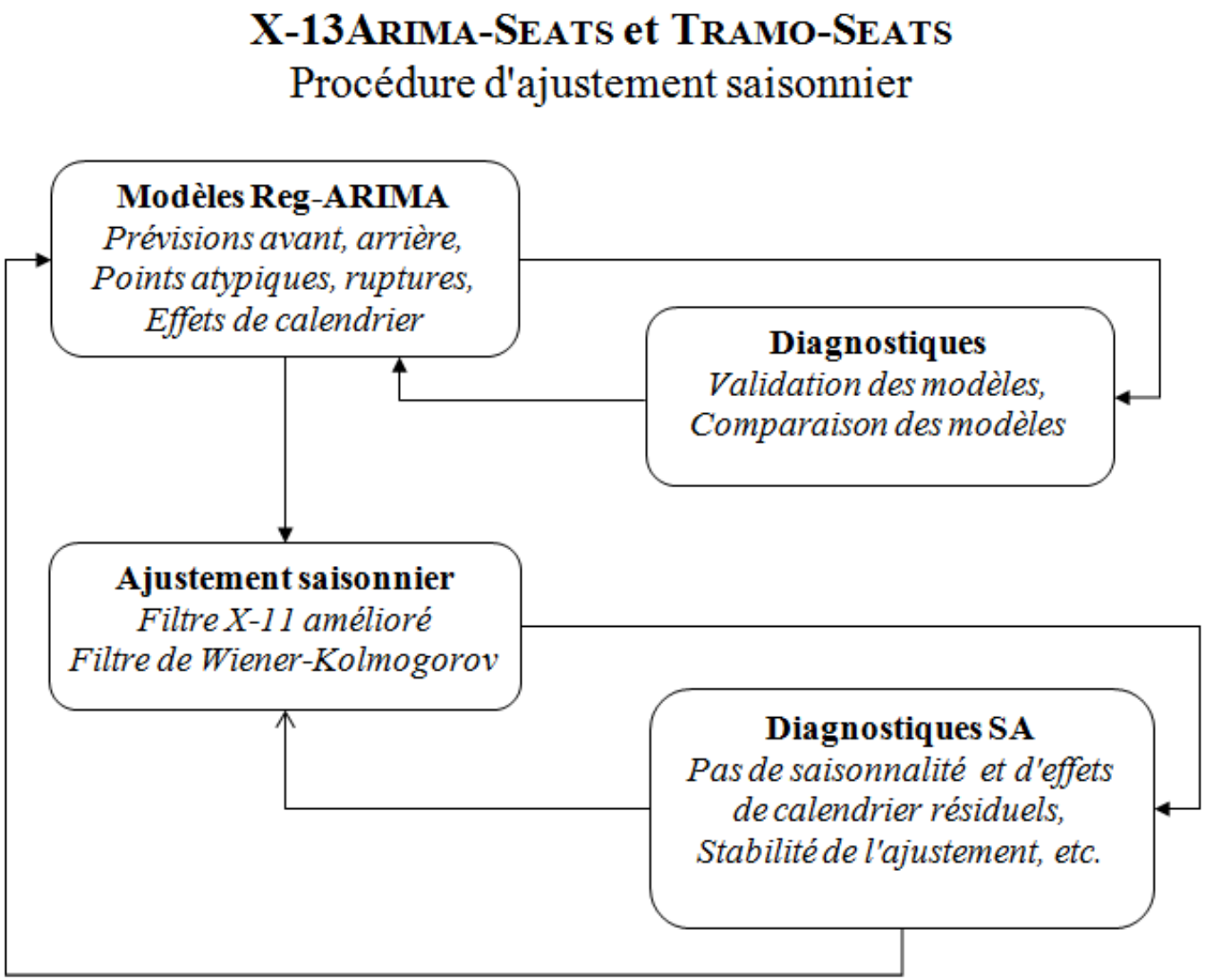
Journées de méthodologie statistique (JMS) - 2018

Du bon usage des modèles Reg-ARIMA en désaisonnalisation

Dominique Ladiray et Alain Quartier-la-Tente
Département des Méthodes Statistiques
Insee, Seasonal Adjustment Centre of Excellence (SACE)

Introduction à la procédure de désaisonnalisation



Écriture mathématique du Reg-ARIMA

Écriture mathématique du modèle Reg-ARIMA en désaisonnalisation :

$$\left.\begin{array}{ll}\text{Additif :} & Y_t \\ \text{Multiplicatif :} & \log(Y_t)\end{array}\right\} = \underbrace{\beta_0 LY_t + \beta_1 WD_t}_{\text{Régresseurs JO}} + \underbrace{\sum_i \gamma_i O_{i,t}}_{\text{Ruptures}} + \underbrace{\varepsilon_t}_{\sim ARIMA}$$

Objectif de l'étude : illustrer des problèmes d'instabilité des estimations avec des exemples sur :

- la correction de l'effet année bissextile (*leap year*)
- l'estimation de ruptures (*outliers*)
- l'identification du modèle ARIMA

Sommaire

1. Correction de l'effet année bissextile

1.1 Quand et comment corriger l'effet année bissextile ?

1.2 Méthodologie de l'étude

1.3 Exemples

1.4 Résultats des simulations

2. Correction des ruptures

3. Identification du modèle ARIMA

4. Conclusion et recommandations

Quand faut-il le corriger ?

Année bissextile (*leap year*) : un jour en plus en février $\simeq$ 4 ans

$\rightarrow$ prise en compte de l'effet « longueur du mois » : c'est un effet de calendrier

Quand le corriger ?

D'après les *guidelines* sur l'ajustement saisonnier, le faire lorsque :

- il y a un sens économique à le faire
- l'effet est stable et statistiquement significatif

Étude des IPI européens (1330 séries) : l'effet *leap year* existe (mais pas toujours mesurable du fait de la collecte)

Deux méthodes pour le corriger

1. Avec le modèle Reg-ARIMA

$$LY_t = \begin{cases} 0,75 & \text{si } t \text{ est un mois de février bissextil} \\ -0,25 & \text{si } t \text{ est un mois de février non bissextil} \\ 0 & \text{sinon} \end{cases}$$

2. Avec une correction *a priori* en multipliant la série initiale par :

$$\begin{cases} \frac{28,25}{29} \simeq 0,974 & \text{mois de février bissextil} \\ \frac{28,25}{28} \simeq 1,009 & \text{mois de février non bissextil} \\ 1 & \text{sinon} \end{cases}$$

[Bell, 1992] : les deux méthodes équivalentes si schéma multiplicatif et valeur estimée proche de 0,035 ($\simeq \frac{29}{28} - 1$, valeur attendue)

$\rightarrow$ Étude des estimations de la 1re méthode

Méthodologie utilisée

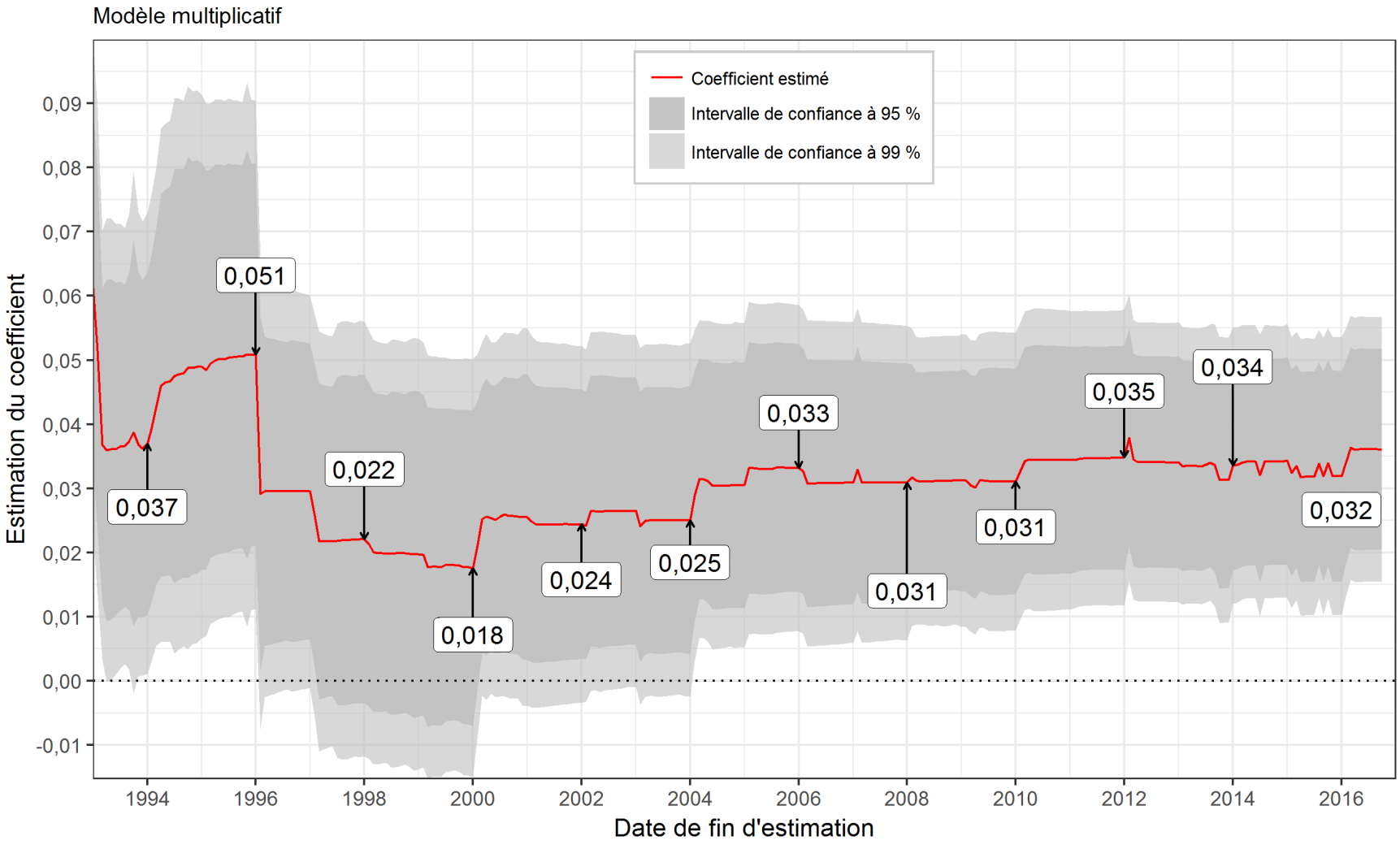
Méthodologie : modèle **identifié** sur l'ensemble de la période (ARIMA, outliers, etc.) et estimation mois par mois sur le passé en **figeant** la date de début d'estimation

On considère qu'il y a convergence lorsque le coefficient estimé reste :

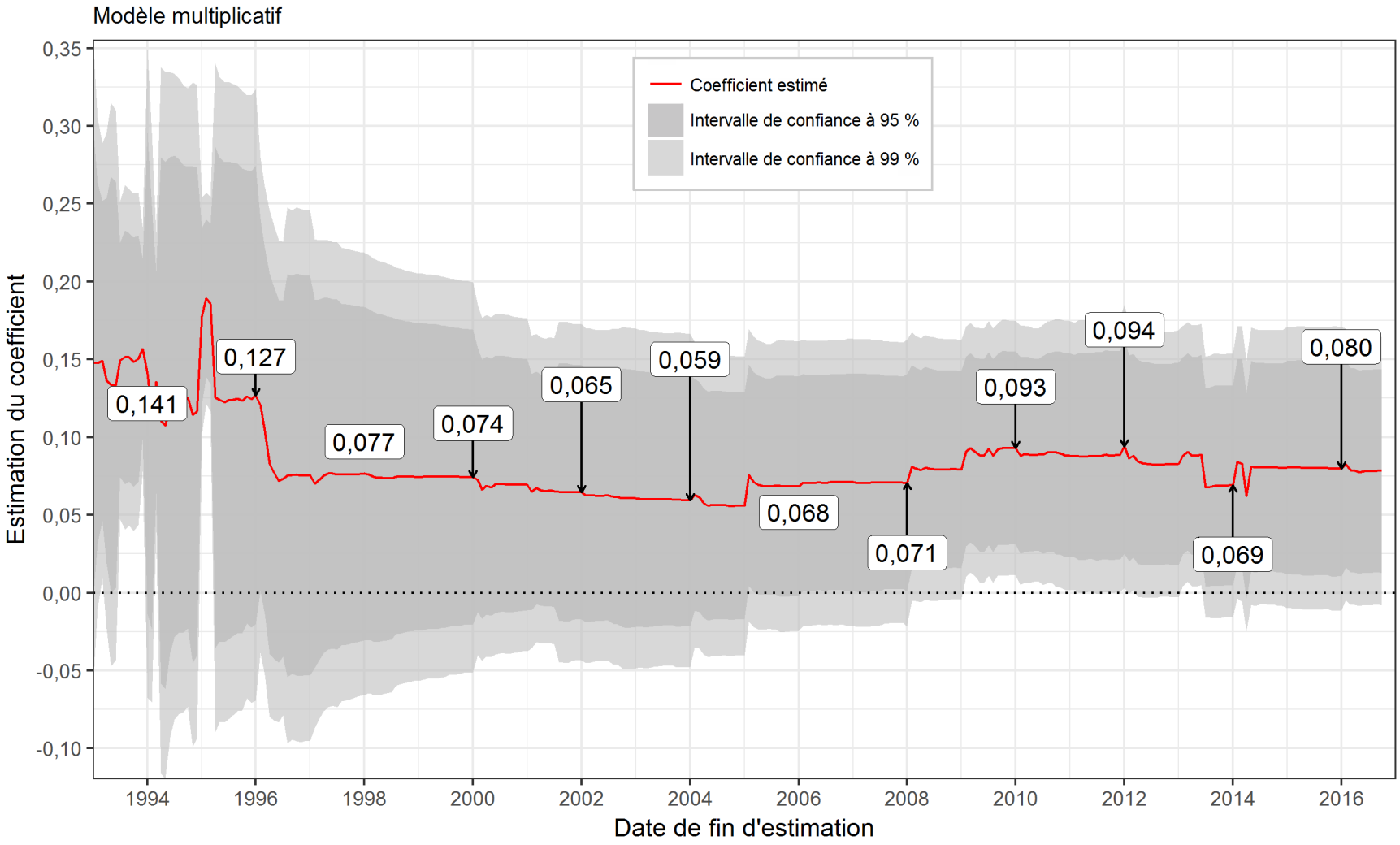
- positif
- non significativement différent de la dernière valeur
- significative : stabilité du choix de corriger

$\Longrightarrow$ IPI européen : 410 séries convergent

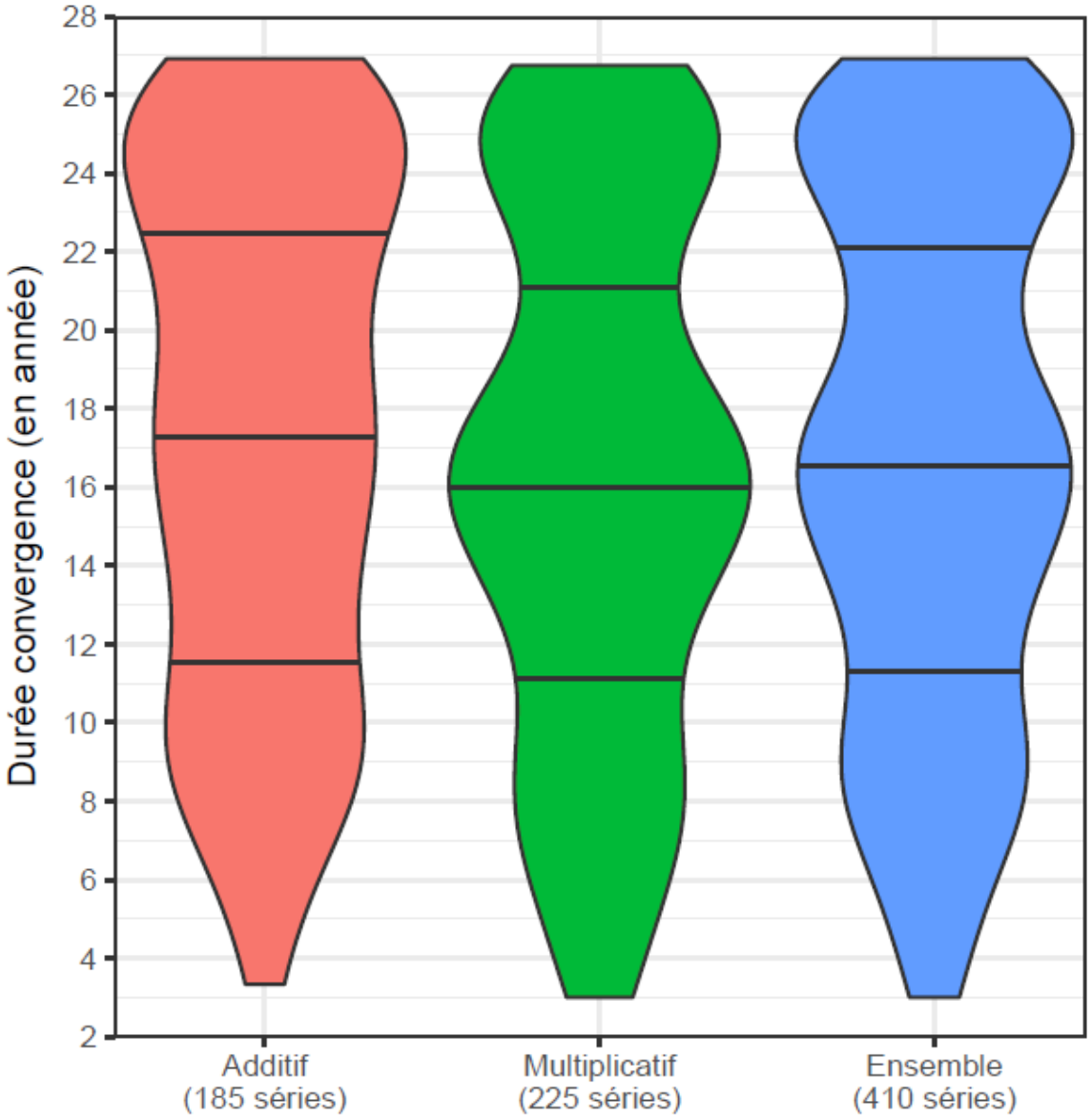
Exemples (1/2) : IPI FR-0610 (extraction de pétrole brut)



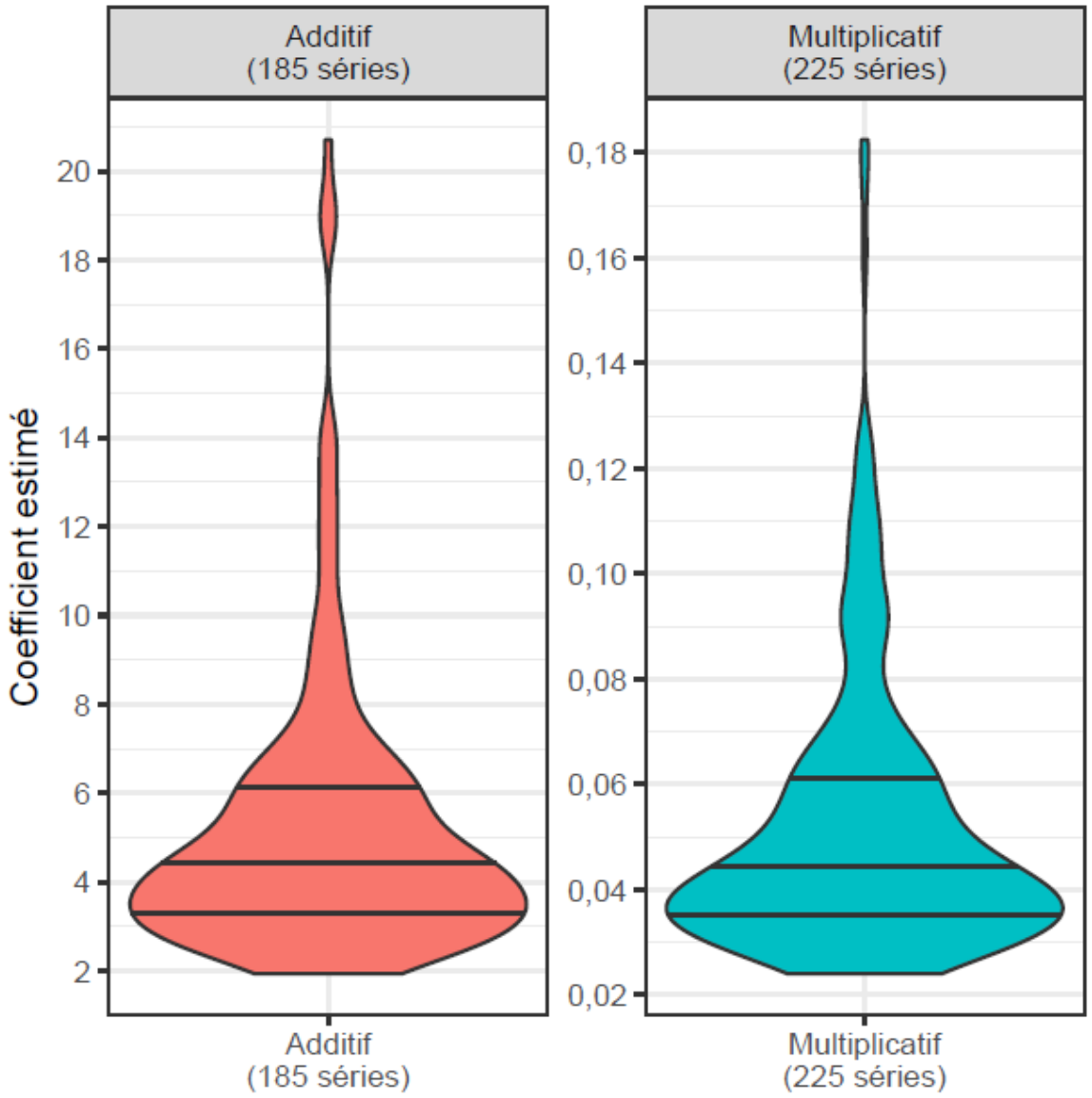
Exemples (2/2) : IPI FR-1391 (Fabrication d'étoffes à mailles)



Une convergence plutôt lente. . .



... Vers une valeur pas toujours cohérente



Comparaison des deux méthodes de correction

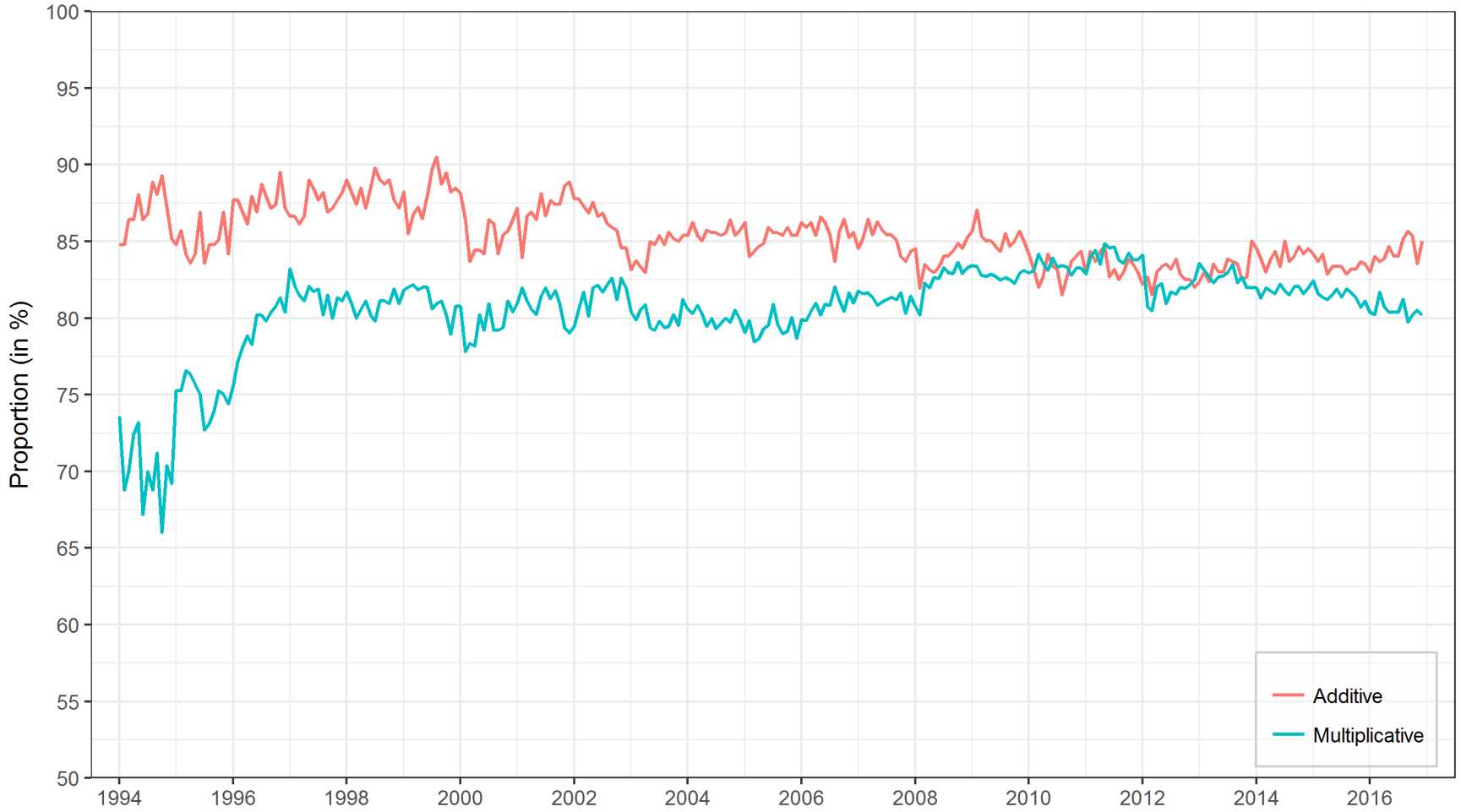


FIGURE 1 – Pourcentage des séries pour lesquelles l'AICC de la méthode 2 (pré-ajustement du LY) est inférieur à l'AICC méthode 1 (régresseur LY)

Sommaire

1. Correction de l'effet année bissextile

2. Correction des ruptures

2.1 Les différentes ruptures étudiées

2.2 Méthodologie de l'étude

2.3 Exemple

2.4 Résultats des simulations

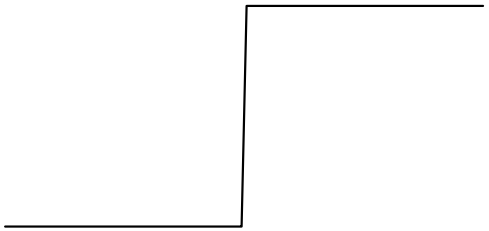
3. Identification du modèle ARIMA

4. Conclusion et recommandations

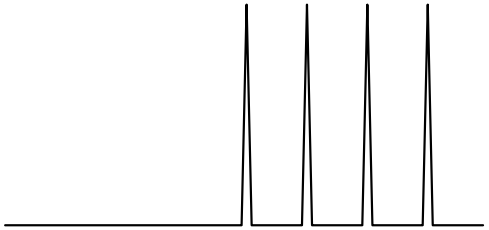
Les principaux types d'*outliers*

Point atypique
Additive outlier (AO)

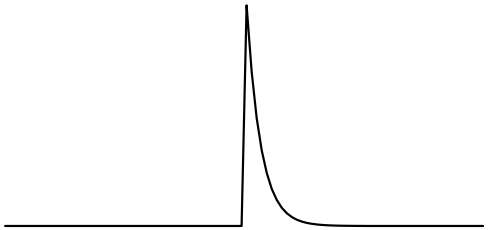
Changement de niveau
Level Shift (LS)



Rupture dans la composante saisonnière
Seasonal Outlier (SO)



Changement transitoire de niveau
Transitory Change (TC)



Méthodologie utilisée

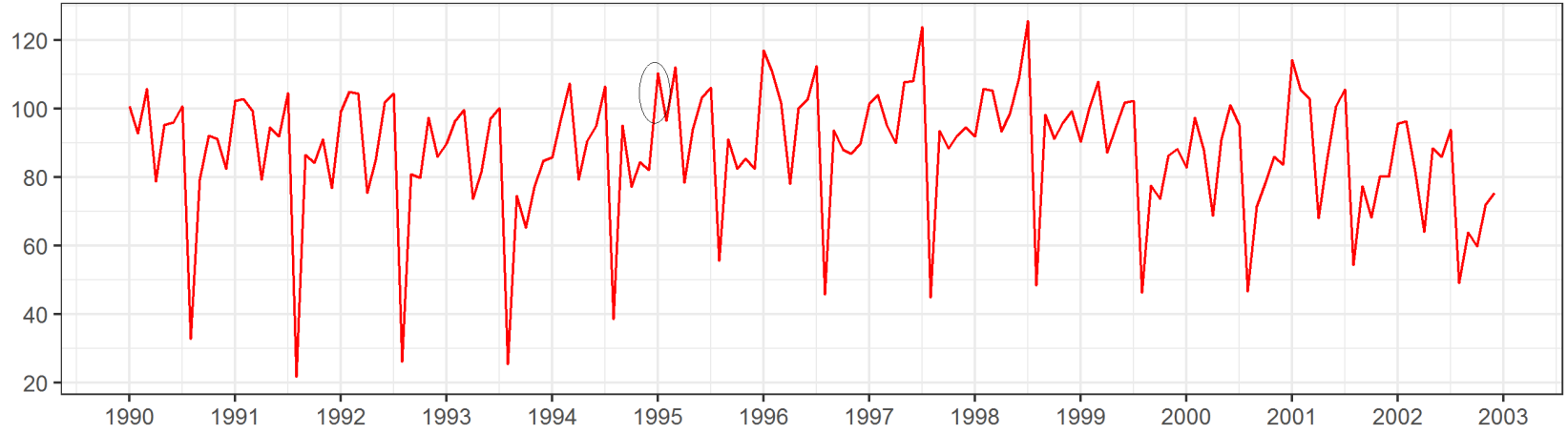
Sur les IPI européens :

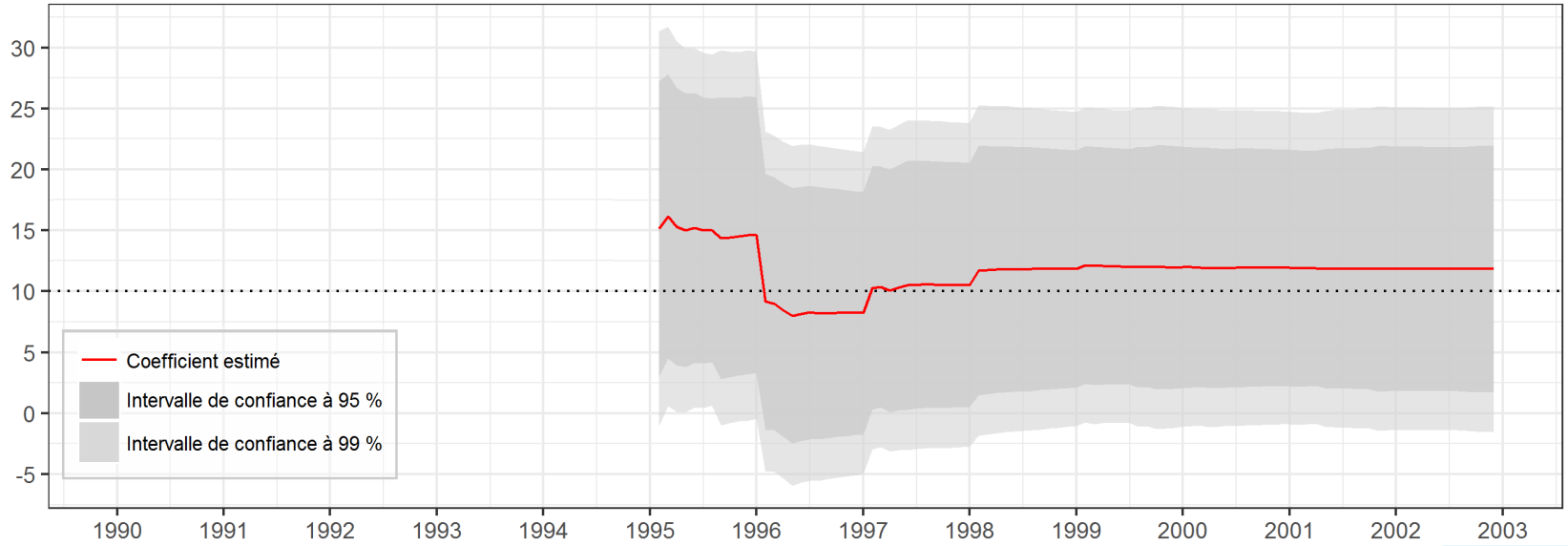
1. **identification** et **estimation** du modèle sur 13 ans
2. simulation d'une rupture 5 ans après la date de début de niveau 10 pour un modèle **additif**
3. estimation du coefficient de la rupture en figeant les estimations de tous les autres paramètres et la date de début d'estimation

On considère qu'il y a convergence lorsque :

$$\left| \frac{\text{valeur estimée}}{\text{dernière valeur estimée}} - 1 \right| < 5\ \%$$

Exemple d'un AO pour la série IPI IT-1413 (Fabrication de vêtements de dessus)





Une convergence plutôt lente. . .

... Mais pas toujours vers la bonne valeur

	Minimum	25 %	50 %	75 %	Maximum
Modèles additifs					
Additive outlier (AO)	-73,4	8,0	11,0	14,4	45,9
Level Shift (LS)	-33,9	6,3	9,2	12,8	95,0
Seasonal outlier (SO)	-83,5	5,9	8,1	10,4	34,3
Transitory Change (TC)	-56,9	6,9	10,2	14,2	133,5

Sommaire

1. Correction de l'effet année bissextile

2. Correction des ruptures

3. Identification du modèle ARIMA

4. Conclusion et recommandations

Identification de deux modèles équivalents

On reprend le même modèle de base sous deux formes différentes mathématiquement équivalentes :

1. Le régresseur *leap year* intégré dans le système des régresseurs jours ouvrables
2. Le régresseur *leap year* introduit comme régresseur externe

$\rightarrow$ étude du modèle automatique

Modèles automatiques différents

Le régresseur LY est dans les effets de calendrier

Summary

Estimation span: [1-1990 - 11-2016]
323 observations
Trading days effects (7 variables)
3 detected outliers

Arima model
[(2,0,0)(0,1,1)]

	Coefficients	T-Stat	P[|T| > t]
Phi(1)	-0,5256	-9,46	0,0000
Phi(2)	-0,2878	-5,17	0,0000
BTheta(1)	-0,7913	-20,56	0,0000

Correlation of the estimates

	Phi(1)	Phi(2)	BTheta(1)
Phi(1)	1,0000	-0,7388	-0,0184
Phi(2)	-0,7388	1,0000	0,0489
BTheta(1)	-0,0184	0,0489	1,0000

Leap year

	Coefficients	T-Stat	P[|T| > t]
Leap year	4,3861	2,65	0,0085

Le régresseur LY est dans les régresseurs externes

Summary

Estimation span: [1-1990 - 11-2016]
323 observations
No trading days effects
8 detected outliers

Arima model
[(0,1,1)(0,1,1)]

	Coefficients	T-Stat	P[|T| > t]
Theta(1)	-0,5051	-10,08	0,0000
BTheta(1)	-0,7533	-18,80	0,0000

Correlation of the estimates

	Theta(1)	BTheta(1)
Theta(1)	1,0000	0,0280
BTheta(1)	0,0280	1,0000

User variables

	Coefficients	T-Stat	P[|T| > t]
Leap year	4,5569	2,92	0,0038

Sommaire

1. Correction de l'effet année bissextile

2. Correction des ruptures

3. Identification du modèle ARIMA

4. Conclusion et recommandations

Conclusion et recommandations (1/2)

Simulations **critiquables** et **améliorables** mais mettent en évidence la **potentielle instabilité** des modèles Reg-ARIMA souvent utilisés comme boîtes noires

Instabilités ont un **effet limité** sur la CVS-CJO... mais ont un effet sur l'histoire à **court terme** et sur les **révisions** !

Algorithmes automatiques des méthodes X-13ARIMA-SEATS et TRAMO-SEATS très **importants** et très **utiles**

Conclusion et recommandations (2/2)

Spécifier le modèle **au préalable** au niveau de la **série** :

- baser les procédures de choix en s'appuyant sur un raisonnement d'abord économique (attention aux séries trop longues)
- ne pas utiliser les méthodes comme des boîtes noires... Sinon, vous serez comme ce statisticien qui...

Merci de votre attention

> *« Il se sert des statistiques comme un ivrogne d'un réverbère : pour se soutenir et non pour s'éclairer. »*

Citation largement attribuée à Andrew Lang (1844-1912)