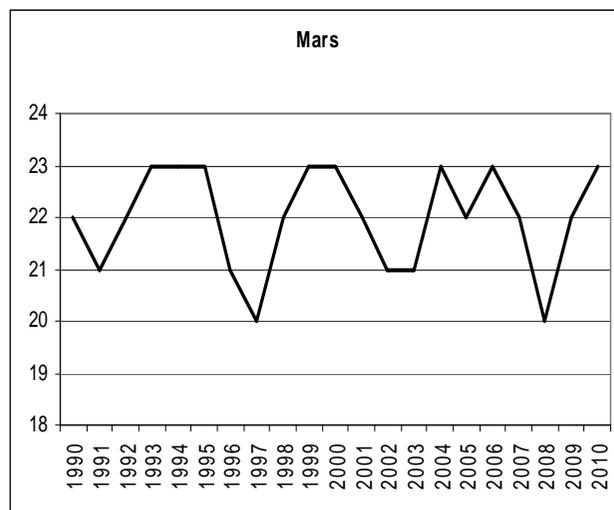
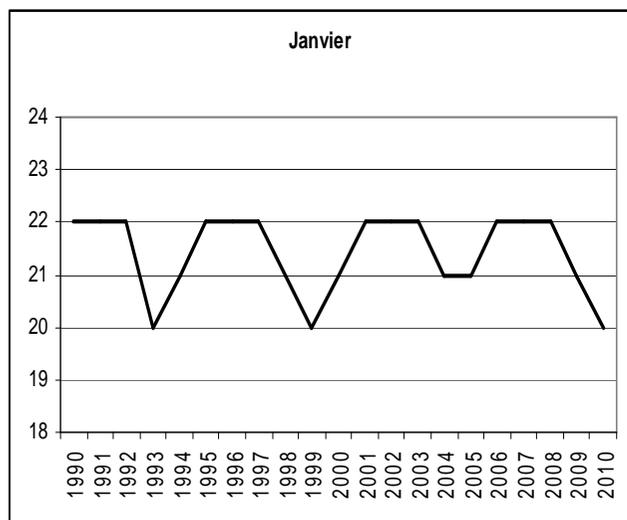
Exemple : nombre de jours ouvrables par trimestre

Pour un trimestre donné, le nombre de jours ouvrables peut être très différent d'une année à l'autre. Cet effet potentiel n'est pas supprimé par la correction de variations saisonnières

	Q1	Q2	Q3	Q4
2011	64	62	64	63
2012	65	60	64	64
2013	63	60	65	63
2014	63	60	64	64
2015	63	60	65	64
2016	63	63	64	63
Moyenne 1940-2010	63,54	61,32	64,30	63,59

Exemple : nombre de jours ouvrables par mois

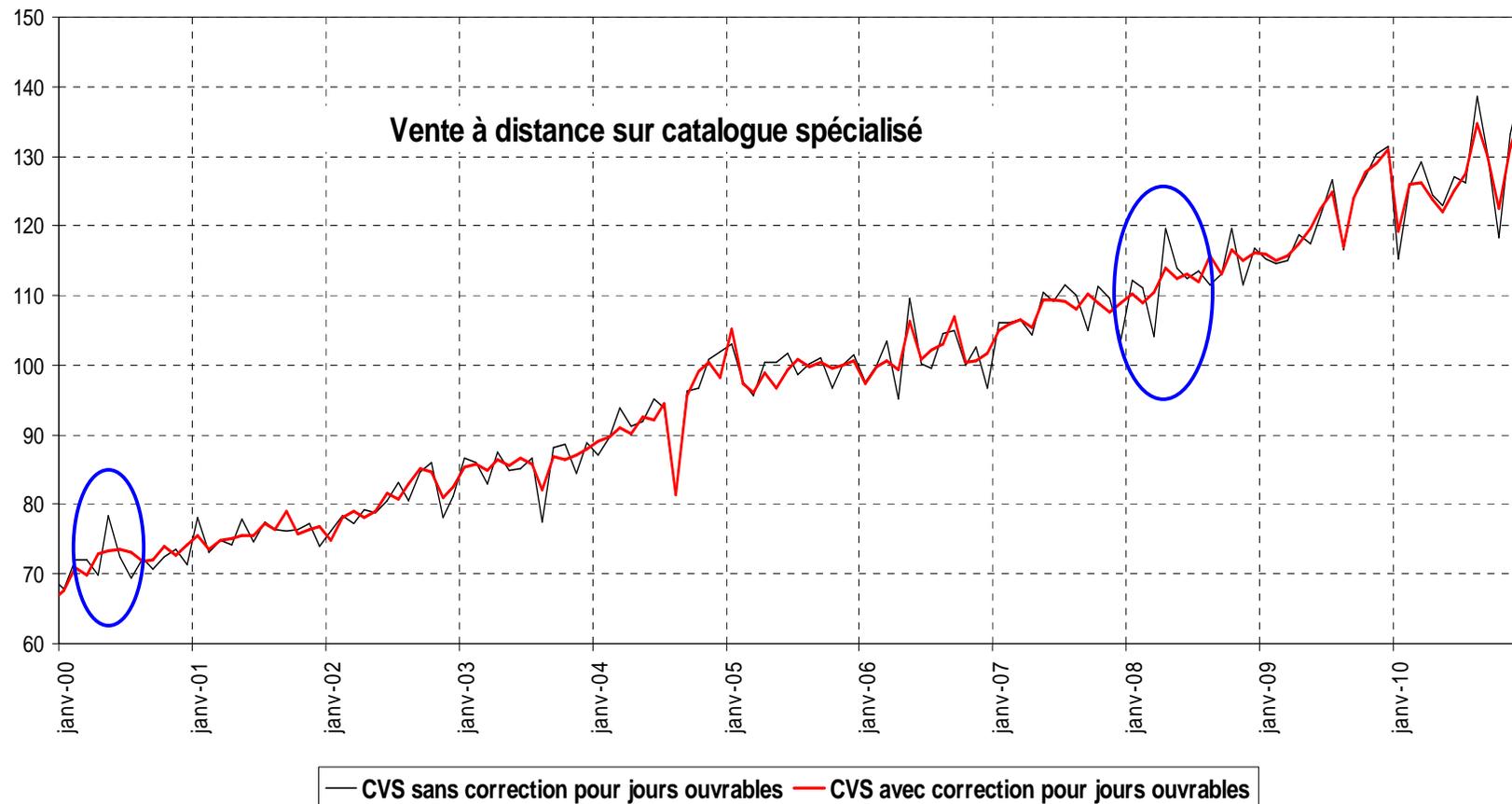
Exemple : variation du nombre de jours ouvrables sur les mois de janvier, mars et mai sur la période 1990-2011



Impact de la correction des effets de calendrier

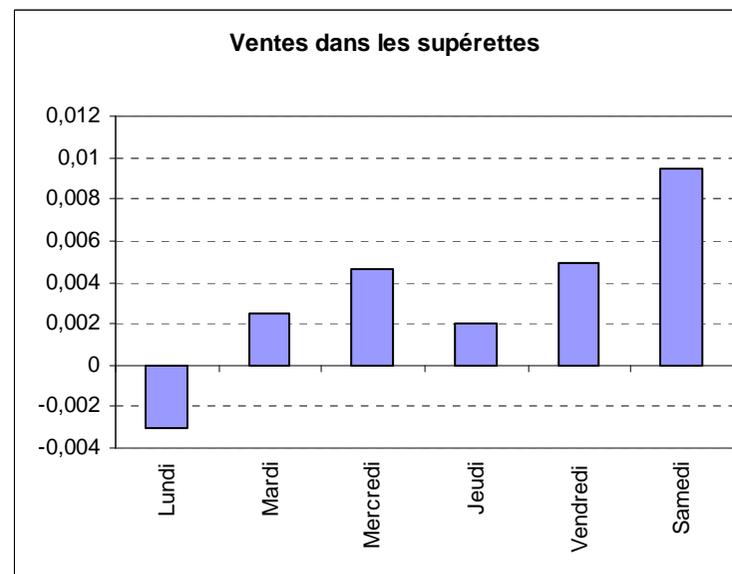
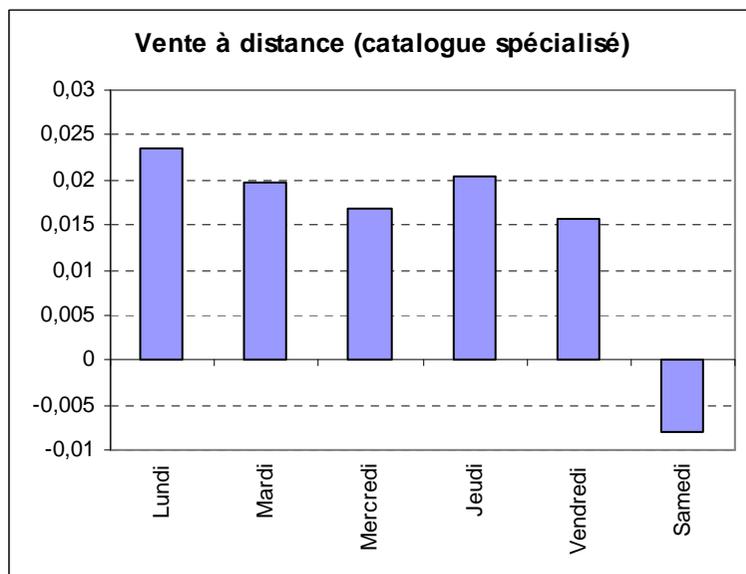
Correction des effets de “jours ouvrables”

Ex. : Série de la vente à distance sur catalogue spécialisé



Des effets de calendrier très différents d'un secteur d'activité à l'autre

Exemples de coefficients CJO pour les indices de chiffre d'affaires dans 2 secteurs du commerce de détail



Idée : Classification des coefficients CJO

Classification des coefficients CJO

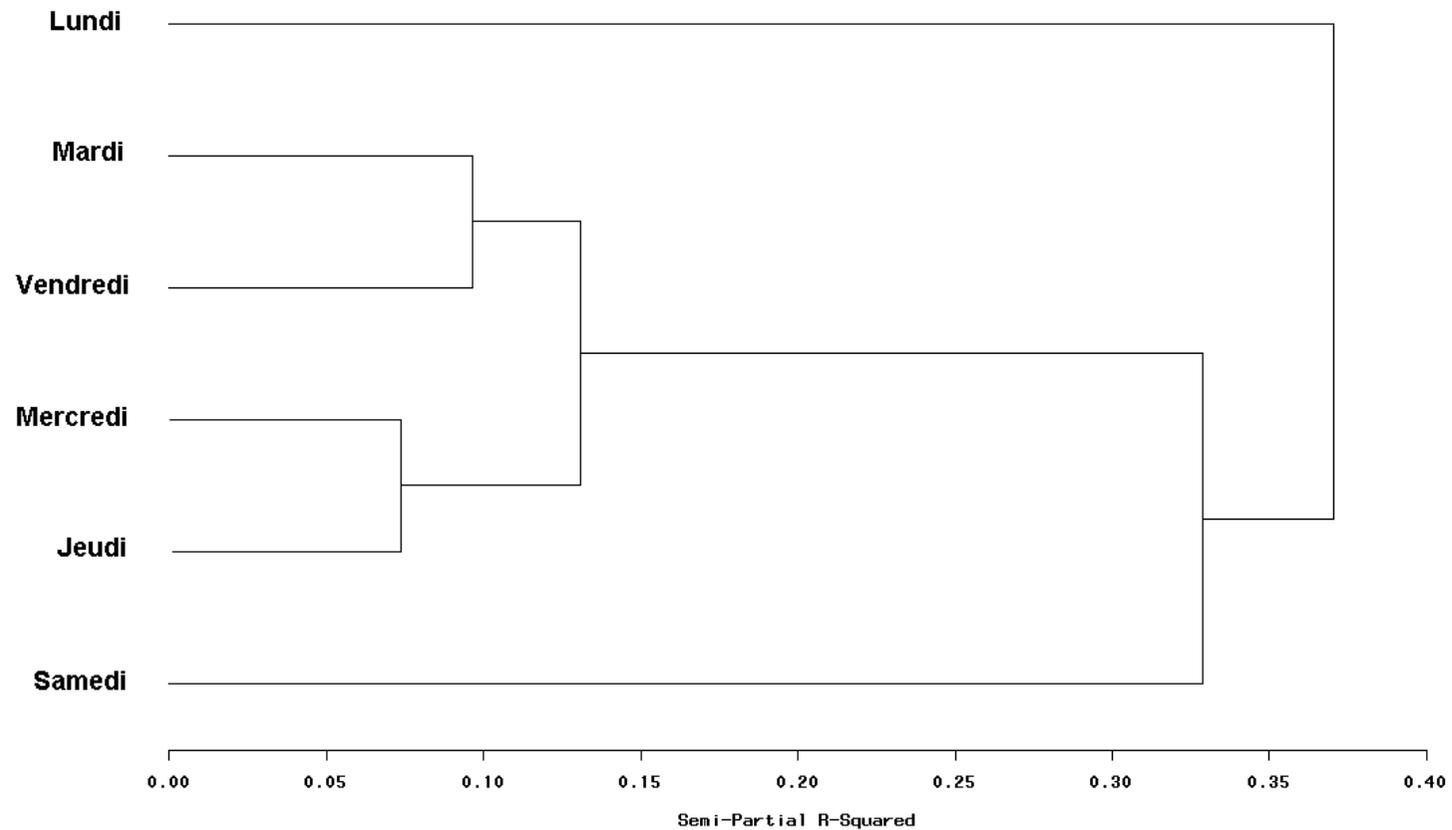
Traitement des 57 séries d'indices de chiffres d'affaires dans le commerce de détail par X12-ARIMA avec les 6 régresseurs distinguant les lundis, mardis,..., samedis (+ LeapYear).

Classification ascendante hiérarchique des coefficients estimés.

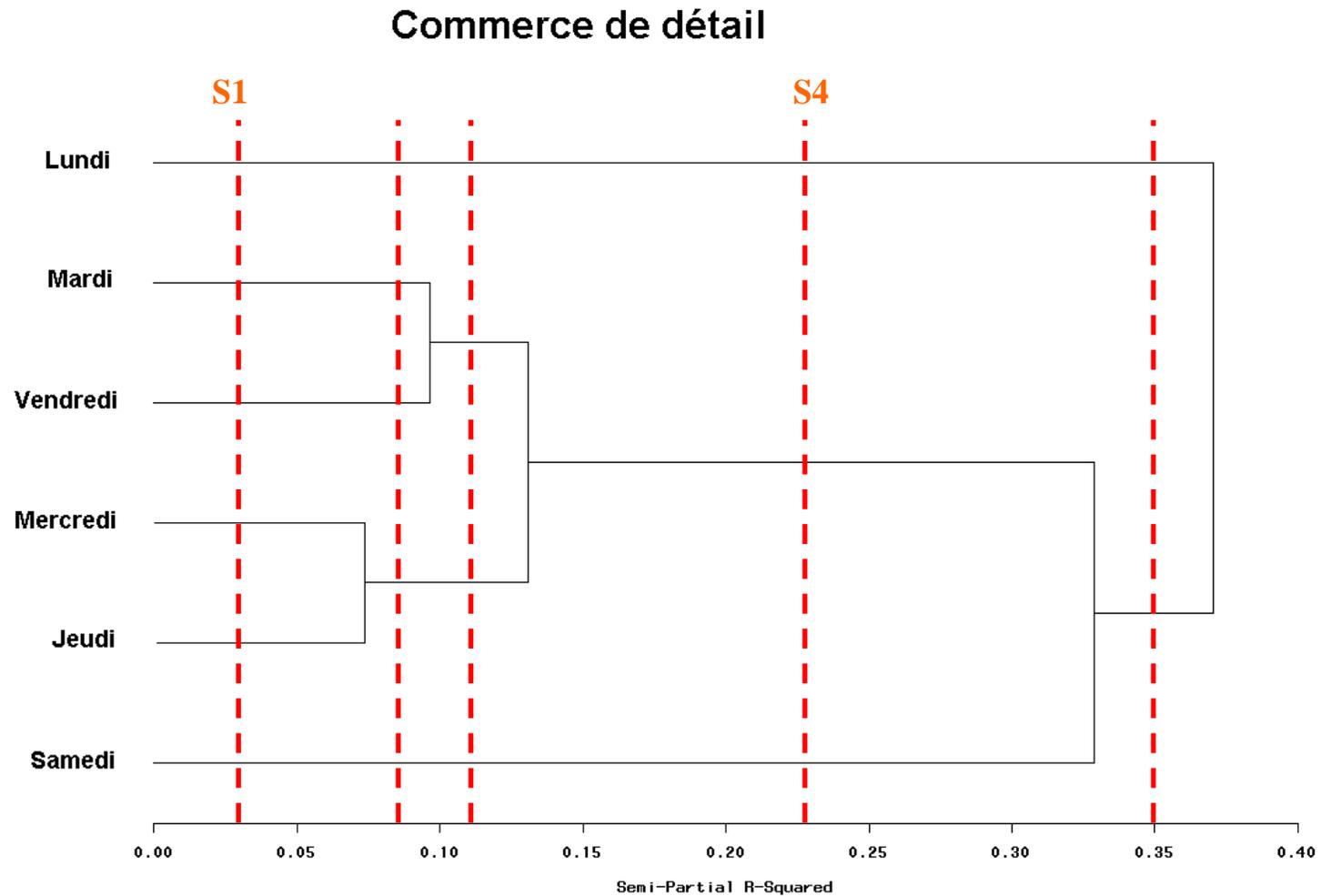
Individus = types de jours (6).

Arbre de classification des coeffs CJO pour les indices de CA dans 57 secteurs du commerce détail

Commerce de détail



Arbre de classification des coeffs CJO pour les indices de CA dans 57 secteurs du commerce détail



Modélisation des effets de calendrier (1)

Modèle REG-ARIMA de base pour des effets fixes de jours ouvrables :

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

X_t : série brute,

N_{it} : Nombre de lundis ($i=1$), Mardis ($i=2$) ... à la date t ,

ε_t suit un modèle ARIMA.

α_i : effet moyen du i -ème jour de la semaine sur la variable X .

Modélisation des effets de jours ouvrables (2)

Afin de résoudre des problèmes de colinéarité et d'éliminer la saisonnalité dans les régresseurs, le modèle est transformé de la façon suivante (voir Bell, Hilmmer (1983) et Bell (1995)) :

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

$$\text{Avec : } \bar{\alpha} = \frac{1}{7} \sum_{i=1}^7 \alpha_i \quad \text{et} \quad \beta_i = \alpha_i - \bar{\alpha}$$

$\bar{\alpha}$ est l'effet moyen d'un jour quelconque

β_i est l'effet spécifique d'un jour de type i et $\sum_{i=1}^7 \beta_i = 0$

$$LY_t = \begin{cases} 0.7575 & \text{si c'est un mois de février à 29 jours} \\ -0.2425 & \text{si c'est un mois de février à 28 jours} \\ 0 & \text{sinon} \end{cases}$$

Un modèle plus parcimonieux dans les logiciels de désaisonnalisation

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

Contraste entre jours de semaine et week-end (samedi+dimanche) :

$$\beta_6 = \beta_7 \text{ and } \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$$

$$\sum_{i=1}^7 \beta_i = 0 \Rightarrow \beta_6 = \beta_7 = -\frac{5}{2} \beta_1$$

Si bien que...

$$X_t = \beta_1 \left[\sum_{i=1}^5 N_{it} - \frac{5}{2} (N_{6t} + N_{7t}) \right] + \bar{\alpha}LY_t + \varepsilon_t$$

L'insertion du calendrier national dans la modélisation

L'effet d'un jour de type i sur la variable X peut être différent si ce jour est férié ou s'il ne l'est pas.

L'introduction du calendrier national conduit à un modèle à 15 variables :

$$X_t = \sum_{i=1}^{14} \beta_i N_{it} + \bar{\alpha} LY_t + \varepsilon_t$$

Où $\bar{\alpha} = \frac{1}{14} \sum_{i=1}^{14} \alpha_i$ et $\beta_i = \alpha_i - \bar{\alpha}$

N_{it} : Nombre de lundis non fériés ($i=1$), mardis non fériés ($i=2$), ..., dimanches non fériés ($i=7$), lundis fériés ($i=8$), mardis fériés ($i=9$), ..., dimanches fériés ($i=14$) à la date t .

Exemple d'un modèle plus parcimonieux

Hypothèses :

Impact équivalent des dimanches travaillés et fériés

Impact de n'importe quel jour férié équivalent à celui d'un dimanche

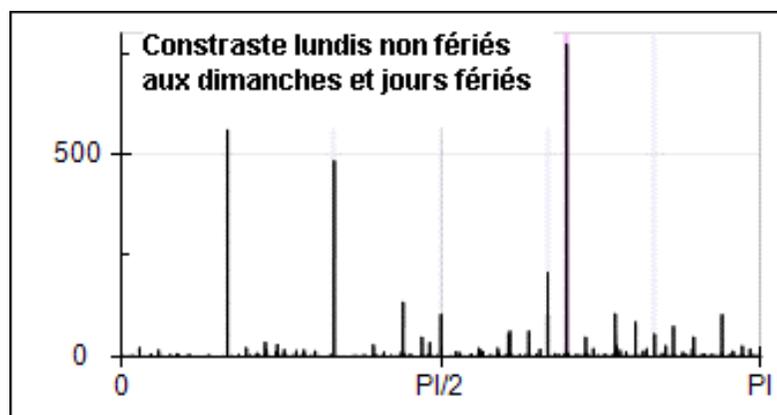
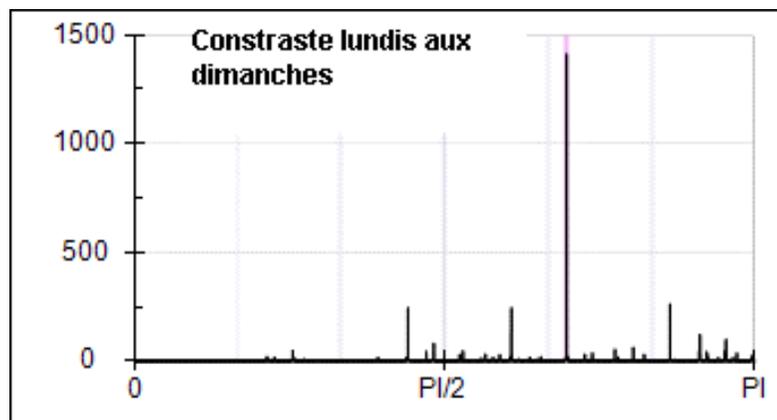
$$\beta_7 = \beta_8 = \beta_9 = \dots = \beta_{13} = \beta_{14}$$

Et un modèle "simplifié" ...

$$X_t = \sum_{i=1}^6 \beta_i \left(N_{it} - \frac{1}{8} \sum_{i=7}^{14} N_{it} \right) + \bar{\alpha}LY_t + \varepsilon_t$$

!!! Le contraste ne supprime plus la composante saisonnière (Moyennes de long terme des deux membres du contraste distinctes) → Il faut supprimer la composante saisonnière des régresseurs

Périodogrammes des contrastes sans et avec prise en compte des jours fériés



Application : Tests de divers jeux de régresseurs pour les indices de chiffres d'affaires

S0: LY_t seul (tous les jours de la semaine similaires)

S1: 6 variables: Lu, Ma, Me, Je, Ve, Sa *non fériés*; Contraste: Di + Jours *fériés*

S2: 1 variable: (Lu+Ma+Me+Je+Ve) *non fériés*; Contraste: Sa+Di+ Jours *fériés*

S3: 5 variables: Lu, Ma, Me, Je, Ve *non fériés*; Contraste: Sa+Di + Jours *fériés*

S4: 3 variables: Lu, (Ma+Me+Je+Ve), Sa *non fériés*; Contraste: Di+Jours *fériés*

La variable LY_t (effet année bissextile) toujours présente également dans les modèles avec S1, S2, S3, S4.

Comment choisir parmi les différents jeux de régresseurs?

Régression Reg-ARIMA appliquée avec le jeu le plus complet de régresseurs (S1). Modèles ARIMA et outliers identifiés automatiquement avec S1 sont ensuite utilisés pour l'estimation des coefficients chaque jeu de régresseurs à tester.

Sélection du jeu le plus pertinent de régresseurs grâce à un algorithme fondé sur deux critères :

Test de Fisher d'égalité des coefficients → confirmer ou infirmer l'équivalence des effets des divers jours de la semaine

Critère AICC → info sur la qualité du modèle Reg-ARIMA pour chaque jeu de régresseurs testé : on réduit le nombre de régresseurs à condition que l'AICC soit plus faible.

Résultats : Distribution (en %) des jeux de régresseurs sélectionnés par « grands » secteurs d'activité

	Pas d'effet JO	S0 LY_t seul	S1	S2	S3	S4
Commerce de détail (57)	5	2	21	4	7	61
Commerce de gros (89)	17	4	8	27	16	28
Services (138)	26	1	7	32	14	19
Industrie (122)	15	0	14	47	25	0 (non testé dans l'industrie)

Qualité de la procédure de sélection des régresseurs ?

% de séries d'indices de chiffres d'affaire du commerce de détail présentant des pics spectraux résiduels aux fréquences de JO (diagnostic de Demetra+):

Pas de régresseur JO	S0	S1	S2	S3	S4	Jeu de régresseurs retenu
61 %	67 %	9 %	39 %	14 %	16 %	12 %

→ Une procédure satisfaisante...