



## 7 - Problèmes d'estimation du modèle Reg-ARIMA

ALAIN QUARTIER-LA-TENTE  
Lemna, Insee

1. Problèmes liés à la qualité des résidus
2. Modèle regARIMA et séries longues
3. Conclusion

## Questions de positionnement

---

Quelles sont les hypothèses sur les résidus du modèle regARIMA et sont-elles importantes ?

Avoir plus de données permet-il d'améliorer l'estimation du modèle regARIMA ?

Quelle est la longueur optimale pour l'estimation d'un modèle ?

# Sommaire

---

## 1. Problèmes liés à la qualité des résidus

1.1 Les hypothèses sur les résidus du modèle regARIMA

1.2 Exemple sur une série de l'IPI

## 2. Modèle regARIMA et séries longues

## 3. Conclusion

# Les hypothèses sur les résidus

---

Les estimations du modèle regARIMA sont **consistantes** (convergent et sans biais) et **efficaces** (de variance minimale), si les résidus  $\varepsilon_t$  sont :

- *décorrélés* :  $\forall t \neq t' : \text{Cov}(\varepsilon_t, \varepsilon_{t'}) = 0$  (*autocorrélés* sinon)
- *homoscédastiques* :  $\forall t, t' : \mathbb{V}[\varepsilon_t] = \mathbb{V}[\varepsilon_{t'}]$  (*hétéroscédastiques* sinon)
- distribuées selon une *loi normale*

# Conséquences des problèmes sur les résidus

Problème sur les résidus	Estimation des coefficients	Estimation de la variance des coefficients (pour tests de significativité)	Tests invalidés
Autocorrélation	biaisée	biaisée	<ul style="list-style-type: none"> <li>- Student</li> <li>- Hétéroscédasticité</li> <li>- Normalité</li> <li>- Qualité des prévisions</li> </ul>
Heteroscédasticité	non biaisée	biaisée	<ul style="list-style-type: none"> <li>- Student</li> <li>- Normalité</li> <li>- Qualité des prévisions</li> </ul>
Normalité	non biaisée	non biaisée	<ul style="list-style-type: none"> <li>- Student</li> <li>- Qualité des prévisions</li> </ul>

## Sources des problèmes de spécification des résidus

---

L'autocorrélation des résidus peut provenir :

- d'erreurs de mesure : si les données sont interpolées toujours à la même date, un biais systématique peut être observé
- problème de variable omise : il manque une variable explicative importante
- mauvaise spécification : par exemple dans le cas d'une équation non linéaire mais polynomiale
- effet de l'habitude : biais systématique du fait de l'optimisme
- d'un lissage artificiel des données trimestrielles sur données annuelles



L'autocorrélation peut souvent être corrigé en augmentant l'ordre AR mais elle peut aussi provenir de la présence de certains points atypiques (LS, SO)

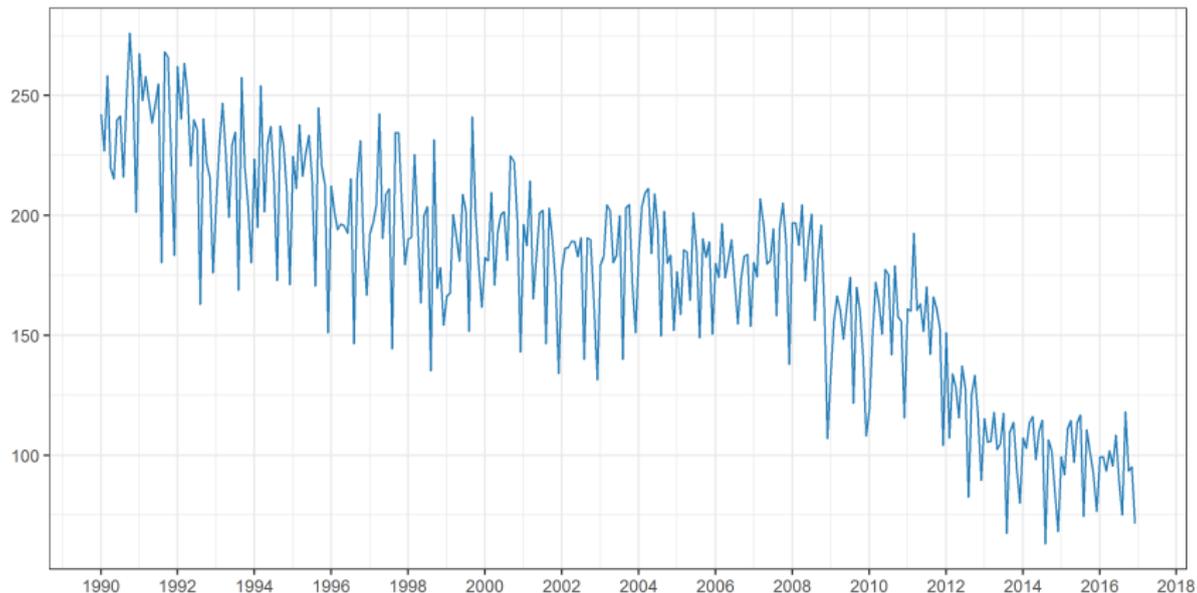


Les points atypiques affectent l'hétéroscédasticité (AO, TC) et la non-normalité

## Exemple (1/4)

Un modèle  $ARIMA(0,1,1)(0,0,0)$  est-il plausible ?

IPI branche 0811 (série brute)



## Exemple (2/4)

Que peut-on dire sur les régresseurs JO ?

### Arima model

[(0,1,1)(0,0,0)].

	<b>Coefficients</b>	<b>T-Stat</b>	<b>P[ T  &gt; t]</b>
Theta(1)	-0,8417	-27,90	0,0000

### Regression model

User-defined calendar variables

	<b>Coefficients</b>	<b>T-Stat</b>	<b>P[ T  &gt; t]</b>
REG1_AC1	0,0052	0,34	0,7331
REG1_AC2	0,0224	1,48	0,1387
REG1_AC3	0,0237	1,58	0,1159
REG1_AC4	0,0200	1,25	0,2113
REG1_AC5	0,0122	0,80	0,4249
REG1_AC6	-0,0225	-1,37	0,1704
REG1_LPY	0,0473	0,76	0,4465

Joint F-Test = 2,70 (0,0098)

### Prespecified outliers

	<b>Coefficients</b>	<b>T-Stat</b>	<b>P[ T  &gt; t]</b>
AO (2-2012)	-0,2728	-1,91	0,0573
TC (12-2008)	-0,3779	-3,33	0,0010

## Exemple (3/4)

Avec un modèle bien spécifié :

### Arima model

[(2,2,1)(1,1,1)].

	Coefficients	T-Stat	P[ T  > t]
Phi(1)	0,5620	9,98	0,0000
Phi(2)	0,2396	4,29	0,0000
Theta(1)	-1,0000	-51,88	0,0000
BPhi(1)	-0,2398	-3,85	0,0001
BTheta(1)	-0,9457	-28,71	0,0000

### Correlation of the estimates

	Phi(1)	Phi(2)	Theta(1)	BPhi(1)	BTheta(1)
Phi(1)	1,0000	0,4630	-0,0181	0,1076	0,1149
Phi(2)	0,4630	1,0000	0,0365	0,1196	0,0481
Theta(1)	-0,0181	0,0365	1,0000	0,0275	0,3949
BPhi(1)	0,1076	0,1196	0,0275	1,0000	-0,0921
BTheta(1)	0,1149	0,0481	0,3949	-0,0921	1,0000

### Regression model

User-defined calendar variables:

	Coefficients	T-Stat	P[ T  > t]
REG1_AC1	0,0182	3,22	0,0014
REG1_AC2	0,0211	3,76	0,0002
REG1_AC3	0,0246	4,28	0,0000
REG1_AC4	0,0316	5,12	0,0000
REG1_AC5	0,0108	1,89	0,0594
REG1_AC6	-0,0177	-2,94	0,0035
REG1_LPY	0,0450	1,95	0,0523

## Exemple (4/4)

Quel impact sur ma série désaisonnalisée ?

IPI branche 0811 (séries CVS)



# Sommaire

---

## 1. Problèmes liés à la qualité des résidus

## 2. Modèle regARIMA et séries longues

### 2.1 Hypothèse de stabilité des coefficients

### 2.2 Exemples

## 3. Conclusion

# Hypothèse de stabilité du modèle regARIMA

---

Le modèle regARIMA est un modèle de régression linéaire. Il suppose :

- que les coefficients sont stables dans le temps
- que la structure des résidus (modèle ARIMA) est constante dans le temps

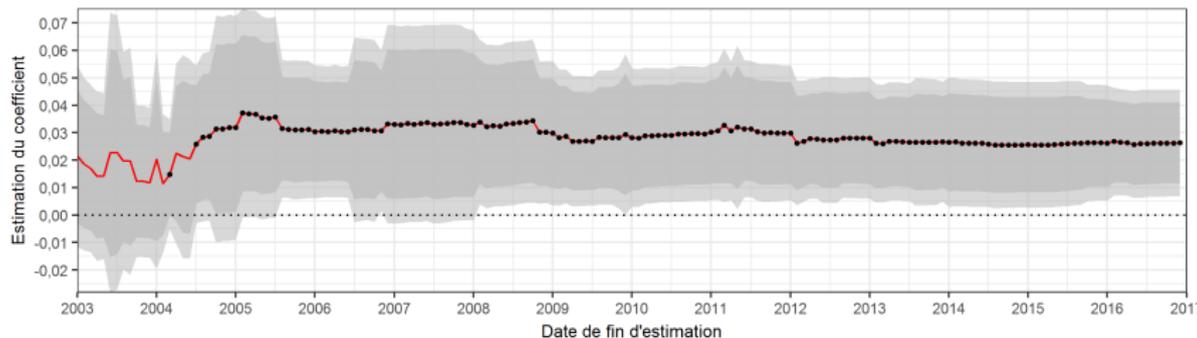


Est-ce plausible ? Quelle est la durée nécessaire pour avoir une estimation *stable* ?

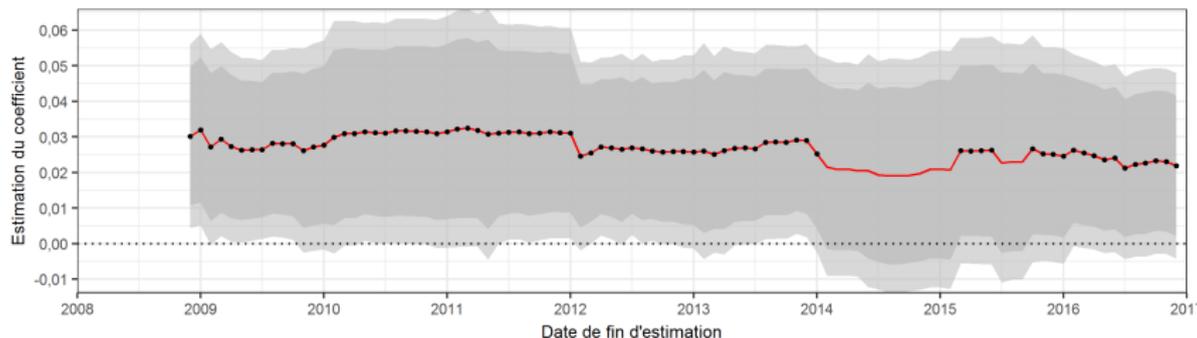
➡ Étudions les estimations des régresseurs JO et la stabilité de la décision de CJO

# Ici estimation du leap year relativement stable...

Estimation du régresseur leap year (FR-G452), date de début fixée

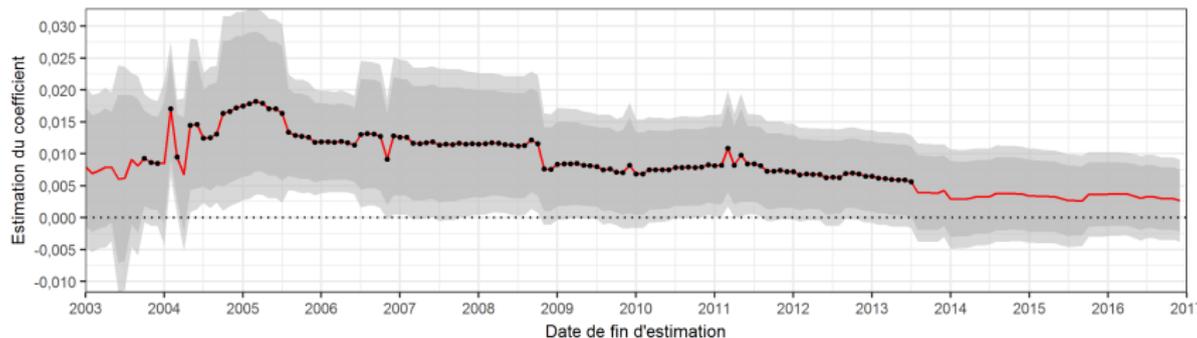


Estimation roulante sur 10 ans du régresseur leap year (FR-G452)

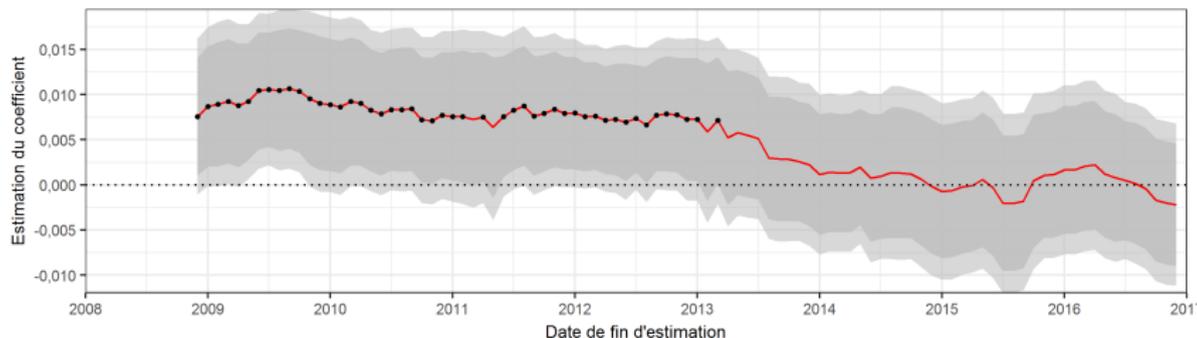


# ... Mais pas le régresseur mercredi

Estimation du régresseur mercredi (FR-G452), date de début fixée



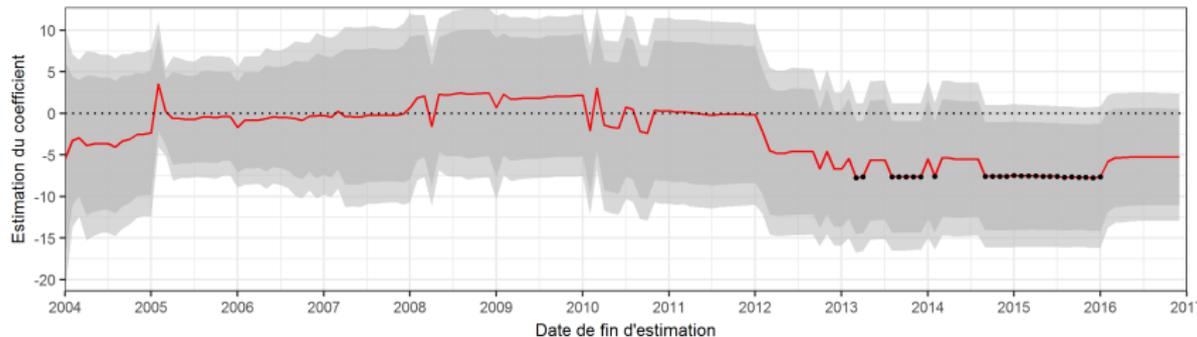
Estimation roulante sur 10 ans du régresseur mercredi (FR-G452)



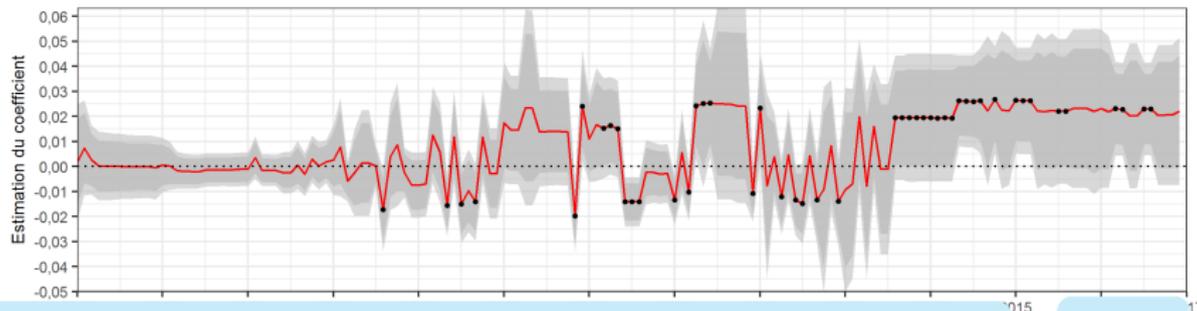
# Estimations du LY pas toujours stable

Estimations du LY toujours non significatives avec estimations roulantes

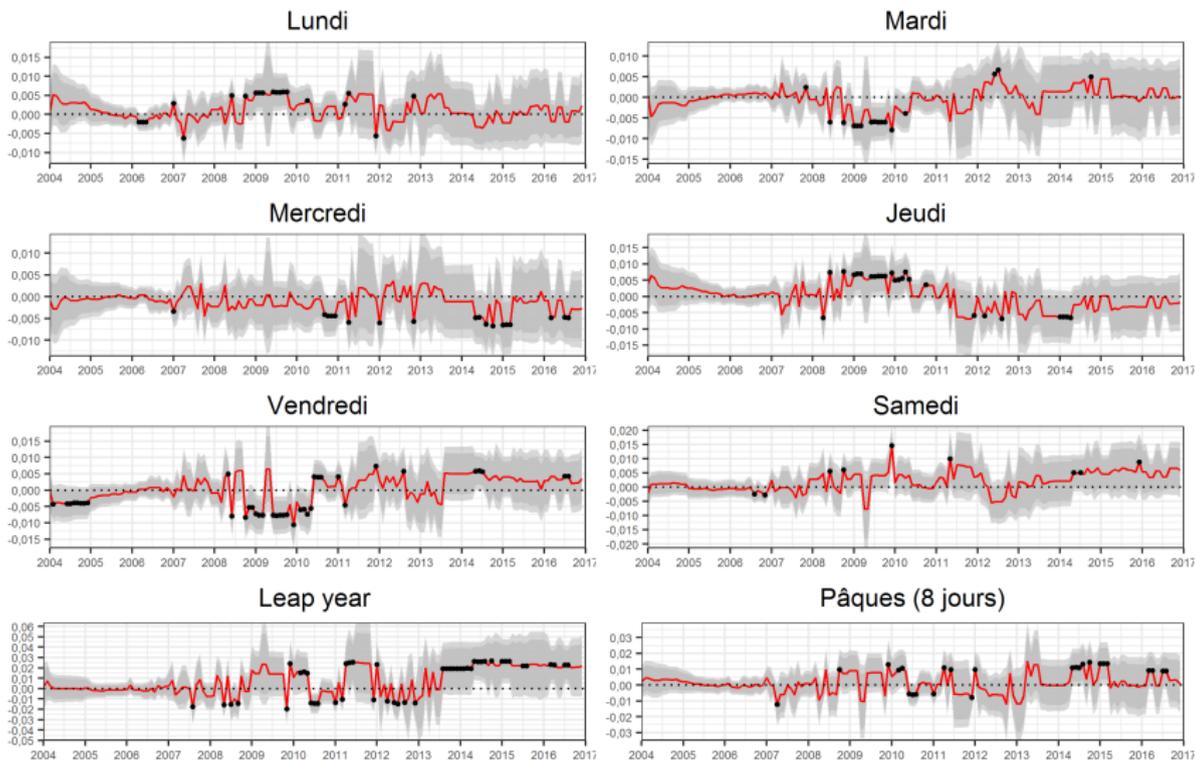
Estimation du régresseur leap year (RO-G467), date de début fixée



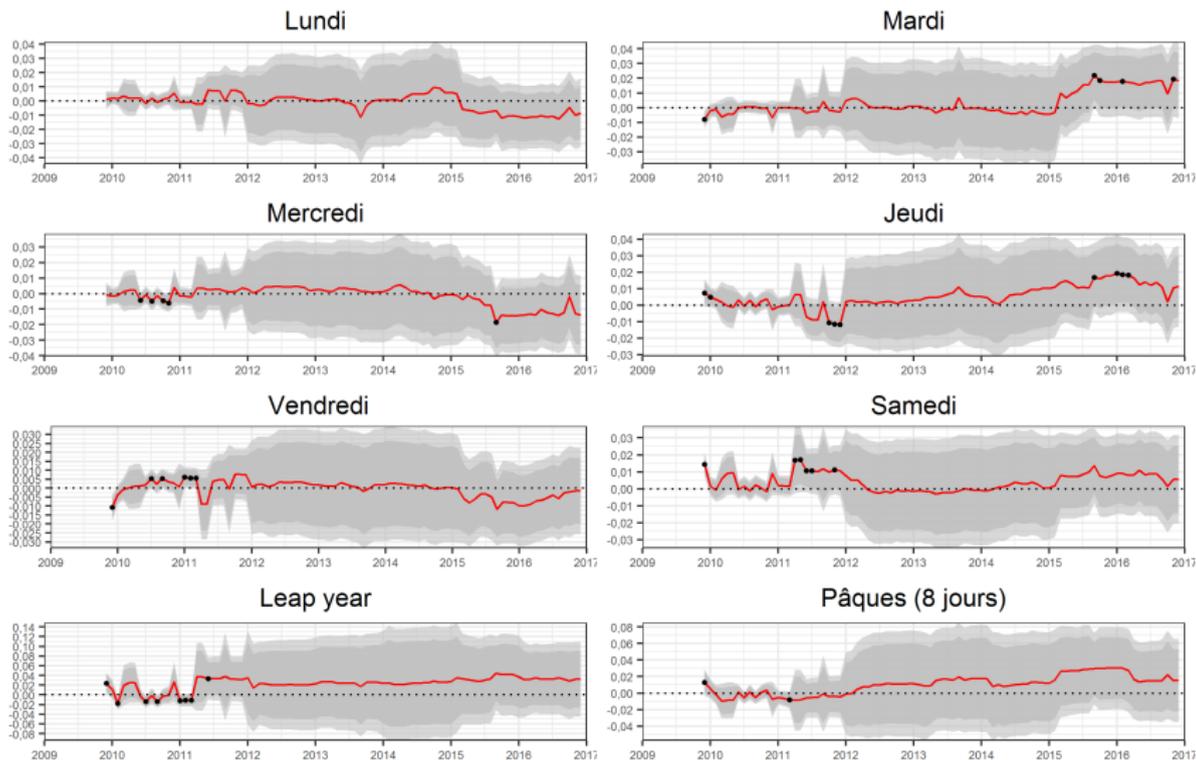
Estimation du régresseur leap year (BG-472), date de début fixée



# Cas compliqué (1/2) : EL G473



# Cas compliqué (2/2) : estimations roulantes



# Sommaire

---

1. Problèmes liés à la qualité des résidus

2. Modèle regARIMA et séries longues

**3. Conclusion**

3.1 Bibliographie

# Les essentiels

---

- une bonne spécification des résidus est importante pour interpréter le modèle regARIMA
- résidus autocorrélés  $\implies$  hétéroscédasticité  $\implies$  non normalité
- attention aux séries longues : l'hypothèse de stabilité du modèle est généralement fausse pour les séries de plus de 20 ans
- attention aux séries courtes : les estimations sont généralement moins précises (plus grande variance des estimateurs) et peuvent être fortement révisées

# Bibliographie

---



Ladiray D., Quartier-la-Tente A. (2018), Du bon usage des modèles Reg-ARIMA en désaisonnalisation, Actes des 13<sup>e</sup> Journées de Méthodologie Statistique, [http://www.jms-insee.fr/2018/S05\\_1\\_ACTEv3\\_QUARTIERLATENTE\\_JMS2018.pdf](http://www.jms-insee.fr/2018/S05_1_ACTEv3_QUARTIERLATENTE_JMS2018.pdf).