



Désaisonnaliser les séries très longues par sous-période

HIEN PHAM ET ALAIN QUARTIER-LA-TENTE
Division Recueil et Traitement de l'Information
Département des Méthodes Statistiques

Objectifs de l'article

Problème de la désaisonnalisation des séries de plus de 20 ans :

- société, institutions et comportements économiques évoluent (ex : travail le dimanche)
- un modèle, reflet d'un processus "moyen" sur 20 ans est sous-optimal
- en particulier sur le début et la fin de séries

Répondre aux questions :

1. Couper et traiter les séries par sous-période, qu'est-ce qui change ?
2. Couper à quelle date ?
3. Quelle longueur de recouvrement entre les sous-périodes ?
4. Comment raccorder les sous-périodes ?

Réponses illustrées avec les séries de l'IPI

Tests et résultats empiriques sur les 205 séries de l'IPI de niveau classes (4 chiffres)

Longueur : 27 ans, janvier 1990 - décembre 2017

Outil : JDemetra+

Désaisonnalisation en deux étapes avec X13-ARIMA

Pré-ajustement - modèle Reg-ARIMA : correction des jours ouvrables, correction des points atypiques, prévision pour prolonger la série linéarisée

Décomposition par des moyennes mobiles : module X11

Sommaire

1. Couper les séries, qu'est-ce qui change ?

2. Couper à quelle date ?

3. Quelle longueur de recouvrement ?

4. Comment raccorder ?

Couper change les résultats et améliore les modèles

Couper en 2 sous-périodes en janvier 2005, “le Passé” et “le Courant”, avec 3 ans de recouvrement de chaque côté

Comparer les modèles “données complètes” et “le Courant”, **sur les mêmes données**

Couper change les résultats et améliore les modèles

Couper en 2 sous-périodes en janvier 2005, “le Passé” et “le Courant”, avec 3 ans de recouvrement de chaque côté

Comparer les modèles “données complètes” et “le Courant”, **sur les mêmes données**

Période d'intérêt : les 6 dernières années, 2012 - 2017

Différences entre les deux versions de CVS : pour 2/3 des séries tx d'évolution commentés différents (1 pt ou plus en moyenne)

Couper change les résultats et améliore les modèles

Couper en 2 sous-périodes en janvier 2005, “le Passé” et “le Courant”, avec 3 ans de recouvrement de chaque côté

Comparer les modèles “données complètes” et “le Courant”, **sur les mêmes données**

Période d'intérêt : les 6 dernières années, 2012 - 2017

Différences entre les deux versions de CVS : pour 2/3 des séries tx d'évolution commentés différents (1 pt ou plus en moyenne)

Couper les séries améliore le pré-ajustement :

- correction de jours ouvrables plus efficace
- meilleurs résidus Reg-ARIMA : normalité, homoscédasticité et absence d'autocorrélation

En revanche, la qualité de la phase de décomposition reste inchangée

Sommaire

1. Couper les séries, qu'est-ce qui change ?

2. Couper à quelle date ?

2.1 Guidelines : couper à la fin de série

2.2 Et tester les longueurs de recouvrement

3. Quelle longueur de recouvrement ?

4. Comment raccorder ?

Recommandations sur les révisions de “fin de séries”

Lignes directrices sur l'ajustement saisonnier d'Eurostat (Guidelines)

Chapitre 4 sur les politiques de révision

Section 4.3, longueur de séries à réviser en régime courant :

- passé figé, seule la fin de série est révisée
- date de révision au plus tôt 3 ans avant la période où les données brutes restent révisables
- date doit tenir compte de l'estimation finale avec filtre symétrique au point de coupure

Fin de séries de l'IPI

Appliquer à l'IPI :

2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
										Données brutes révisables

Annotations :

- 3 ans minimum avant période de révision données brutes (point de coupure en jan 2014)
- 5 ans pour un filtre symétrique M3x9, dernier point du passé figé (point de coupure en dec 2011)

- 2017 données brutes révisables, donc CVS au plus tôt : janvier 2014
- M3x9 nécessite 5 ans de chaque côté du point central estimé
 - estimation finale décembre 2011 avec 2012 à 2016 définitives
 - couper en janvier 2012

Fin de séries pour l'IPI : janvier 2012 à décembre 2017, 6 ans

Guidelines et les séries longues

Section 6.2 traite de la gestion des séries longues :

- désaisonnaliser par sous-période
- prévoir des plages de recouvrement entre sous-périodes
- chaque sous-période doit être longue d'au moins 8 ans
- choisir la longueur sur la base des tests et examens graphiques

Date de coupure et date d'estimation

Appliquer à l'IPI :

2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Passé figé					Recouvrement pour estimer le passé					
???	Recouvrement pour estimer la fin de série				Fin de série					
↑	Date de début d'estimation ??				Date de Coupure					
			Min : 2 ans		jan 2012					
5 ans pour un filtre symétrique au point de coupure										

Date de coupure : janvier 2012, date de mise à jour des CVS avec nv_x modèles

Date de début d'estimation : janvier ???? , ? ans de recouvrement

Un recouvrement de 2 ans minimum, selon les Guidelines

Or, pour que janvier 2012 soit estimé avec un filtre symétrique il faut 5 ans de recouvrement

Sommaire

1. Couper les séries, qu'est-ce qui change ?
2. Couper à quelle date ?
- 3. Quelle longueur de recouvrement ?**
 - 3.1 Comparer les fenêtres de longueurs différentes
 - 3.2 S'appuyer sur les critères de qualité et de stabilité
 - 3.3 Algorithme de sélection de fenêtres
4. Comment raccorder ?

Les fenêtres comparées

Les fenêtres longues de 6 à 20 ans

15 fenêtres au total, avec recouvrement de 0 à 14 ans

Petites fenêtres retenues pour tester la stabilité des modèles

Un modèle automatique estimé avec JDemetra+ sur chaque fenêtre

Compare l'effet de ces modèles **sur les mêmes données**, 6 ans de la fin de série

- critères de qualité **relatifs à la fin de série**
- critères de stabilité des modèles

Les critères de qualité

6 indicateurs issus de JDematra+ résumés en 1 score :

- saisonnalité résiduelle (peu discriminant)
- jours ouvrables résiduels (pénalise fenêtres les plus courtes et les plus longues)
- qualité résidus ARIMA, normalité, homoscedasticité et indépendance (favorise fenêtres courtes)
- qualité de la prévision : test *out of sample* de moyenne des erreurs nulle (pénalise fenêtres courtes)

Les critères de qualité

6 indicateurs issus de JDematra+ résumés en 1 score :

- saisonnalité résiduelle (peu discriminant)
- jours ouvrables résiduels (pénalise fenêtres les plus courtes et les plus longues)
- qualité résidus ARIMA, normalité, homoscedasticité et indépendance (favorise fenêtres courtes)
- qualité de la prévision : test *out of sample* de moyenne des erreurs nulle (pénalise fenêtres courtes)

Valeurs des indicateurs 3 (Bad), 4 (Severe) et 0 sinon

Poids de 15, sauf normalité et homoscedasticité, poids 5.

Pour chaque série, on détermine une plage de score minimal : au moins 3 fenêtres contiguës

Idée : fenêtres proches avec modèles de qualité équivalente - ajouter une année (2018) ne devrait pas modifier cette appréciation

Les critères de stabilité

4 paramètres du modèle :

- le schéma de composition
- l'ordre de différenciation
- l'ordre AR
- l'ordre MA

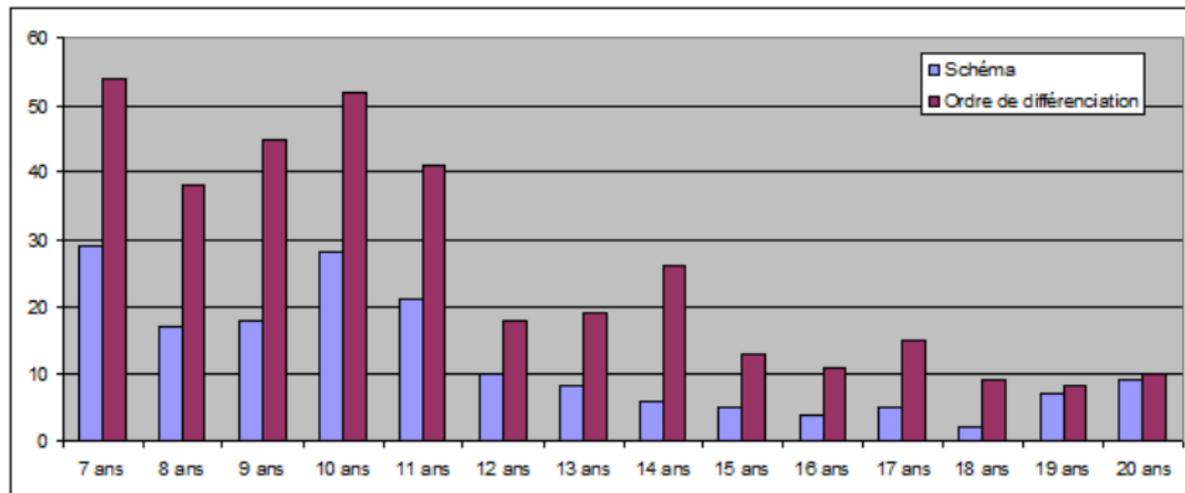
Plages de stabilité : au moins 3 fenêtres contiguës

Sinon on libère la contrainte sur les ordres AR et MA

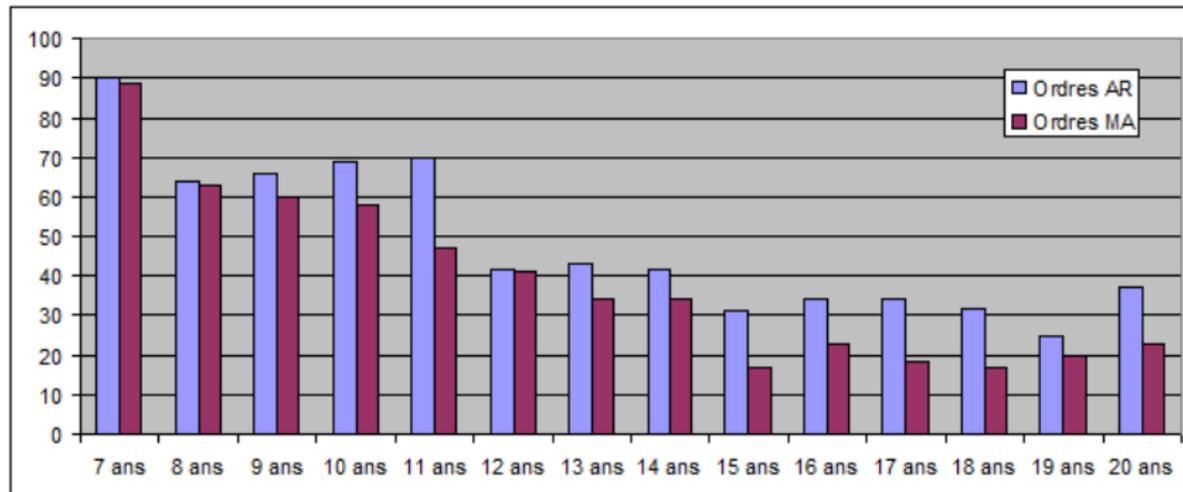
Idée : fenêtres proches avec mêmes paramètres, dynamique de la série bien approximée et modèle proche des données.

Changements de paramètres plus fréquents sur les fenêtres courtes 6 à 10 ans

Stabilité du schéma et ordre de différenciation



Stabilité des ordres AR et MA

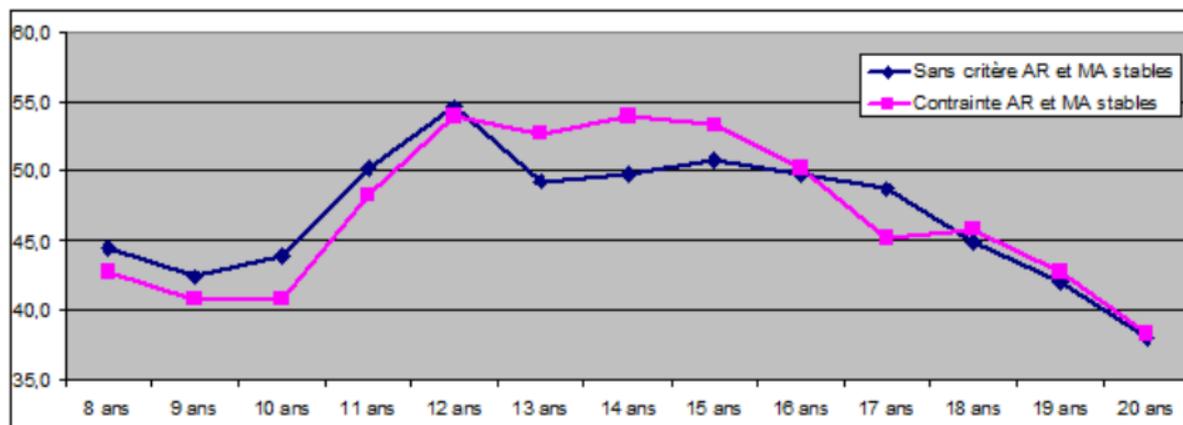


Algorithme appliqué à chaque série

	6 ans	7 ans	8 ans	9 ans	10 ans	11 ans	12 ans	13 ans	14 ans	15 ans	16 ans	17 ans	18 ans	19 ans	20 ans
ET1 : plages de scores min	■	■	■	■	■	■	■			■	■	■	■		
ET2 : plages de stabilité		■	■	■	■	■	■	■	■			■	■	■	
ET3 : intersection		■	■	■	■	■	■					■	■		
ET4 : proximité de fin série		■	■	■	■	■	■								
ET5 : 2 ans mini de recouvrement			■	■	■	■	■								

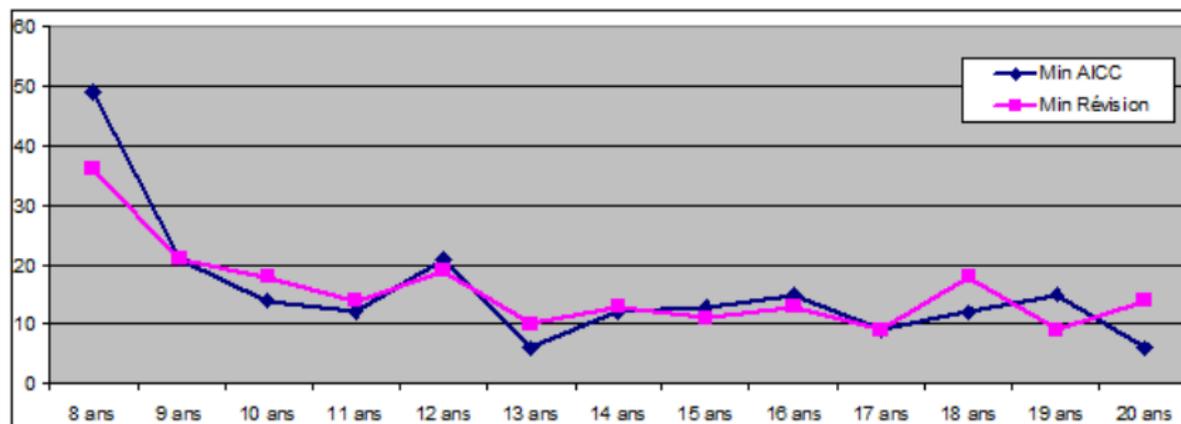
Résultats sur l'IPI - cumul des plages

Choix d'une fenêtre pour l'ensemble des séries à partir des plages : fenêtre 12 ans, soit 6 ans de recouvrement



Résultats sur l'IPI - critères supplémentaires

Choix d'une fenêtre pour chaque série : 8 ans, soit 2 ans de recouvrement



Sommaire

1. Couper les séries, qu'est-ce qui change ?
2. Couper à quelle date ?
3. Quelle longueur de recouvrement ?
4. **Comment raccorder ?**

Juxtaposer ou chaîner

Juxtaposer : CVS du passé (figée) jusqu'au décembre 2011, puis “coller” la CVS de fin de série à partir du point janvier 2012

- avantage : préserve le niveau de la fin de série
- inconvénient : évolution 2011 - 2012 intègre l'effet changement de méthode
- à privilégier pour les séries dont le niveau est une information d'intérêt

Juxtaposer ou chaîner

Juxtaposer : CVS du passé (figée) jusqu'au décembre 2011, puis "coller" la CVS de fin de série à partir du point janvier 2012

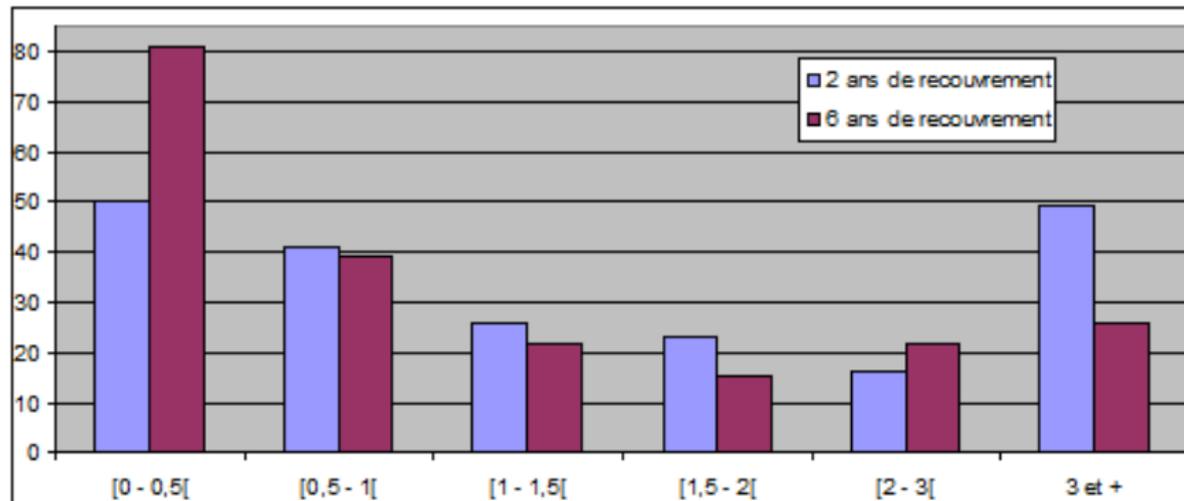
- avantage : préserve le niveau de la fin de série
- inconvénient : évolution 2011 - 2012 intègre l'effet changement de méthode
- à privilégier pour les séries dont le niveau est une information d'intérêt

Chaîner : CVS du passé (figée) jusqu'au décembre 2011, à partir de ce point on chaîne les évolutions mensuelles obtenues sur la CVS de fin de série

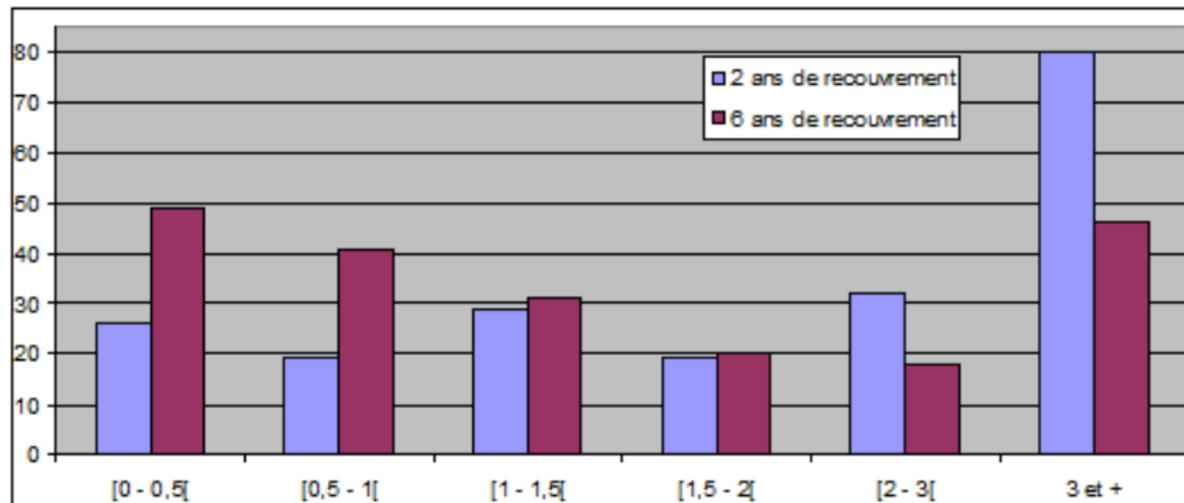
- avantage : évolution décembre 2011 - janvier 2012 interprétable
- inconvénient : possible décalage de niveau sur la CVS de fin de série
- à privilégier pour les séries d'indices

Risque de rupture au point de raccordement dans les deux cas

Recouvrement long réduit la rupture - niveau



Recouvrement long réduit la rupture - évolution



Conclusion

Couper les séries longues change les commentaires sur les CVS et améliore le pré-ajustement, en particulier la correction des jours ouvrables

Conclusion

Couper les séries longues change les commentaires sur les CVS et améliore le pré-ajustement, en particulier la correction des jours ouvrables

Couper en fin de série, 6 ans pour l'IPI, en tenant compte de :

- la période de révision des données brutes
- filtre symétrique au point de coupure

Conclusion

Couper les séries longues change les commentaires sur les CVS et améliore le pré-ajustement, en particulier la correction des jours ouvrables

Couper en fin de série, 6 ans pour l'IPI, en tenant compte de :

- la période de révision des données brutes
- filtre symétrique au point de coupure

Tester la longueur de recouvrement, algorithme permettant de :

- comparer les différentes longueurs
- Prendre en compte la qualité et la stabilité du modèle

Pour l'IPI, 6 ans de recouvrement : filtre symétrique au point de coupure et moins de rupture pour raccorder

Conclusion

Couper les séries longues change les commentaires sur les CVS et améliore le pré-ajustement, en particulier la correction des jours ouvrables

Couper en fin de série, 6 ans pour l'IPI, en tenant compte de :

- la période de révision des données brutes
- filtre symétrique au point de coupure

Tester la longueur de recouvrement, algorithme permettant de :

- comparer les différentes longueurs
- Prendre en compte la qualité et la stabilité du modèle

Pour l'IPI, 6 ans de recouvrement : filtre symétrique au point de coupure et moins de rupture pour raccorder

Résultats empiriques propres à l'IPI, mais démarche généralisable aux séries longues : choix des critères adaptés aux données