

DÉSAISONNALISER LES SÉRIES TRÈS LONGUES PAR SOUS-PÉRIODE

GAINS ET CHOIX DE LA LONGUEUR DE TRAITEMENT - EXEMPLE DES SÉRIES DE L'IPI

Hien PHAM, Alain QUARTIER-LA-TENTE

INSEE, Département des Méthodes Statistiques

`hien.pham@insee.fr alain.quartier-la-tente@insee.fr`

Mots-clés. Désaisonnalisation, ARIMA, Séries longues, Correction des variations saisonnières, Correction des jours ouvrables.

Résumé

Sur longue période, les institutions, les normes de sociétés ainsi que les comportements des agents économiques évoluent, induisant des changements dans la dynamique des séries économiques que l'on étudie. Ainsi, on admet l'idée que les séries temporelles d'une longueur dépassant les 20 années présentent en général un profil changeant. Ces changements pouvant concerner les effets de calendrier (activités les dimanches et jours fériés par exemple), la saisonnalité ou la corrélation temporelle entre observations éloignées modélisée par une fonction ARIMA. Dans ce cas, réaliser une désaisonnalisation sur les séries longues risque de produire des résultats sous-optimaux, en particulier sur les périodes les plus récentes et les plus anciennes.

Le premier objectif de cet article est d'examiner si couper les séries longues permet effectivement d'améliorer la qualité de la désaisonnalisation. Les résultats qui sont présentés s'appuient sur une étude empirique des séries de l'indice de la production industrielle (IPI) qui sont longues de 27 années, de janvier 1990 à décembre 2017. En coupant ces séries en deux sous-périodes au point de janvier 2005 afin de désaisonnaliser par sous-période, nous constatons que la qualité de la phase de pré-ajustement s'améliore. Cette phase concerne la détection des *outliers*, la correction des effets de calendrier et l'estimation d'un modèle ARIMA. En revanche, on constate qu'il n'y a pas de différence sur l'appréciation de la qualité de la décomposition réalisée par la méthode non-paramétrique X11, quand bien même les résultats de la décomposition obtenus sont différents.

Par ailleurs, le choix de la date de coupure est lui-même un problème à résoudre. Ce choix peut être guidé par des informations sur les changements de gestion des données, sur les révisions de méthodes statistiques appliquées, sur les chocs connus de nature économique et institutionnelle ou encore il peut simplement résulter de contraintes de production. Autrement dit, le choix peut être effectué à l'aide d'informations qualitatives existant en dehors des données et de manière indépendantes des données. L'autre objectif de cet article est justement d'examiner s'il est possible de proposer un moyen de choisir une date de coupure qui soit fondée sur l'étude des données.

Abstract

Over a long period of time, the norms of the society as well as the behaviors of the economic agents evolve, inducing changes in the dynamics and the structure of the economic series. Thus, long series of more than 20 years generally have a changing pattern. Performing seasonal adjustment on such series may produce sub-optimal results, mainly in the most recent and the initial parts of the series.

The first objective of this article is to show that cutting long series improves the quality of seasonal adjustment, especially the working days adjustment and the Reg-ARIMA modeling. Then, we will describe a method for choosing a cut-off date by checking the quality of the process and the stability of the models estimated.

Table des matières

Introduction	1
1 Couper les séries longues change les résultats et améliore les modèles	2
1.1 Une histoire différente sur la fin de série	2
1.1.1 Écarts sur les niveaux des CVS	2
1.1.2 Différences sur les taux d'évolution	3
1.1.3 Signes d'évolution parfois contraires	4
1.2 Couper les séries améliore les modèles	5
1.2.1 La correction des jours ouvrables plus efficace	5
1.2.2 Moindre autocorrélation des résidus des modèles Reg-ARIMA	6
1.2.3 Qualité de la décomposition inchangée	6
1.3 Conclusion de la partie	10
2 Méthode pour le choix d'une fenêtre optimale	11
2.1 Que disent les <i>guidelines</i> sur la fin de série ?	11
2.2 Méthode basée sur la comparaison des fenêtres de données	12
2.3 Critères pour le choix des fenêtres	13
2.3.1 Les critères de qualité résumés en un score	13
2.3.2 Les critères de stabilité du schéma et du modèle ARIMA	14
2.4 Description de l'algorithme et résultats sur les séries de l'IPI	16
3 Comment raccorder les CVS de deux sous-périodes au point de coupure	17
3.1 Concaténer les CVS	18
3.2 Chaîner sur une période commune	19
3.3 Chaîner sur une année commune	20
Conclusion	21
A Expression et spécification des modèles Reg-ARIMA	22
B Choix a priori d'une fenêtre d'estimation	22
B.1 Recherche de ruptures en 2009	23
B.2 Conclusion	24
C Changements de profil et conséquences sur la modélisation	24
C.1 Changements sur les effets du calendrier	24
C.2 La détection des <i>outliers</i>	26
C.3 L'identification d'un modèle ARIMA	28
D Exemples de séries IPI avec une plage de fenêtres d'estimation optimale	30
E Description de l'Indice de la production industrielle	34
Références	36

Liste des tableaux

1	Méthode de calcul de la note associée à chaque indicateur	14
2	Poids associé à chaque indicateur pour le calcul d'un score global	14
3	Nombre de <i>outliers</i> de la branche 1052 (fabrication de glaces et sorbets)	27
4	Nombre de <i>outliers</i> de la branche 1412 (fabrication de vêtements de travail)	28

Table des figures

1	Fréquence des écarts relatifs en niveau par intervalle	3
2	Fréquence des écarts sur les taux d'évolution par intervalle	4
3	Fréquence des différences de signes d'évolution par intervalle	4
4	Fréquence des problèmes de jours ouvrables résiduels par type de modèle et par sous-période	5
5	Rejets au test d'absence d'autocorrélation par modèle et par sous-période	6
6	Distribution des statistiques $M7$ selon le modèle sur la sous-période « le Courant »	7
7	Distribution des statistiques $Q-M2$ selon le modèle sur la sous-période « le Courant »	8
8	Fréquence des choix de filtres saisonniers selon la fenêtre de données considérée	9
9	Mauvaise qualité de la statistique $M7$ selon la moyenne mobile et la longueur des séries	10
10	Mauvaise qualité de la statistique $Q - M2$ selon la moyenne mobile et la longueur des séries	10
11	Fréquence de changement du schéma et des ordres de différenciation par fenêtre	15
12	Fréquence de changement des ordres ARMA par fenêtre	15
13	Fréquence relative des fenêtres retenues dans les plages, selon le critère de stabilité	17
14	Fréquence des fenêtres retenues pour chaque série, selon le critère AICC ou de révision	18
15	Fréquence des écarts sur les estimations du niveau du point de janvier 2012 par intervalle	19
16	Fréquence des écarts sur les évolutions décembre 2011 - janvier 2012 par intervalle	19
17	Rupture du coefficient jours ouvrables, exemple de la branche 2920 (industrie de la fabrication de carrosseries et de remorques)	25
18	Dérive linéaire du coefficient jours ouvrables, exemple de la branche 2829 (fabrication de machines diverses d'usage général)	26
19	Multiple changements de sens d'évolution du coefficient jours ouvrables, exemple de la branche 2434 (branche tréfilage à froid)	26
20	Nombre d' <i>outliers</i> détectés par sous-période et selon le modèle	27
21	Série IPI branche 1052 (fabrication de glaces et sorbets)	27
22	Série IPI branche 1412 (fabrication de vêtements de travail)	28
23	Série IPI branche 1722 (fabrication d'articles en papier à usage sanitaire ou domestique)	30
24	Série IPI branche 1032	30
25	Série IPI branche 1042	31
26	Série IPI branche 1073	31
27	Série IPI branche 1083	31
28	Série IPI branche 1200	32
29	Série IPI branche 1392	32
30	Série IPI branche 1722	32
31	Série IPI branche 1724	33
32	Série IPI branche 2015	33
33	Série IPI branche 2041	33

Introduction

Une des questions les plus souvent posées par les producteurs de séries corrigées des variations saisonnières et corrigées des jours ouvrables (CVS)¹ est « combien faut-il de données pour faire une désaisonnalisation? ». La réponse usuelle est « autant que possible » puisque plus nous avons d’observations mieux nous pouvons estimer un modèle qui approxime la structure des données. En revanche, le problème inverse est moins souvent évoqué : « peut-on avoir trop de données? ». En effet, les principales méthodes de désaisonnalisation (X13-ARIMA et TRAMO-SEATS) font appel à des modèles de régression linéaire dont les erreurs suivent un modèle ARIMA — *AutoRégressive Integrated Moving Average* — appelé modèle Reg-ARIMA (annexe A). Ces modèles sont utilisés pour :

- « linéariser » la série en supprimant les « non-linéarités » liés aux effets de calendrier et à la présence de points atypiques (*outliers*) ;
- prolonger la série linéarisée par prévision afin de minimiser les révisions des CVS.

Dans le processus de la désaisonnalisation, cette étape correspond à la phase de « pré-ajustement » et précède la « décomposition », où la composante saisonnière est estimée.

Le modèle Reg-ARIMA suppose que la relation entre la variable à expliquer, i.e. la variable à désaisonnaliser, et les variables explicatives (régresseurs jours ouvrables et *outliers*) est constante dans le temps. Cette hypothèse n’est souvent pas vérifiée sur les séries longues pour principalement deux raisons :

1. La maintenance de séries longues est généralement affectée par des changements dans les définitions, de nomenclatures et de méthodologie. Mêmes si les séries sont rétopolées, les méthodes utilisées peuvent créer des changements dans la structure des données.
2. Sous l’effet de modifications des pratiques de production et de l’organisation du travail, l’influence des jours ouvrables sur la production n’est pas constante. Par exemple on assiste depuis une quinzaine d’année à une extension du travail continu le dimanche dans la branche de l’automobile. Plus récemment la loi du 6 août 2015 « pour la croissance de l’activité et l’égalité des chances économiques » a élargi les possibilités d’ouverture des commerces le dimanche.

Les séries longues seraient alors le produit de processus de génération de données différents. Effectuer l’ajustement saisonnier sur l’ensemble de la série peut donc conduire à des résultats sous-optimaux, principalement sur le début et la fin de série.

C’est pourquoi, pour des séries de plus de 20 ans, les lignes directrices sur l’ajustement saisonnier d’Eurostat (*guidelines*)[2] recommandent d’effectuer l’ajustement saisonnier en plusieurs sous-périodes. Cependant les *guidelines* ne donnent pas d’indications précises sur la façon de gérer les séries longues. Comment décider de la longueur des séries à désaisonnaliser? Quels critères utiliser pour appuyer la décision? Sachant que les choix retenus dépendront également des objectifs de diffusion et des contraintes pesant sur la production. L’objectif de cet article est de répondre à ces questions.

La première partie de l’article montre que couper des séries longues améliore les CVS. Les résultats s’appuient sur l’exemple des séries de l’Indice de la Production Industrielle (IPI). La deuxième partie présente une méthode permettant de choisir la longueur de données en vue de désaisonnaliser la fin de série. Enfin, la dernière partie aborde la question du raccordement des sous-périodes. En annexe figure une description des types de changements dans les données (dérive linéaire ou rupture) qui seraient susceptibles de produire des résultats sous-optimaux sur des séries longues.

1. La notation usuelle est CVS-CJO, mais dans la suite de l’article nous noterons simplement CVS.

1 Couper les séries longues change les résultats et améliore les modèles

L'objectif de cette partie est de comparer les CVS des séries de l'IPI (cf. annexe E) provenant de deux processus de désaisonnalisation, l'un en utilisant l'ensemble des données disponibles et l'autre en coupant les séries. Les séries étudiées sont au nombre de 205² et sont longues de 27 ans, allant de janvier 1990 à décembre 2017. Les désaisonnalisations sont effectuées avec la méthode X13-ARIMA sous le logiciel JDemetra+³. La modélisation Reg-ARIMA est automatique, c'est-à-dire que le schéma de composition, les *outliers*, ainsi que le modèle ARIMA sont automatiquement déterminés par le logiciel [4]. La correction des jours ouvrables est introduite après un test de significativité, avec le jeu de 7 régresseurs comportant les 6 variables de contraste, les lundis aux samedis contre les dimanches et jours fériés, et la variable années bissextiles. Les filtres utilisés pour la décomposition sont également déterminés de manière automatique par JDemetra+.

Les séries de l'IPI sont coupées en janvier 2005 en deux sous-périodes. Cette date est choisie par le service producteur pour des raisons de gestion et aussi afin d'éviter de couper trop près de la crise de 2008. Chaque sous-période est prolongée d'une période de recouvrement de 3 ans après et avant la date de coupure⁴. Ce recouvrement de 3 ans est retenu en raison de la taille de la moyenne mobile $M3 \times 5$, qui est la plus fréquemment utilisée par la méthode X11 pour estimer les coefficients saisonniers. Ainsi, les deux sous-périodes, dénommées par la suite « le Passé » et « le Courant », correspondent respectivement aux fenêtres de données allant de janvier 1990 à décembre 2007 et de janvier 2002 à décembre 2017. Elles sont longues respectivement de 17 et 16 ans.

Nous comparons les deux processus en étudiant les écarts sur les CVS obtenues en sortie. Seuls les écarts mesurés sur les 6 dernières années, de janvier 2012 à décembre 2017, seront présentés, puisque les données les plus consultées par les utilisateurs sont relatives à la fin de série. Si les CVS obtenues présentent peu d'écarts, choisir l'une ou l'autre option aura peu d'incidence sur la publication des résultats et des commentaires économiques associés. La décision de couper ou non les séries s'appuiera alors sur les coûts relatifs des deux options. En revanche, si les résultats sont différents pour une partie des séries, il s'agira de décider quel traitement est « meilleur ». Pour arbitrer il sera alors nécessaire de s'appuyer sur des indicateurs de qualité qui seront à préciser.

1.1 Une histoire différente sur la fin de série

Les CVS obtenues avec les données complètes sont confrontées à celles obtenues sur la sous-période « le Courant », en ne retenant que les 6 dernières années. Trois indicateurs d'écart sont mesurés : les écarts en niveau, les écarts en évolution et les différences dans les signes d'évolution.

1.1.1 Écarts sur les niveaux des CVS

On note :

2. Séries au niveau des classes de la NAF rév. 2, qui est le niveau de désaisonnalisation directe et le premier niveau de publication des CVS. Les niveaux plus agrégés sont obtenus de manière indirecte par agrégation.

3. Logiciel de désaisonnalisation utilisé en production pour les séries de l'IPI et pour la très large majorité des séries CVS publiées par l'Insee et les services statistiques ministériels.

4. La date de COUPURE est définie comme la date à partir de laquelle (ou jusqu'à laquelle, s'il s'agit des données du passé) les CVS sont mises à jour avec de nouveaux modèles. Elle ne correspond pas à la date de début (ou de fin, s'il s'agit du passé) d'estimation qui prend en compte la période de recouvrement.

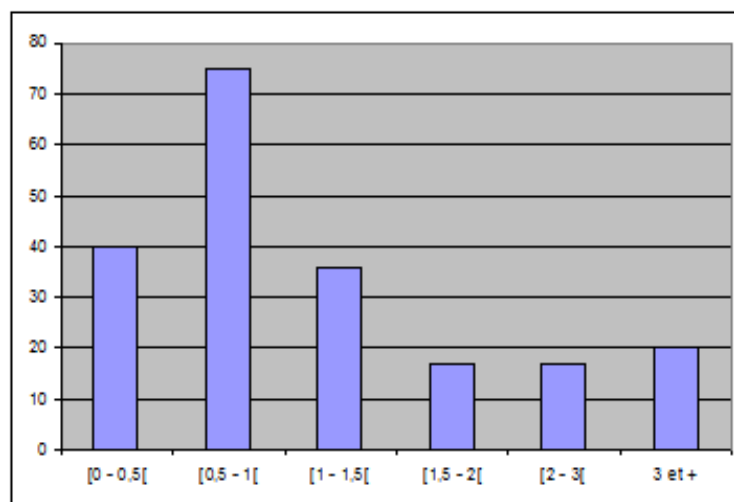
- X_t une série IPI ;
- T le nombre de points utilisés, soit 72 pour une série mensuelle sur 6 années ;
- $SA1_t$ (respectivement $SA2_t$) la CVS obtenue à la date t avec une désaisonnalisée réalisée avec des données complètes (respectivement une désaisonnalisation de la sous-période « le Courant »).

Pour chaque série X_t , les écarts en niveau sont calculés de la façon suivante⁵ :

$$EN(X_t) = \frac{1}{T} \sum_{k=1}^T \frac{|SA1_k - SA2_k|}{SA1_k} \times 100$$

On obtient ainsi 205 valeurs EN , qui mesurent pour chaque série un écart moyen sur les 6 dernières années entre les deux versions de CVS. Pour 90 séries, soit 44 % de l'ensemble, l'écart moyen en niveau est de 1 point ou plus (graphique 1). Pour 10 % des séries l'écart est supérieur à 3 points.

GRAPHIQUE 1 – Fréquence des écarts relatifs en niveau par intervalle



Note de lecture : Sur les 6 dernières années, 40 séries de l'IPI ont un écart moyen en niveau inférieur à 0,5 point.

Un écart important en niveau peut se produire après une différence ponctuelle sur le taux d'évolution, quand bien même sur la fin de série les deux versions de CVS évoluent de façon similaire. Il convient par conséquent de comparer également les taux d'évolution.

1.1.2 Différences sur les taux d'évolution

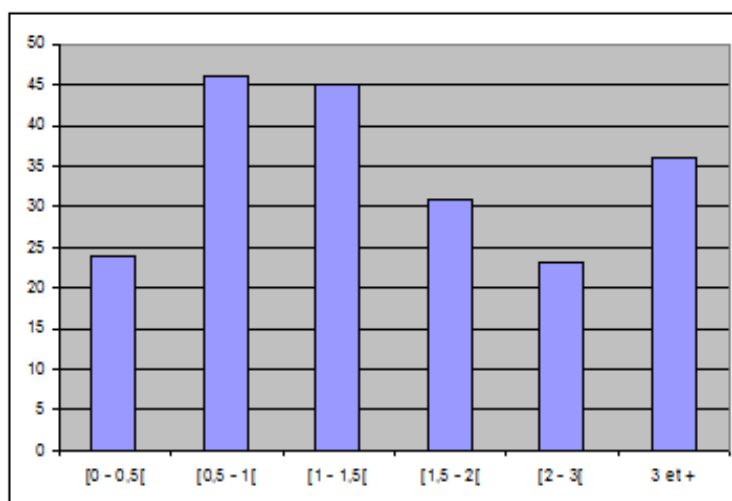
La moyenne des différences sur les taux mensuel d'évolution entre les deux versions de CVS est obtenue de la façon suivante :

$$ETC(X_t) = \frac{1}{T} \sum_{k=2}^T \left| \frac{SA1_k - SA1_{k-1}}{SA1_{k-1}} - \frac{SA2_k - SA2_{k-1}}{SA2_{k-1}} \right| \times 100$$

Les différences sur les taux d'évolution sont très marquées (graphique 2). Pour les deux tiers des séries les taux d'évolution diffèrent en moyenne de plus de 1 point. Ceci induit une appréciation des évolutions économiques différente pour une part importante des séries. Pour une série sur cinq les différences en évolution dépassent même les 3 points en moyenne.

5. Il s'agit d'un écart relatif. En effet une différence de 2 points d'indice n'a pas la même signification si l'on la rapporte à un niveau de 200 ou de 50.

GRAPHIQUE 2 – Fréquence des écarts sur les taux d'évolution par intervalle



Note de lecture : Sur les 6 dernières années, 24 séries de l'IPI présentent un écart moyen sur les taux d'évolution inférieur à 0,5 point.

1.1.3 Signes d'évolution parfois contraires

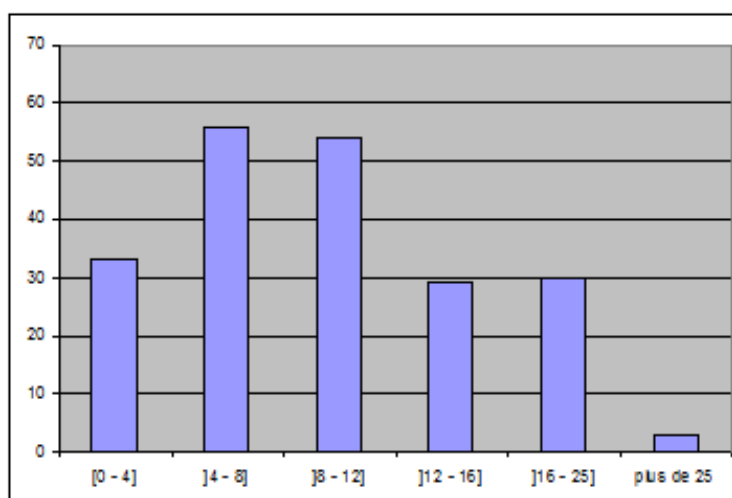
Enfin, on mesure ici la fréquence à laquelle les deux versions de CVS donnent des évolutions de sens contraires, indépendamment du taux d'évolution.

$$CohS(X_t) = \frac{1}{T} \sum_{k=2}^T I_{(SA1_k - SA1_{k-1})(SA2_k - SA2_{k-1}) < 0} \times 100$$

Avec $I_{(SA1_k - SA1_{k-1})(SA2_k - SA2_{k-1}) < 0} = 1$ si $(SA1_k - SA1_{k-1})(SA2_k - SA2_{k-1}) < 0$, 0 sinon.

Un $CohS$ égal à 8 indique des différences de signes d'évolution pour 8 % des points sur les 6 dernières années d'une série. C'est-à-dire une fois par an en moyenne. Le niveau de 16 % correspond à des différences de signes se produisant deux fois par an.

GRAPHIQUE 3 – Fréquence des différences de signes d'évolution par intervalle



Note de lecture : Sur les 6 dernières années, 33 séries de l'IPI ont des différences de signes d'évolution sur 4 % des points ou moins.

Pour trois séries sur cinq, les évolutions mensuelles des deux versions de CVS sont de signes contraires au moins une fois par an (graphique 3). Pour une série sur cinq, les différences de signes se produisent au moins deux fois par an.

En conclusion, couper les séries change les résultats. Selon l'option retenue, l'appréciation des évolutions économiques serait différente pour une part importante des séries. Toutefois, quelle version de CVS est-elle préférable ?

1.2 Couper les séries améliore les modèles

Le choix entre les options s'appuiera sur les critères permettant de juger de la qualité respective des deux procédures. Cependant, nous veillerons à comparer les indicateurs résultant de modèles différents, mais qui devront porter sur un même ensemble de données. Ainsi, pour chaque série, le modèle estimé à l'aide des données complètes sera évalué en l'appliquant à deux sous-périodes, « le Passé » et « le Courant ». Les deux jeux d'indicateurs ainsi obtenus seront alors comparables à ceux produits par les modèles estimés par sous-période, respectivement « le Passé » et « le Courant ».

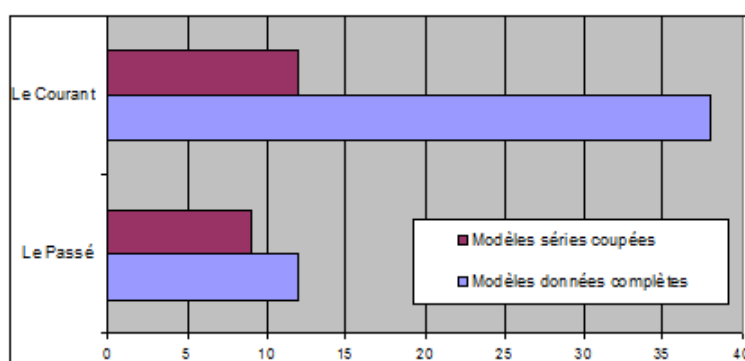
Les indicateurs considérés concernent la qualité de la phase de pré-ajustement ainsi que celle de la décomposition. Les deux critères les plus importants sont relatifs aux problèmes de saisonnalité résiduelle et de jours ouvrables résiduels.

1.2.1 La correction des jours ouvrables plus efficace

Le problème de saisonnalité résiduelle est peu présent sur les séries IPI (une dizaine de séries)⁶ et il n'y a pas de différence significative entre les deux processus. Par conséquent, la saisonnalité résiduelle n'est pas un critère suffisant pour choisir entre les deux options.

En revanche, les problèmes de jours ouvrables résiduels sont plus fréquents et plus discriminants (graphique 4). Si l'on considère la sous-période « le Courant », la présence de jours ouvrables résiduels est 3 fois plus fréquente lorsqu'on utilise le modèle issu des données complètes (38 séries contre 12).

GRAPHIQUE 4 – Fréquence des problèmes de jours ouvrables résiduels par type de modèle et par sous-période



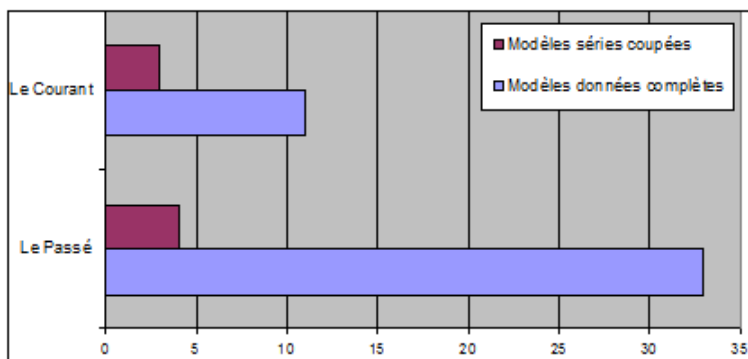
Note de lecture : Sur la sous-période « le Courant », 38 séries présentent des effets de jours ouvrables résiduels avec les modèles « données complètes », contre 12 sur séries coupées.

6. Le test QS sur la CVS est ici notre indicateur de référence [3]. La statistique QS permet d'examiner s'il existe des autocorrélations positives de l'ordre de la saisonnalité dans la série désaisonnalisée (ordre 4 pour les séries trimestrielles et 12 pour les séries mensuelles).

1.2.2 Moindre autocorrélation des résidus des modèles Reg-ARIMA

En théorie, un modèle Reg-ARIMA est valide lorsque ses résidus vérifient notamment les hypothèses de normalité, d'homoscédasticité et d'absence d'autocorrélation. Les résultats des tests montrent que sur ces 3 critères, les modèles issus des séries coupées produisent systématiquement moins d'anomalies. Les différences sont en particulier très marquées si l'on considère le problème d'autocorrélation des résidus (graphique 5), qui est le critère le plus important parmi les trois⁷.

GRAPHIQUE 5 – Rejets au test d'absence d'autocorrélation par modèle et par sous-période



Note de lecture : Sur la sous-période « le Courant », 11 séries présentent des résidus autocorrélés avec les modèles « données complètes », contre 3 sur séries coupées.

Les critères relatifs aux problèmes de jours ouvrables résiduels et à la qualité des résidus Reg-ARIMA conduisent à privilégier la désaisonnalisation des séries de l'IPI par sous-période. Ce constat rejoint l'idée selon laquelle un modèle à paramètres et à coefficients constants estimés sur longue période conduit à des ajustements sous-optimaux. Couper des séries longues améliore le pré-ajustement, mais qu'en est-il de l'estimation des coefficients saisonniers dans la phase de décomposition ?

1.2.3 Qualité de la décomposition inchangée

La décomposition X11 s'effectue avec des filtres choisis de manière automatique⁸ en évaluant l'importance relative des composantes tendance-cycle et saisonnier par rapport au bruit (l'irrégulier). Changer la longueur des données peut donc influencer sur le choix des filtres automatiques finaux, et en conséquence modifier les résultats de la décomposition⁹.

Sur la sous-période « le Courant », on constate que la moyenne mobile de Henderson choisie par X11 change pour 22 % des séries lorsque le choix est fait sur les données tronquées au lieu des données complètes. Quant au filtre saisonnier final, un choix différent est fait pour 16 % des séries¹⁰. Les différences de filtres contribuent aux différences sur les résultats présentées dans la

7. En effet, en présence d'autocorrélation des résidus les estimations des coefficients du modèle Reg-ARIMA sont biaisées. L'hétéroscédasticité et la non normalité ne biaisent pas les estimations des coefficients [8].

8. Les moyennes mobiles de Henderson pour l'estimation de la tendance-cycle et moyennes mobiles appliquées à la composante saisonnier-irrégulier pour l'estimation des coefficients saisonniers.

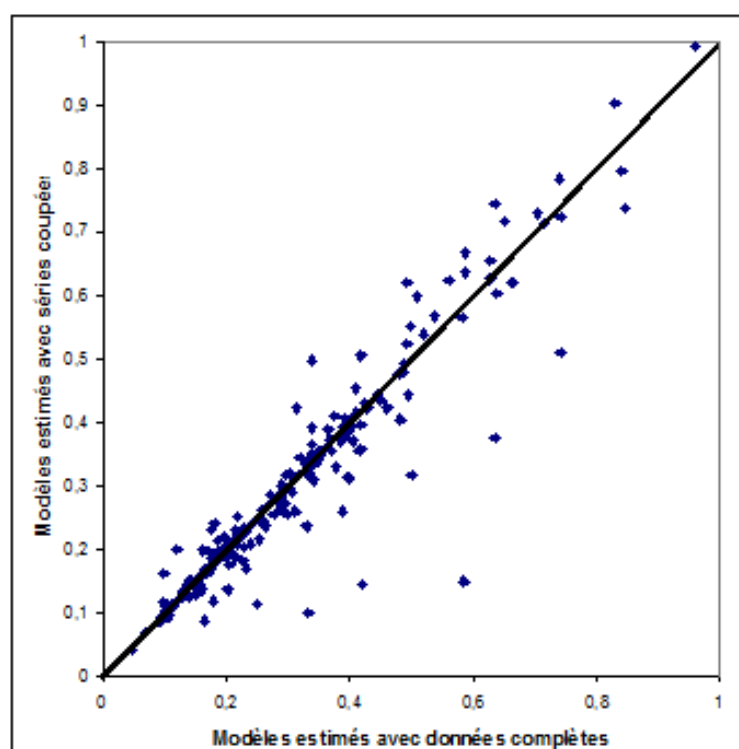
9. X11 effectue des corrections automatiques sur les points de l'irrégulier qui sont détectés comme atypiques [6]. Les différences sur les séries linéarisées peuvent modifier le nombre de points corrigés par X11, ainsi que le niveau de correction de chacun de ces points. Ce traitement de points atypiques affecte les résultats de la décomposition. Toutefois dans cet article, nous négligeons l'impact de ces corrections.

10. Au total, le changement de filtre de décomposition concerne un tiers des séries, soit la moyenne mobile de Henderson, soit le filtre saisonnier, soit les deux à la fois.

section 1.1. Toutefois, on constate que le choix de couper ou non une série ne modifie pas notre appréciation de la qualité de la phase de décomposition.

Pour juger de la qualité de la décomposition, on s'appuie traditionnellement sur les statistiques M (M1 à M11) et Q (synthèse des statistiques M) calculées par X11 [7]. Ces statistiques renseignent sur la qualité de décomposition réalisée par X11. Elles comparent la taille de l'irrégulier (bruit) au regard de celles des composantes tendance-cycle et saisonnalité (signaux) et déterminent la présence d'une saisonnalité évolutive dans la série. Une série peu bruitée comportant des signaux forts et stables dans le temps sera plus aisée à désaisonnaliser. En privilégiant les deux plus importantes statistiques, la M7 et la Q-M2¹¹, on constate que leurs distributions issues des deux traitements sont proches (graphiques 6 et 7).

GRAPHIQUE 6 – Distribution des statistiques M7 selon le modèle sur la sous-période « le Courant »



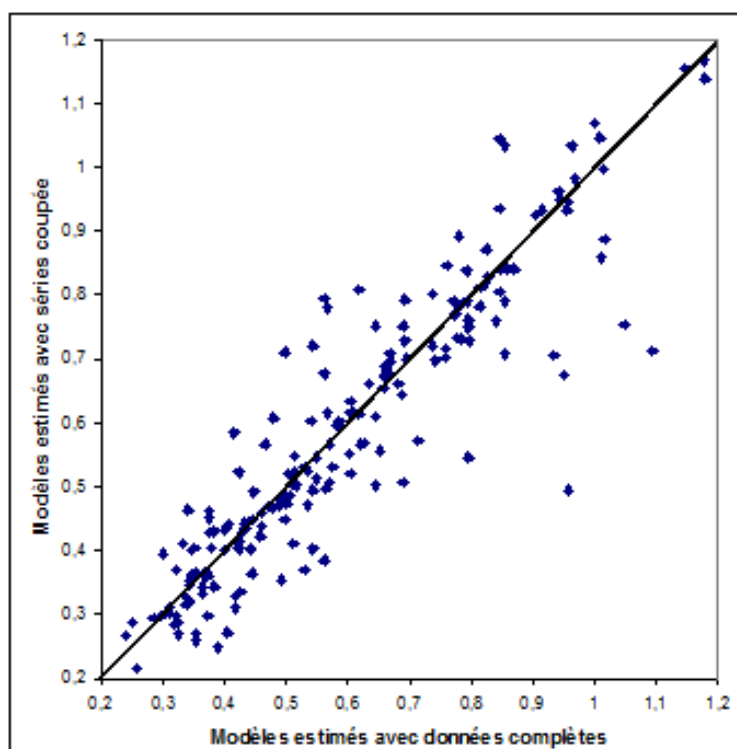
Note de lecture : Chaque point représente une série. Au-dessus de la bissectrice la statistique M7 du modèle « données complètes » est plus petite que celle des séries coupées (donc meilleure). En dessous de la bissectrice c'est l'inverse. La statistique M7 est considérée comme « bonne » lorsqu'elle est inférieure à 1.

Ce résultat paraît cohérent. En effet, contrairement aux paramètres des modèles Reg-ARIMA qui dépendent de l'ensemble des données considérées, l'estimation des coefficients saisonniers dans la phase de décomposition est réalisée sur des sous-périodes par les moyennes mobiles. Ainsi, sur une série longue, le profil saisonnier du début de la série ne contribue pas aux coefficients saisonniers calculés sur la fin de série, et inversement. La qualité de la décomposition n'est donc pas affectée par le problème des séries longues¹². De plus, couper une série avec une période de

11. M7 indique si la saisonnalité est identifiable par X11. La Q-M2 est une synthèse de l'ensemble des statistiques M, excluant la statistique M2.

12. Les résultats présentés ici ne sont pas suffisants pour expliquer les distributions inchangées des statistiques M7 et Q-M2. Les éléments complémentaires seront apportés en fin de cette section.

GRAPHIQUE 7 – Distribution des statistiques Q-M2 selon le modèle sur la sous-période « le Courant »



Note de lecture : Au-dessus de la bissectrice la statistique Q-M2 du modèle « données complètes » est plus petite que celle des séries coupées (donc meilleure). La statistique Q-M2 est considérée comme « bonne » lorsqu'elle est inférieure à 1.

recouvrement suffisamment longue ne modifie l'estimation des coefficients saisonniers que dans la mesure où la série linéarisée change¹³ et que X11 opère un changement de filtres¹⁴.

Nous avons vu que couper les séries de l'IPI modifie le choix des filtres saisonniers pour une série sur 5. Le graphique 8 montre par ailleurs que la moyenne mobile $M3 \times 5$ est utilisée pour une très large majorité des séries, quelle que soit la fenêtre de données considérée. En revanche, la moyenne mobile $M3 \times 9$ est extrêmement rare. Par ailleurs, nous avons également souligné que la qualité de la décomposition, observée à l'aide des statistiques $M7$ et $Q - M2$, reste globalement inchangée que l'on coupe ou non les séries¹⁵.

La stabilité des choix des filtres saisonniers n'explique pas seul ce résultat. En effet, pour les séries de l'IPI, la qualité de la décomposition ne dépend pas de filtres choisis, mais de la longueur de données sur laquelle on examine les statistiques M et Q .

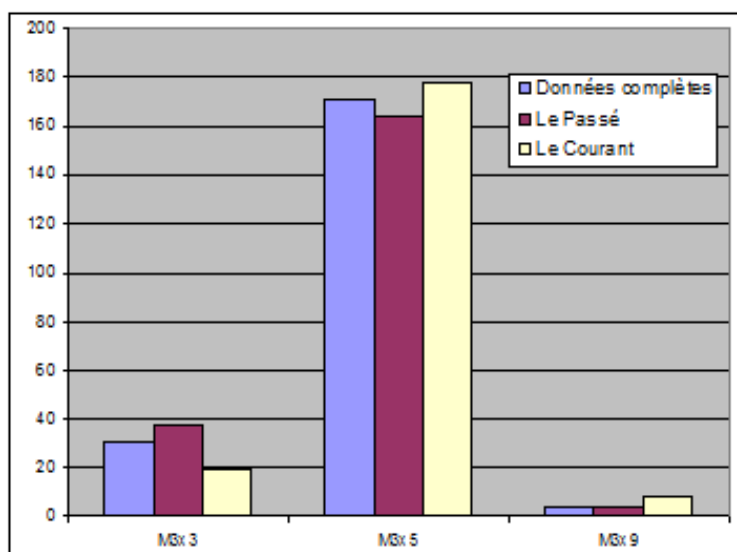
Pour une fenêtre de données fixée, nous avons comparé 3 jeux de décompositions. Chaque jeu

13. Le modèle Reg-ARIMA est modifié pour 60 % des séries, même sans tenir compte des différences sur les outliers détectés. Le test de significativité des régresseurs de calendrier change de conclusion rarement (9 % des séries). Le schéma de composition est plutôt stable, il change pour 19 % des séries. Tandis que le modèle ARIMA est différent pour 47 % des séries, sachant que c'est la partie régulière qui est instable, 40 % des séries contre 12 % pour la partie saisonnière.

14. On a vu que cela affecte un tiers des séries. Au total, 74 % des séries ont changé de paramètres de pré-ajustement et/ou de décomposition entre le modèle estimé avec données complètes et le modèle courant.

15. Nous avons comparé deux choix de modèles et de filtres différents, ceux déduits sur données complètes versus ceux de la sous-période « le Courant ». Cependant l'examen de la qualité des résultats porte sur la même fenêtre de données, en l'occurrence « le Courant ».

GRAPHIQUE 8 – Fréquence des choix de filtres saisonniers selon la fenêtre de données considérée



Note de lecture : La moyenne mobile $M3 \times 3$ est choisie par X11 pour 30 séries quand le choix est évalué sur les données complètes, 37 séries quand la sous-période « le Passé » est considérée et 19 pour la sous-période « le Courant ».

est obtenu en imposant un même filtre saisonnier aux 205 séries de l'IPI, soient respectivement les moyennes mobiles $M3 \times 3$, $M3 \times 5$ et $M3 \times 9$. Par conséquent, pour une fenêtre de données, nous avons 615 résultats sur lesquels on peut examiner la qualité de la décomposition. Enfin, ces mêmes jeux de tests sont produits sur 15 différentes fenêtres de données, de longueur allant de 6 à 20 ans.

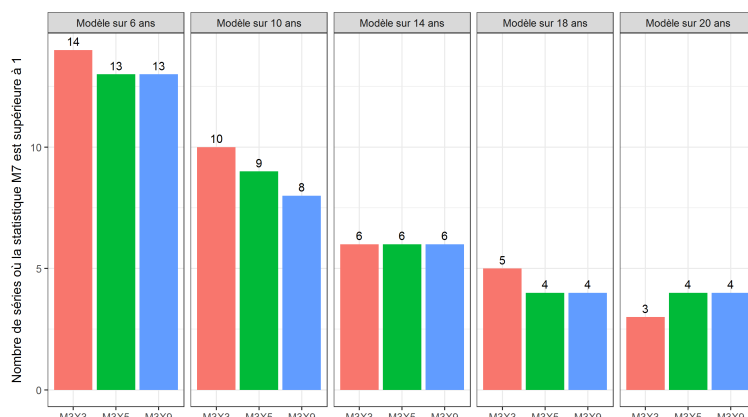
Le constat tiré de ces jeux d'essais est que, quelle que soit la fenêtre de données considérée, la qualité des statistiques $M7$ et $Q - M2$ n'est pas différenciée entre les moyennes mobiles $M3 \times 3$ et $M3 \times 5$ (graphiques 9 et 10)¹⁶. Le choix de la moyenne mobile $M3 \times 9$ n'affecte pas non plus la statistique $M7$, mais tend à accroître la valeur de la statistique $Q - M2$ (donc la dégrade) par rapport à la moyenne mobile $M3 \times 5$, et cela uniquement sur les fenêtres de données longues (à partir de 13 ans). Compte tenu du fait que le choix automatique de la moyenne mobile $M3 \times 9$ par X11 est rare (graphique 8), les changements de moyennes mobiles s'opèrent entre les $M3 \times 3$ et $M3 \times 5$, qui agissent de la même façon sur la qualité de la décomposition.

Enfin, ces tests montrent également que la qualité de la décomposition dépend de la longueur de données considérée. Les statistiques $M7$ et $Q - M2$ diminuent avec la longueur des données¹⁷, c'est-à-dire qu'elles s'améliorent. Ce résultat est vérifié quel que soit le filtre saisonnier utilisé.

16. Ces graphiques sont produits à titre d'illustration, les résultats ont été obtenus par régression des statistiques $M7$ et $Q - M2$ sur la taille des moyennes mobiles.

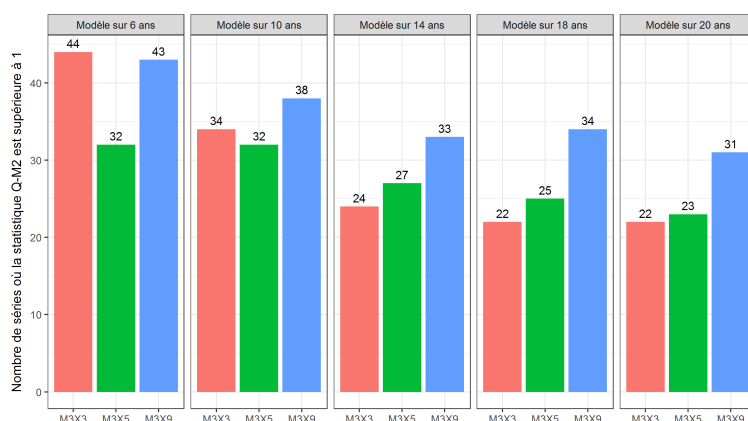
17. Pour être précis, si on compare les statistiques $M7$ et $Q - M2$ issues de la décomposition de deux fenêtres de données, par exemple 2008-2017 (10 ans) et 2011-2017 (7 ans), on trouverait que les statistiques obtenues sur les données 2008-2017 sont globalement meilleures que celles obtenues sur les données 2011-2017. Toutefois, cela ne signifie pas que la partie 2011-2017 de la série dans la fenêtre 2008-2017 soit mieux décomposée.

GRAPHIQUE 9 – Mauvaise qualité de la statistique $M7$ selon la moyenne mobile et la longueur des séries



Note de lecture : Sur la fenêtre de données longue de 6 ans, 14 séries ont une statistique $M7$ supérieure à 1 quand la moyenne mobile $M3 \times 3$ est utilisée et 13 séries sont dans ce cas quand ce sont les moyennes mobiles $M3 \times 5$ et $M3 \times 9$.

GRAPHIQUE 10 – Mauvaise qualité de la statistique $Q - M2$ selon la moyenne mobile et la longueur des séries



Note de lecture : Sur la fenêtre de données longue de 6 ans, 44 séries ont une statistique $Q - M2$ supérieure à 1 quand la moyenne mobile $M3 \times 3$ est utilisée, 32 séries quand il s'agit de la moyenne mobile $M3 \times 5$ et 43 séries pour la moyenne mobile $M3 \times 9$.

1.3 Conclusion de la partie

Pour résumer cette première partie, traiter les séries longues par sous-période modifie les séries CVS produites, ainsi que les commentaires économiques associés. Les différences sur les CVS sont dues d'une part aux différences sur les modèles Reg-ARIMA de la phase de pré-ajustement, et, d'autre part, aux changements de filtres utilisés pour la décomposition. Si les modèles Reg-ARIMA sont globalement de meilleure qualité quand on coupe les séries — problèmes de jours ouvrables résiduels et ceux d'autocorrélation des résidus moins fréquents — la qualité de la décomposition en revanche n'est pas modifiée. Cela implique notamment que traiter les séries longues par sous-période permet une meilleure correction des jours ouvrables et de réduire les risques de révisions dues à l'ajout d'un nouveau point (meilleures prévisions du modèle Reg-ARIMA).

2 Méthode pour le choix d'une fenêtre optimale

Dans la partie 1, nous avons conclu que couper les séries de l'IPI en 2005 avec une période de recouvrement de 3 ans de chaque côté permet d'améliorer la qualité de pré-ajustement. Ainsi sur la sous-période « le Courant », longue de 16 ans, les modèles Reg-ARIMA se comportent mieux que ceux obtenus avec 27 ans. La question que l'on peut soulever ici est, 16 ans, est-ce toujours trop long (ou au contraire trop court) ? Autrement dit, au lieu de 16 ans, si l'on avait retenu que 10 ans, n'aurions-nous pas encore de meilleurs modèles de pré-ajustement¹⁸ ? Quelle serait la longueur minimale permettant de produire des modèles de qualité ?

Pour répondre à ces questions, nous privilégions la fin de série, qui fait généralement l'objet de commentaires dans les publications. On s'interrogera ainsi sur la quantité de données nécessaire et suffisante pour assurer un modèle de qualité sur la fin de série. Mais d'abord, comment pourrions-nous définir et caractériser une « fin de série » ?

2.1 Que disent les *guidelines* sur la fin de série ?

Dans la section 4.3 des lignes directrices (*guidelines*), Eurostat recommande de choisir une date de coupure qui correspond à au moins 3 années précédant l'année où les données brutes restent révisables, de mettre à jour les CVS uniquement à partir de la date choisie et de figer tous les résultats antérieurs. Cette situation est à maintenir pendant 5 ans, avant de choisir de nouveau une nouvelle date de coupure¹⁹. Pour les séries de l'IPI, les données brutes sont révisables sur la dernière année²⁰. La dernière année de données étant 2017, selon les *guidelines* la date de coupure serait donc au plus tôt au janvier 2014.

Toutefois, selon la même section 4.3 des *guidelines*, le choix de la date doit également tenir compte de la longueur de données nécessaire pour obtenir l'estimation finale par un filtre symétrique du point de coupure²¹. La moyenne mobile la plus longue qui peut être choisie par X11 est la moyenne mobile $M3 \times 9$. Celle-ci est utilisée pour extraire la composante saisonnière et nécessite 5 années de données de chaque côté du point central à estimer²². De ce fait si l'on exclut l'année 2017, dont les données brutes restent révisables, les 5 années à considérer sont donc 2012 à 2016. Ainsi, quel que soit le filtre choisi par X11, l'estimation des coefficients saisonniers des points antérieurs à décembre 2011 serait alors réalisée avec des filtres symétriques sur données brutes définitives. Tous les points du passé, jusqu'au décembre 2011 inclus, sont donc en estimation finale et ne sont plus en principe affectés par des révisions. On peut donc figer la CVS jusqu'au point de décembre 2011. Le point de janvier 2012 en revanche, reste révisable si la valeur brute de janvier 2017 est mise à jour. On souhaite donc que la CVS soit révisable à partir du point de janvier 2012 inclus. Ainsi couper les séries de l'IPI en janvier 2012, c'est-à-dire que l'on ne met à jour les CVS qu'à partir de janvier 2012, répondrait aux recommandations de la section 4.3 des *guidelines*. C'est ainsi que nous allons désormais définir notre « fin de séries », la période allant de janvier 2012 à décembre 2017, soit 6 ans.

18. Pour la correction des jours ouvrables par exemple, on estimerait des coefficients plus proches de ceux attendus sur la fin de séries.

19. « *A starting date for the earliest revision of the seasonally adjusted data should be set at the beginning of a year, three years before the revision period of the unadjusted data. This date should be kept fixed for up to five years.* »

20. La période de révision des données brutes est en réalité plus courte, mais cela ne change rien au choix de la date de coupure.

21. « *This choice takes into account the extent of revisions of unadjusted data as well as the normal convergence properties of seasonal adjustment filters and the period for the filter to become symmetric.* »

22. La moyenne mobile la plus fréquemment utilisée est cependant la $M3 \times 5$ qui nécessite 3 années de données de chaque côté du point central à estimer. Par ailleurs, le filtre final de X11 obtenu par composition de l'ensemble des moyennes mobiles utilisées est longue de 14 ans, i.e. 7 ans de chaque côté. Néanmoins les coefficients du filtre sur les extrémités convergent très vite vers 0.

Cependant, le chapitre 4 des *guidelines* est relatif aux politiques de révision des CVS. Il permet de déterminer la date à partir de laquelle mettre à jour les CVS, mais n'indique pas la longueur de données à utiliser pour estimer la CVS de la fin de séries. La section 6.2, qui concerne la gestion des séries longues, donne quelques indications : « *Perform the seasonal adjustment on partially overlapping sub-periods, each possibly longer than seven years, selected by means of tests and graphical inspection* ».

Finalement, les recommandations des *guidelines* supposent que l'on couperait les séries en janvier 2012, en prévoyant une période de recouvrement avec le passé. La longueur totale, fin de série et période de recouvrement incluse, devrait contenir au moins 8 années de données. Le choix de la longueur, doit s'appuyer sur des tests et des analyses graphiques. Malheureusement les *guidelines* ne précisent pas lesquels. C'est ce problème que nous allons considérer.

2.2 Méthode basée sur la comparaison des fenêtres de données

Le problème du choix de la longueur d'estimation revient à choisir la longueur de la période de recouvrement, puisque la fin de série est fixée par défaut à 6 ans. La section 6.2 des *guidelines* suggère que cette période de recouvrement est au minimum de 2 ans, or cela paraît insuffisant. En effet, le coefficient saisonnier du point de janvier 2012 ne sera estimé avec un filtre symétrique que si l'on dispose des données depuis janvier 2007, c'est-à-dire 5 ans de recouvrement.

La longueur de recouvrement à retenir dépendra du profil de la série, et il n'y a pas lieu de penser que l'on parviendra au même choix pour toutes les séries. Néanmoins, pour une gestion simplifiée en production il sera nécessaire de faire un choix unique pour toutes les séries. Dans cette partie, nous allons commencer par considérer chaque série de manière indépendante et étudier comment on peut choisir la longueur de recouvrement pour chacune d'elles.

La démarche que l'on propose ici consiste à comparer, pour chaque série et à l'aide de critères préalablement choisis, les différentes fenêtres de données en faisant varier la longueur de la période de recouvrement. Plus précisément, nous comparerons les modèles automatiques estimés par JDemetra+ sur chacune des fenêtres de données. L'utilisation de la modélisation automatique est justifiée dans le contexte de la production. En effet, après le choix d'une fenêtre de données, la première étape à mettre en oeuvre en production consistera à chercher un modèle adapté aux données, en l'initialisant par un modèle automatique estimé par JDemetra+²³. Si le modèle automatique de départ est déjà un « bon » modèle, on aura alors peu ou pas de travail supplémentaire à fournir pour assurer une production de qualité. Dans cette optique, il s'agira de choisir pour chaque série la fenêtre de données qui donnera le « meilleur » modèle automatique estimé par JDemetra+.

Les fenêtres de données que l'on considère sont longues de 6 à 20 ans, c'est-à-dire en utilisant entre 0 et 14 ans de recouvrement, soit 15 fenêtres au total²⁴. Enfin, ce qui nous importe n'est pas la qualité portée sur toute la longueur d'une fenêtre, mais celle relative à la fin de série. Ainsi, nous allons considérer les modèles automatiques estimés sur chaque fenêtre de données, mais l'examen de la qualité des modèles sera réalisé uniquement au regard de la fin de série. Il reste à préciser les critères qui vont permettre de décider quel est le « meilleur » modèle et quelle longueur de recouvrement retenir.

23. Il ne serait pas pertinent de reprendre le modèle utilisé précédemment en production, puisque la principale raison qui motive la décision de couper les séries est que les modèles obtenus sur des séries longues sont sous-optimaux et non adaptés à la fin de série.

24. Sachant que la longueur minimale de recouvrement préconisée par les *guidelines* est de 2 ans, ou de 3 à 5 ans si l'on souhaite une décomposition avec un filtre symétrique au point de coupure, les premières fenêtres de données seront de fait exclues du choix. Les fenêtres courtes sont testées uniquement dans le but de vérifier la stabilité des paramètres des modèles lorsque l'on est proche du point de coupure.

2.3 Critères pour le choix des fenêtres

Les critères de sélection que nous avons retenus sont de deux types :

- les indicateurs statistiques permettant de juger de la qualité des modèles ;
- les critères de stabilité de certains paramètres du modèle.

Par construction, les choix des critères à retenir vont influencer le choix de la fenêtre « optimale » d'estimation. Ils peuvent donc varier en fonction des contraintes de production et des besoins des producteurs. La liste des indicateurs retenus dans le cadre de cette étude est donnée à titre indicatif et n'a pas vocation à servir de référence : ce sont ceux qui, dans le cadre de l'étude de la désaisonnalisation des séries de l'IPI, nous paraissaient les plus pertinents et qui étaient facilement calculables avec JDemetra+ [5].

2.3.1 Les critères de qualité résumés en un score

La qualité des modèles est jugée selon trois critères : l'absence de saisonnalité résiduelle, l'absence d'effet calendrier résiduel et la qualité du modèle Reg-ARIMA. Les deux premiers critères correspondent à la finalité de la production des CVS : corriger des variations saisonnières et des effets de calendrier. Le troisième critère correspond à des contraintes de diffusion : on souhaite limiter les révisions. Les révisions sont dues à l'ajout de nouveaux points et à la qualité du modèle Reg-ARIMA : en fin de série, les CVS sont estimées à l'aide de valeurs prévues par le modèle.

Pour la sélection de la fenêtre temporelle six indicateurs sont retenus :

- un test de saisonnalité résiduelle (test QS) ;
- un test d'effet de calendrier résiduel (f-test) ;
- trois tests sur la qualité des résidus du modèle Reg-ARIMA : un test d'indépendance (test de Ljung-Box), un test d'homoscédasticité (test de Ljung-Box sur le carré des résidus) et un test de normalité (test de Doornik-Hansen).
- Un test sur la qualité des prévisions en *out of sample* – hors échantillon – (moyenne nulle des erreurs de prévision).

La saisonnalité résiduelle concerne peu de séries, le test correspondant ne sera pas un critère de différenciation fort entre les fenêtres. En revanche, le test d'effet de calendrier résiduel pénalisera les fenêtres trop courtes et les fenêtres trop longues. En effet, si la période d'estimation est courte alors les estimations peuvent être biaisées et si la période d'estimation est longue les coefficients estimés ne refléteraient pas les effets de calendrier sur la fin de la série. Dans les deux cas la correction des jours ouvrables (CJO) serait sous-optimale et un effet résiduel subsisterait.

Les critères sur la qualité des résidus, et en particulier celui sur l'indépendance des résidus, vont favoriser les fenêtres courtes. Cela vient du fait que sur les courtes périodes le modèle Reg-ARIMA retenu reflétera davantage la structure des données récentes. À l'inverse, le test sur la qualité de la prévision pénalisera les fenêtres courtes puisque dans ce cas les estimations seront moins stables.

À chaque indicateur est associée une note entre 0, 3 et 4 en fonction de la p-valeur du test associé. Dans le logiciel de désaisonnalisation JDemetra+ chaque indicateur est associé à une modalité (« bon », « incertain », « mauvais » et « sévère ») et c'est cette modalité qui est utilisée pour calculer la note :

- lorsqu'un indicateur est considéré comme « bon » on lui associe la note de 0 ;
- lorsqu'un indicateur est considéré comme « incertain » on lui associe également la note de 0²⁵ ;
- lorsqu'un indicateur est considéré comme « mauvais » on lui associe la note de 3 ;

25. Une autre alternative consiste à attribuer la note 1. Le score serait un peu plus sévère et on réduirait le nombre de fenêtres avec un score minimal.

— lorsqu'un indicateur est considéré comme « sévère » on lui associe la note de 4 ;
C'est-à-dire que plus la note sera faible plus l'indicateur sera considéré comme satisfaisant.
Le tableau 1 détaille, pour les indicateurs retenus, les seuils utilisés pour calculer les notes.

TABLEAU 1 – Méthode de calcul de la note associée à chaque indicateur

Notes et modalités associées à l'indicateur	0 (« bon »)	0 (« incertain »)	3 (« mauvais »)	4 (« sévère »)
QS et f-test	$p > 0,05$	$0,01 < p \leq 0,05$	$0,001 < p \leq 0,01$	$p \leq 0,001$
Tests sur les résidus	$p > 0,1$	$0,01 < p \leq 0,1$	$p \leq 0,01$	
Test sur les prévisions	$p > 0,1$	$0,01 < p \leq 0,1$	$p \leq 0,01$	

Note de lecture : $p = p$ -valeur.

Les hypothèses nulles des tests sont toujours « le modèle est correctement spécifié » : absence d'effet résiduel, résidus normaux, indépendants, homoscédastiques et moyenne des prévisions nulle.

Toutes les notes sont synthétisées en un score qui tient compte de l'importance des différents indicateurs (un problème de normalité des résidus n'est pas aussi important qu'un problème de saisonnalité résiduelle). Ainsi, chaque indicateur est associé à un poids : le score final est obtenu en effectuant une moyenne pondérée de la note associée à l'indicateur et de son poids (tableau 2)²⁶. Le score reflète bien la qualité d'un modèle sur la fin de la série mais estimé sur une fenêtre. Les scores des différentes fenêtres sont donc comparables puisqu'ils portent sur les mêmes données.

TABLEAU 2 – Poids associé à chaque indicateur pour le calcul d'un score global

Indicateur	Poids
Saisonnalité résiduelle	15
Effets jours ouvrables résiduels	15
Autocorrélation des résidus	15
Homoscédasticité des résidus	5
Normalité des résidus	5
Moyenne des erreurs de prévision	15

2.3.2 Les critères de stabilité du schéma et du modèle ARIMA

On considère qu'un modèle est stable si les paramètres identifiés ne varient pas lorsqu'on les estime sur plusieurs fenêtres consécutives. Ainsi, l'ajout d'une année de données supplémentaires présentera moins de risques de modifier le nombre et la nature des paramètres à estimer. Rappelons que le choix d'une longueur de recouvrement a pour objectif de produire un modèle qui sera utilisé pour désaisonnaliser les séries qui seront complétées chaque mois ou trimestre par de nouvelles données. Avec un modèle stable, on aura l'assurance que la longueur de recouvrement choisie est adaptée et permet d'identifier un modèle qui décrit bien le processus de génération de données²⁷. Dans ce cas, les bénéfices en production seront les faibles révisions. En effet, un modèle mal identifié produirait des prévisions de mauvaises qualités, qui dans le cadre d'une

26. Un autre critère à considérer est la moyenne des erreurs de prévision au carré (MSE). Toutefois dans le cas des séries de l'IPI l'ajout de ce critère ne modifie pas la conclusion.

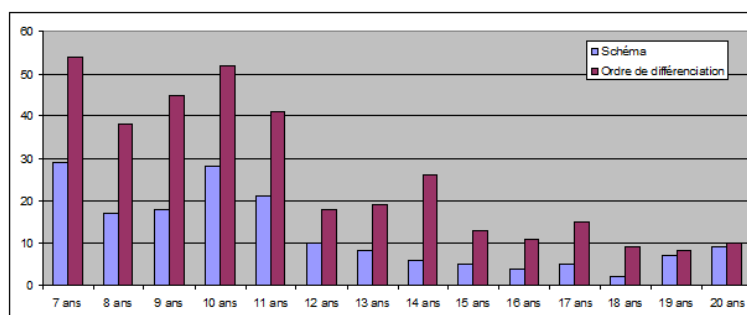
27. En toute rigueur un modèle identifié de manière stable dans le temps ne prouve pas l'absence de biais. Toutefois, sans élément indiquant l'existence d'un biais, une période d'estimation produisant un modèle stable est préférable à la situation où la structure et les ordres du modèle varient sur des fenêtres de données consécutives.

production mensuelle induirait des révisions importantes à chaque nouveau point ajouté. En outre, lors d'une campagne annuelle on pourrait également avoir de fortes révisions liées à la ré-identification du modèle.

Dans quels cas l'identification d'un modèle serait instable ? Une hypothèse raisonnable serait que sur les fenêtres courtes, on risque de ne pas disposer suffisamment de données pour identifier correctement un modèle et cela résulterait par l'instabilité des paramètres estimés sur les fenêtres successives. Sur les séries de l'IPI, en comparant les paramètres estimés sur une fenêtre à ceux estimés sur la fenêtre précédente, on constate effectivement que sur les fenêtres courtes de 6 à 10 ans la fréquence des changements de paramètres est plus importante.

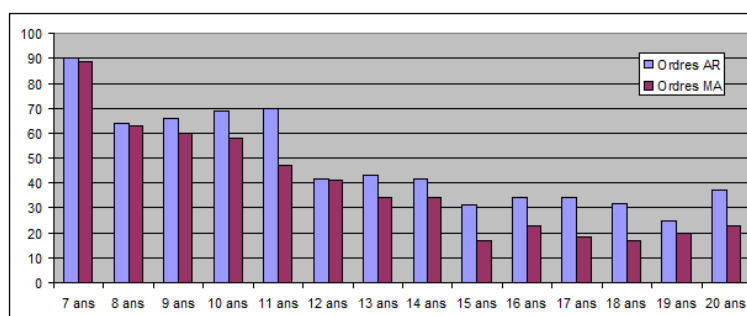
Les deux paramètres qui nous semblent les plus importants pour définir la stabilité sont ceux qui permettent de stationnariser les séries, à savoir l'ordre de différenciation dans un modèle ARIMA et la transformation logarithmique, c'est-à-dire le schéma de composition. Le graphique 11 montre que pour les séries de l'IPI, le schéma ainsi que l'ordre de différenciation changent plus fréquemment sur les fenêtres courtes jusqu'à 10 ans et deviennent plus stables à partir de 11 ans.

GRAPHIQUE 11 – Fréquence de changement du schéma et des ordres de différenciation par fenêtre



Note de lecture : Sur la fenêtre des 7 ans, par rapport à la fenêtre précédente (6 ans), 29 séries ont changé de schéma de composition et 54 d'ordre de différenciation.

GRAPHIQUE 12 – Fréquence de changement des ordres ARMA par fenêtre



Note de lecture : Sur la fenêtre des 7 ans, par rapport à la fenêtre précédente (6 ans), 90 séries ont changé d'ordre AR et 89 d'ordre MA.

Lorsque l'on considère les variations des ordres AR et MA, on obtient le même constat (graphique 12). Ces paramètres deviennent plus stables à partir de 11 ans. Concernant la correction des jours ouvrables par l'introduction des régresseurs dans le modèle, les résultats du test de significativité changent également plus fréquemment sur les fenêtres courtes et deviennent plus stables à partir de 12 ans.

Dans nos simulations sur les séries de l'IPI, les paramètres que nous retenons comme critère de stabilité des modèles sont le schéma de composition, l'ordre de différenciation, l'ordre AR, ainsi que l'ordre MA. Les deux premiers types de paramètres sont prioritaires, c'est-à-dire qu'en cas de l'absence de stabilité simultanée des 4 paramètres, la stabilité du schéma et de la différenciation contribue au choix des fenêtres²⁸. La stabilité des ordres ARMA est importante dans la mesure où le modèle ARIMA sert de cadre pour la détection des *outliers*, ainsi que pour l'estimation des coefficients relatifs aux effets de jours ouvrables. Un modèle ARMA stable favorise la stabilité des autres paramètres de pré-ajustement.

Un autre paramètre important dans les modèles Reg-ARIMA est la correction des jours ouvrables. Cependant nous ne retenons pas ce paramètre dans le critère de stabilité, dans la mesure où dans une procédure habituelle de production le choix de corriger ou non des effets de jours ouvrables et le choix du jeu de régresseurs devraient être faits en amont du choix de la longueur de recouvrement.

Le critère de stabilité induit la sélection de fenêtres contiguës, qui forment une plage de fenêtres avec un modèle « stable ». Un modèle est stable si une plage comporte au moins 3 fenêtres. Ainsi sur 15 fenêtres examinées, nous sommes susceptible de trouver plusieurs plages de stabilité. Dans ce cas, nous retenons la plage comportant les fenêtres les plus courtes qui devrait mieux refléter la structure de la fin de série.

2.4 Description de l'algorithme et résultats sur les séries de l'IPI

L'algorithme de sélection s'appuie en priorité sur les indicateurs de qualité résumés dans le score. La première étape consiste à identifier pour chaque série les fenêtres pour lesquelles les scores sont inférieurs ou égaux à un seuil, puis à les regrouper en plages de fenêtres contiguës. Le seuil du score est progressivement augmenté tant que l'on ne dispose pas au moins d'une plage de 3 fenêtres contiguës²⁹. La contrainte d'avoir des fenêtres contiguës permet de s'assurer d'une certaine stabilité de la qualité. L'idée est que les longueurs de recouvrement proches devraient produire une désaisonnalisation de qualité similaire. Cette condition permet de poser l'hypothèse selon laquelle prolonger la fin de série d'une année, en ajoutant l'année 2018 en l'occurrence, ne devrait pas modifier sensiblement la qualité de la désaisonnalisation.

Après la première étape, nous disposons pour chaque série de l'IPI, une ou plusieurs plages de fenêtres contiguës où le score est « minimal ». L'étape suivante consiste à extraire de ces plages, les fenêtres qui répondent en plus au critère de stabilité du modèle Reg-ARIMA. À l'issue de cette deuxième étape on obtient ainsi pour chaque série une plage plus restreinte de fenêtres, qui sont toutes équivalentes au regard des critères de qualité et de stabilité. Enfin, les fenêtres longues de 6 et de 7 ans sont exclues dans tous les cas afin de respecter la longueur minimale de 8 ans recommandée par les *guidelines*³⁰.

Finalement pour chaque série, on sait déterminer une plage de fenêtres contiguës, qui produiraient des désaisonnalisations avec des modèles similaires et de qualité proche (voir les exemples de séries dans l'annexe D). Les plages de fenêtres ainsi déterminées sur les 205 séries de l'IPI sont partiellement ou totalement disjointes. Aucune des 15 fenêtres est incluse dans l'ensemble des

28. Une autre alternative consiste à retenir comme critère de stabilité seulement le schéma de composition et l'ordre de différenciation. Dans ce cas l'hypothèse retenue serait que les ordres AR et MA identifiés de manière automatique sur des fenêtres proches, même différents, auraient une incidence négligeable sur les résultats de pré-ajustement de la série brute et sur la prévision. La conséquence de ce choix est que la plage de stabilité est plus large.

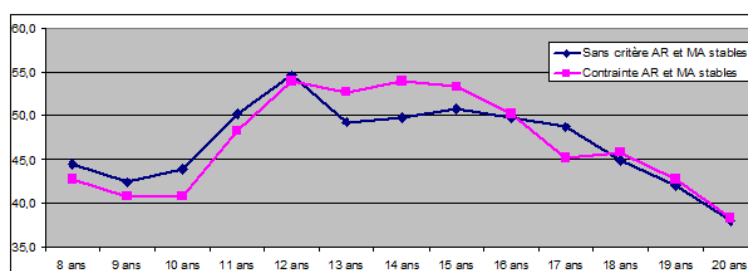
29. Le seuil est par défaut égal au plus petit score obtenu sur les 15 fenêtres de données. En général ce minimum est 0, c'est-à-dire absence de toute anomalie.

30. Une autre alternative consiste à imposer une longueur minimale de 11 ans afin de s'assurer qu'au point de coupure le coefficient saisonnier soit estimé avec un filtre symétrique.

205 pages. Or, en production il est souhaitable de ne choisir qu'une longueur de recouvrement pour l'ensemble des séries de l'IPI. Le choix consiste dans ce cas à retenir la fenêtre qui apparaît le plus souvent dans les 205 pages.

Ainsi, la fenêtre optimale pour désaisonnaliser la fin de séries de l'IPI serait alors longue de 12 ans, soit 6 ans de recouvrement. Cependant, on note qu'il existe un plateau constitué des fenêtres 12 à 15 qui ont des fréquences proches (graphique 13). Dans le cas où les ordres AR et MA ne font pas partie du critère de stabilité, la fenêtre de 12 ans se démarque davantage des autres³¹.

GRAPHIQUE 13 – Fréquence relative des fenêtres retenues dans les plages, selon le critère de stabilité



Note de lecture : La fenêtre de 8 ans apparaît 44 fois sur 100 si le critère de stabilité ne contient pas les ordres AR et MA. Cette fréquence est de 43 fois sur 100 dans le cas contraire.

Enfin, on pourrait aussi souhaiter choisir une fenêtre optimale pour chaque série. Partant des plages de fenêtres déterminées précédemment, on pourra ne retenir qu'une seule fenêtre à l'aide d'un critère supplémentaire. Nous produisons ici deux options possibles, l'une consiste à considérer le niveau de révisions des CVS produit par chaque fenêtre³² et à choisir celle qui produit la plus petite révision. L'autre option consiste à retenir la fenêtre qui produit la plus petite valeur de l'AICC³³. Dans les deux options on trouve que la fenêtre la plus souvent retenue est celle qui correspond à 8 ans (graphique 14), soit la longueur minimale recommandée par les *guidelines*.

3 Comment raccorder les CVS de deux sous-périodes au point de coupure

Dans la section 2.1 nous avons vu que selon les recommandations des *guidelines*, il suffirait de couper les séries de l'IPI en janvier 2012 et mettre à jour les 6 dernières années³⁴. Une autre question à laquelle il convient de répondre est comment raccorder les CVS du passé et les CVS

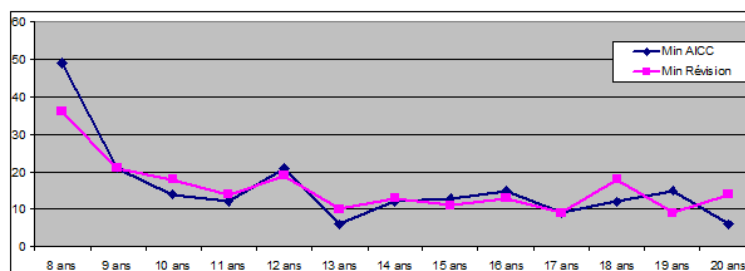
31. Si l'on ajoute dans le score le critère de la moyenne des erreurs de prévision au carré, le plateau des fenêtres de 12 à 15 est davantage marqué.

32. Les révisions des CVS sont ici étudiées uniquement sur la dernière année, c'est-à-dire l'année 2017. Pour cela le modèle identifié jusqu'en décembre 2017 est ré-estimé à chaque mois entre janvier et décembre 2017. Cela permet de mesurer la différence entre la première et la dernière estimation connue des CVS (qui est encore provisoire). L'indicateur retenu correspond alors à la moyenne des différences relatives en niveau entre la première et la dernière estimation.

33. Rappelons que les modèles relatifs à chaque fenêtre sont évalués sur les mêmes données (la fin de série) et que les plages retenues ont un schéma de composition et un ordre de différenciation stable. Par conséquent les valeurs du critère AICC sont comparables.

34. Cela signifie que tous les points CVS antérieurs à janvier 2012 sont figés, sauf en cas de révision des données brutes.

GRAPHIQUE 14 – Fréquence des fenêtres retenues pour chaque série, selon le critère AICC ou de révision



Note de lecture : La fenêtre de 8 ans est choisie pour 49 séries sur les 205 dans le cas de l'AICC. Elle est choisie pour 36 séries quand on utilise le niveau de révisions.

mises à jour sur la fin de série, le point de raccordement étant le point de coupure en janvier 2012.

Les CVS étant estimées avec une période de recouvrement de chaque côté, nous disposons donc de deux estimations du point de décembre 2011, l'une provenant de la précédente estimation (qui sera par la suite figée) et l'autre de l'estimation de la CVS de la fin de série. Il y a peu de chance que ces deux estimations du point de décembre 2011 soient au même niveau. Il en est de même pour le point de janvier 2012, deux estimations bien distinctes. Ainsi, les évolutions mensuelles entre décembre 2011 et janvier 2012 seront également différentes selon si l'on considère la CVS du passé ou la CVS de la fin de série. Raccorder signifie donc gérer la rupture en niveau entre les deux estimations et aussi faire le choix d'une évolution à imputer au point de janvier 2012. Nous présenterons dans cette partie trois façons d'effectuer le raccordement.

Enfin, nous verrons également que l'importance de la rupture en niveau et les différences sur les évolutions dépendent beaucoup de la longueur de la période de recouvrement retenue pour estimer les CVS de chaque sous-période.

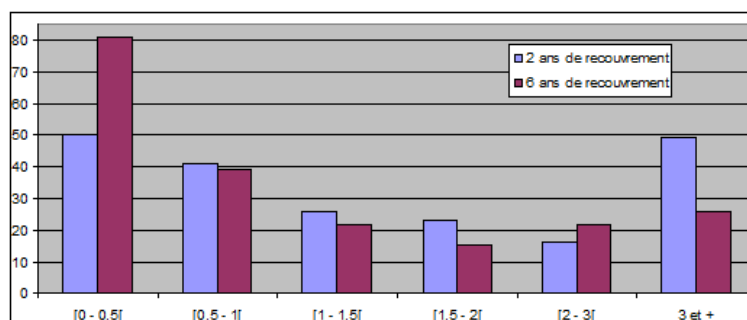
3.1 Concaténer les CVS

La première méthode consiste simplement à concaténer les CVS des deux sous-périodes. La CVS du passé est publiée jusqu'au point de décembre 2011, puis à partir du point de janvier 2012 on publie la CVS obtenue avec le nouveau modèle sans modifier son niveau. Cette méthode suppose que l'on considère la rupture du modèle comme une composante de l'évolution entre décembre 2011 et janvier 2012. L'inconvénient de cette méthode est que les valeurs autour du point de raccordement ne sont plus directement comparables, les évolutions entre 2011 et 2012 reflètent en partie un changement de méthode statistique. Son avantage est qu'elle conserve les niveaux des séries CVS. Elle est donc à privilégier lorsque l'on s'intéresse aux données en niveau et les évolutions entre les années autour du point de raccordement ne sont plus ou pas utilisées.

La concaténation sera d'autant plus « douce » que les niveaux des deux estimations au point de janvier 2012 et celles au point de décembre 2011 sont proches³⁵. Or, plus la période de recouvrement est longue moins ces estimations présentent d'écart. Les résultats de la section 2.4 suggèrent d'estimer la fin de série avec 6 ans de recouvrement sur le passé si l'on considère les intersections des plages de fenêtres ou avec 2 ans seulement si le choix d'une fenêtre optimale pour chaque série s'appuie sur les critères d'AICC et des révisions. Les graphiques 15 et 16 montrent qu'avec 6 ans de recouvrement les estimations au point de janvier 2012 présentent globalement de plus faibles écarts en niveau et en évolution.

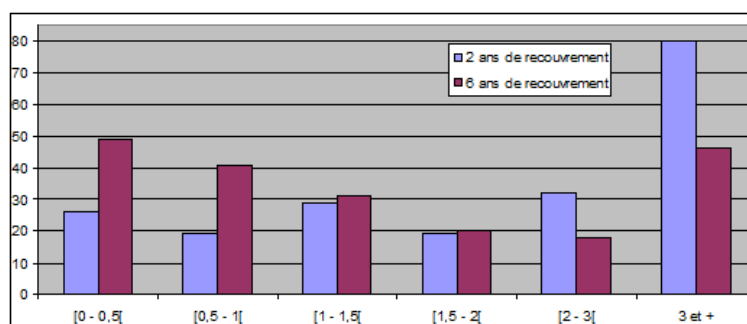
35. Cela implique que les évolutions autour du point de coupure sont proches.

GRAPHIQUE 15 – Fréquence des écarts sur les estimations du niveau du point de janvier 2012 par intervalle



Note de lecture : Avec 2 ans de recouvrement 50 séries présentent un écart en valeur absolue de moins de 0,5 point entre les deux estimations (CVS passé et CVS fin de série) du point de janvier 2012. Avec 6 ans de recouvrement le nombre de séries entrant dans ce critère est de 81.

GRAPHIQUE 16 – Fréquence des écarts sur les évolutions décembre 2011 - janvier 2012 par intervalle



Note de lecture : Avec 2 ans de recouvrement 26 séries présentent un écart en valeur absolue de moins de 0,5 point entre les deux estimations (CVS passé et CVS fin de série) de l'évolution entre décembre 2011 et janvier 2012. Avec 6 ans de recouvrement le nombre de séries entrant dans ce critère est de 49.

3.2 Chaîner sur une période commune

Il s'agit ici de chaîner les évolutions mensuelles de la CVS estimée sur la fin de série à partir du point de décembre 2011 de la CVS du passé. Le coefficient de raccordement est calculé comme le rapport entre les valeurs du point de décembre 2011 issues de la CVS du passé et de la CVS courante. Cette méthode a l'avantage de permettre l'interprétation des évolutions mensuelles entre décembre 2011 et janvier 2012, puisque les évolutions sont à modèles constants. L'inconvénient est qu'elle peut créer une rupture sur la tendance de la CVS : un décalage en niveau est alors créé entre la série brute et la série CVS³⁶. Cette méthode est à privilégier lorsque les statistiques d'intérêt sont les évolutions mensuelles et non le niveau de la série. C'est le cas des séries d'indices en général, et de l'IPI en particulier.

Comme dans le cas de la concaténation, avec le chaînage la rupture en niveau est d'autant plus forte que les séries CVS sont différentes autour du point de raccordement. Ces différences sont en outre renforcées lorsque les *outliers* détectés sont différents sur les périodes de recouvrement entre les modèles « Passé » et « Courant »³⁷. Ainsi, plus la période de recouvrement est longue, plus la rupture créée par le raccordement est faible.

36. Utiliser cette méthode de raccordement revient en effet à effectuer une translation en niveau à partir de janvier 2012 de la CVS estimée sur la fin de série.

37. C'est le cas en particulier des changements de niveau (*Level Shift*), lorsqu'ils sont détectés à proximité du point de coupure dans un modèle et pas dans l'autre.

3.3 Chaîner sur une année commune

La différence avec le cas précédent est que le coefficient de raccordement est calculé en utilisant une année commune entre les deux CVS, en l'occurrence l'année 2011. Il s'agit du rapport des moyennes de l'année 2011 de la CVS du passé et de la CVS courante. Cette méthode a l'avantage de préserver la cohérence entre le niveau des CVS et des séries brutes. L'accroissement moyen annuel entre 2011 et 2012 est cohérent et n'est pas affecté par la rupture de modèle.

Conclusion

En somme, désaisonnaliser des séries très longues (de plus de 20 ans) par sous-période améliore bien la qualité du processus. Cela provient notamment du fait que l'hypothèse de stabilité des modèles Reg-ARIMA n'est plus vérifiée sur des très longues séries. Ainsi, la phase de pré-ajustement du processus de désaisonnalisation est améliorée en désaisonnalisant par sous-période : la correction des effets de jours ouvrables est améliorée ; le modèle obtenu est de meilleure qualité et reflète davantage la dynamique des données et les risques de révisions liées aux mauvaises prévisions du modèle Reg-ARIMA sont diminués. En revanche, même si les séries CVS produites sont différentes, la qualité de la décomposition est inchangée. Ces résultats sont conformes avec les *guidelines* sur l'ajustement saisonnier qui recommandent de désaisonnaliser par sous-période les séries de plus de 20 ans.

Sur la période récente l'enjeu n'est pas sur la date de coupure (i.e. date à partir de laquelle les séries sont révisées) : cette date peut être fixée en suivant les recommandations des sections 4.3 et 6.2 des *guidelines*. Dans le cas de l'IPI, la date de coupure serait janvier 2012³⁸ et elle devrait être maintenue pendant au plus 5 ans. L'enjeu est plutôt sur la longueur de la période de recouvrement à utiliser pour estimer au mieux la fin de la série. Ce choix peut s'appuyer à la fois sur les connaissances *a priori* des données (connaissances de ruptures, contraintes de diffusion, etc.) et sur la prise en compte de critères statistiques. L'algorithme que nous proposons dans cet article permet d'aider à la prise de décision en fournissant des fenêtres qui permettent d'estimer au mieux cette période. À partir de ces fenêtres de données où les modèles sont statistiquement équivalents, d'autres critères peuvent intervenir pour le choix final de la date de début d'estimation, i.e. la longueur de recouvrement : arbitraires (exemple choix de la plus petite fenêtre) ou statistiques (critères d'information ou sur les révisions).

Dans le cadre des séries de l'IPI il ressort de l'algorithme deux longueurs de recouvrement possibles en fonction des critères retenus : 2 ans (i.e. un modèle estimé sur 8 ans, depuis janvier 2010) et 6 ans (i.e. un modèle estimé sur 12 ans, depuis janvier 2006). Ainsi, si l'on s'intéresse à la fin de série, estimer un modèle sur 16 ans, comme testé dans la partie 1, paraît trop long. Entre les deux longueurs de recouvrement possibles, nous recommandons d'utiliser le modèle estimé sur 12 ans. Cela permet d'une part de s'assurer que des filtres symétriques sont utilisés pour l'estimation des CVS au point de janvier 2012 ; d'autre part cela minimise la rupture lors du raccordement des CVS au point de coupure. En revanche, certains modèles Reg-ARIMA ainsi estimés pourraient être affectés par la crise de 2008 et par le changement de nomenclature de 2009. Toutefois, il convient de relativiser l'impact de ce type d'évènements sur la modélisation (cf. annexe B).

Précisons cependant que les résultats empiriques de cette étude, en particulier le choix de la date de coupure en janvier 2012 et la longueur de recouvrement de 6 ans, ne sont évidemment applicables que pour les séries de l'IPI et ne sont pas généralisables.

Enfin, la détermination des plages de fenêtres pour chaque série et le choix d'une fenêtre pour l'ensemble des séries dépendent fortement des critères de qualité et de stabilité que nous avons retenus. Les choix sont donc relatifs à ces critères, qui peuvent être contestés et revus. Par ailleurs, la méthode est testée sur les séries de l'IPI, mais elle est évidemment applicable à toutes les séries longues, à condition de bien examiner les critères à retenir et qu'ils soient pertinents au regard de la nature des données étudiées. Des études similaires seraient intéressantes à mener sur d'autres types de données (les indices de chiffres d'affaires, données des enquêtes de conjoncture, etc.).

38. Cet article n'émet aucune décision relative au choix de production des séries de l'IPI. Nous ne signalons ici que l'option qui répond au mieux aux recommandations européennes. Le choix relatif à la date de coupure effective des séries de l'IPI appartient évidemment à la maîtrise d'ouvrage.

A Expression et spécification des modèles Reg-ARIMA

On adopte les notations suivantes :

Y_t série temporelle observée,

O_{it} régresseurs qui modélisent les outliers,

C_{jt} régresseurs qui modélisent les effets de calendrier.

Un modèle Reg-ARIMA s'écrit de la façon suivante :

$$Y_t = \sum \hat{\alpha}_i O_{it} + \sum \hat{\beta}_j C_{jt} + X_t$$

X_t résidus de régression qui suivent un modèle $ARIMA(p, d, q)(P, D, Q)$.

$$\Phi(B)\Phi_s(B)(I - B)^d(I - B^s)^D X_t = \Theta(B)\Theta_s(B)\varepsilon_t$$

Avec :

B opérateur retard,

s périodicité de la saisonnalité, 4 dans le cas trimestriel et 12 dans le cas mensuel,

ε_t un bruit blanc,

$\Phi(B)$ polynôme autorégressif (AR) en B d'ordre p ,

$\Phi_s(B)$ polynôme autorégressif en B^s d'ordre P ,

$\Theta(B)$ polynôme moyenne mobile (MA) en B d'ordre q ,

$\Theta_s(B)$ polynôme moyenne mobile en B^s d'ordre Q ,

Dans cet article, les spécifications retenues pour estimer les modèles Reg-ARIMA de chaque série et sur chaque fenêtre sont similaires à celles retenues pour la production, à savoir :

- les effets de pâques sont exclus ;
- les effets de jours ouvrables estimés avec le jeu de 7 régresseurs, ce dernier est retenu que s'il est significatif³⁹ ;
- les outliers détectés sont les additive outliers (AO), transitory changes (TC), level shifts (LS), avec un seuil de détection défini par défaut dans JDemetra+ ;
- les modèles ARIMA sont identifiés et estimés de manière automatique.

Ainsi, dans le cadre de nos tests sur les séries coupées la spécification retenue est cohérente avec ce qui est mis en oeuvre en production, c'est-à-dire que les paramètres pré-fixés sont généralement limités. Enfin, avec 205 séries et 15 fenêtres d'estimation à comparer, cela donne 3 075 modèles différents. Il va de soi que nous n'avons pas engagé un travail en vue de valider chacun de ces modèles et d'effectuer des ajustements de spécification à la main. Le but étant de trouver une fenêtre d'estimation qui donnerait globalement les meilleurs modèles au regard des critères de qualité et de stabilité que nous avons définis. En production, le choix de la fenêtre d'estimation sera suivi logiquement d'un travail plus précis sur les modèles.

B Choix a priori d'une fenêtre d'estimation

On peut poser la réserve suivante : a-t-on besoin d'un algorithme afin de déterminer la longueur d'estimation ? Ne suffit-il pas de choisir une date qui correspond au dernier changement de nomenclature par exemple ?

Bien entendu, le choix peut être guidé par des informations sur les changements de gestion des données, sur les modifications de méthodes statistiques, sur les chocs connus de nature économique ou institutionnelle ou encore il peut simplement résulter de contraintes de production.

39. L'algorithme utilisé actuellement en production pour choisir le jeu de régresseurs est également initialisé de cette façon. Si le jeu de 7 régresseurs s'avère significatif, alors la procédure se poursuit en comparant ce jeu complet aux autres jeux de régresseurs plus parcimonieux, i.e. ayant moins de variables.

Autrement dit, le choix peut être effectué uniquement à l'aide d'informations qualitatives. Dans ce cas le choix sera postulé et décidé indépendamment du mécanisme de génération des données.

Considérons par exemple les changements de base et de nomenclature qui ont eu lieu en 2009 (annexe E). On peut supposer que ces événements sont causes plausibles de changement de modèles. Par conséquent il suffit de choisir comme date de début d'estimation janvier 2009 pour identifier et estimer un modèle sur la fin de série, il s'agit donc de retenir la fenêtre temporelle, janvier 2009 à décembre 2017, longue de 9 ans. Toutefois, avons nous les moyens de valider une telle hypothèse de travail ? Avons nous constaté que effectivement les changements de base et de nomenclature en 2009 ont entraîné des ruptures et/ou des changements de modèles en 2009 ?

B.1 Recherche de ruptures en 2009

En considérant la structure d'un modèle Reg-ARIMA (annexe A), nous allons examiner chaque composante afin de vérifier si elle a un comportement particulier en 2009.

On peut exclure les outliers comme composante pertinente qui justifie que l'on retienne 2009 comme date de début d'estimation. En effet, les outliers de type AO et TC produisent des chocs ponctuels qui n'affectent pas la structure des données. Un changement de niveau de type LS ne fait que décaler le niveau des séries sans changer leur profil. Un changement de profil saisonnier de type SO (Seasonal Outlier) n'est pas motif à couper les séries, car il suffirait de modéliser ce type de rupture dans le modèle Reg-ARIMA ⁴⁰.

Concernant les effets de jours ouvrables, l'annexe C montre que l'on peut distinguer 3 types d'évolution des coefficients de correction des jours ouvrables : une rupture, une dérive linéaire ou des fluctuations sans changement de niveau pérenne. Les dérives linéaires ou des fluctuations ne permettent pas d'attribuer un statut particulier à l'année 2009 relativement à d'autres années. Par ailleurs on n'observe pas de rupture des effets de calendrier en 2009 sur les séries de l'IPI ⁴¹. Les changements de nomenclature et de base en 2009 ne semblent pas affecter les effets de calendrier.

Concernant le modèle ARIMA, est-ce que les changements de nomenclature et de base en 2009 ont modifié la structure d'autocorrélation des séries de l'IPI ? Il est assez peu plausible que cela puisse affecter les ordres de différenciation, une tendance ou une composante saisonnière n'apparaît pas ou ne disparaît pas par simple changement de nomenclature et/ou de base. Par ailleurs, identifier de manière stable la structure d'autocorrélation d'une série (ordres ARMA) nécessite 11 à 12 ans de données comme nous avons vu dans la section 2.3.2, il est donc difficile de conclure qu'il se produit un changement brutal de la structure des données à partir de 2009. Un changement de nomenclature a eu lieu également en 2003, mais la conclusion est la même on n'observe pas de changement plus fréquent des ordres du modèle à cette date - le graphique 11 ne montre pas de comportement particulier des ordres de différenciation autour de la fenêtre de 15 ans, qui correspond à 2003, il en est de même pour les ordres AR et MA.

40. Couper une série au point de rupture de type SO ne corrige pas le changement de profil saisonnier dès lors que l'on prend une période de recouvrement. Si on coupe sans période de recouvrement on induit un déphasage du fait de l'utilisation d'un filtre asymétrique, ce qui n'est pas une bonne option. Par ailleurs les guidelines recommandent de prendre une période de recouvrement et suffisamment longue. Il est donc de loin préférable de modéliser un SO dans le modèle Reg-ARIMA.

41. Les résultats de tests de rupture sur les coefficients de jours ouvrables montrent que 2009 n'est pas une année remarquable. Seules 22 % des séries présentent une rupture en 2009. Les mêmes tests sur les années 2008 et 2010 donnent des résultats similaires, respectivement 24 % et 20 %, et ce sont essentiellement les mêmes séries.

B.2 Conclusion

Bien qu'il soit raisonnable de faire un choix a priori d'une fenêtre d'estimation en s'appuyant sur les informations externes aux données, comme les évènements de gestion, cela reste néanmoins une hypothèse qu'il convient de confronter aux données. La démarche a priori et le choix fondé sur l'étude des données ne sont pas nécessairement deux solutions exclusives. Il convient cependant de ne pas confondre les hypothèses avec les résultats au risque d'exclure trop rapidement d'autres alternatives qui pourraient mieux orienter nos choix. Le développement présenté dans cet annexe a pour but de montrer que l'affirmation selon laquelle un changement de nomenclature et de base entraîne nécessairement une rupture dans les modèles relève avant tout d'une hypothèse et non d'un résultat que l'on est en mesure d'observer avec les moyens et méthodes que l'on dispose. « L'évidence première n'est pas une vérité fondamentale »⁴².

C Changements de profil et conséquences sur la modélisation

Nous examinons ici de quelle manière le profil changeant des séries agit sur les paramètres d'un modèle. On s'appuiera sur des exemples bien choisis des séries de l'IPI en vue d'illustrer des situations que l'on rencontre en pratique et leurs conséquences sur l'estimation des paramètres du modèle. La date de coupure est janvier 2005 avec 3 ans de recouvrement, comme dans la première partie de l'article et qui correspond au choix a priori de la maîtrise d'ouvrage.

C.1 Changements sur les effets du calendrier

Couper les séries de l'IPI améliore la correction des jours ouvrables, à quoi tient ce résultat ? Nous avons déjà souligné que sur longue période il est raisonnable de penser que les effets du calendrier sur les séries économiques évoluent dans le temps. Par conséquent estimer un coefficient correcteur constant sur longue période donnera des résultats sous-optimaux. Nous verrons dans quels cas précisément couper une série améliore la correction des effets du calendrier, et quels gains en efficacité on peut espérer selon le profil d'évolution de ces effets.

Pour simplifier la lecture et la compréhension des résultats, nous modélisons ici les effets de calendrier par un coefficient unique, qui correspond à l'effet de contraste entre les jours de semaine et les week-end et jours fériés. Pour les séries de l'IPI ce choix a du sens dans la mesure où pour la majorité des branches les activités de production sont peu différenciées entre les jours de semaine. Afin d'observer les variations des coefficients du calendrier dans le temps, ces derniers sont estimés de manière successive sur les fenêtres de 8 ans en glissement. Ensuite, on peut classer les profils d'évolution de ces coefficients en trois catégories :

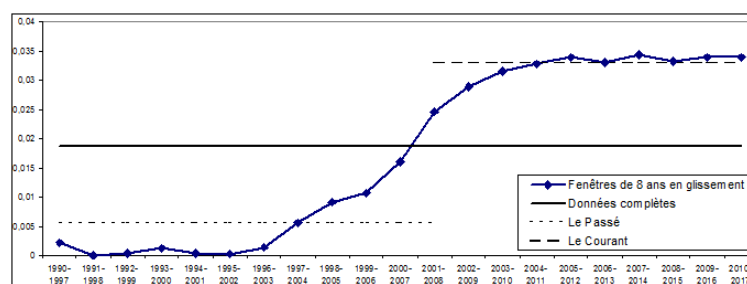
- rupture des coefficients, caractérisée par un niveau stable sur le début de la série, suivi d'un changement brutal de niveau, puis de nouveau un palier stable sur la fin de série ;

42. « Il suffit que nous parlions d'un objet pour nous croire objectifs. Mais, par notre premier choix, l'objet nous désigne plus que nous ne le désignons et ce que nous croyons nos pensées fondamentales sur le monde sont souvent des confidences sur la jeunesse de notre esprit. Parfois nous nous émerveillons devant un objet élu ; nous accumulons les hypothèses et les rêveries ; nous formons ainsi des convictions qui ont l'apparence d'un savoir. Mais la source initiale est impure : l'évidence première n'est pas une vérité fondamentale. En fait, l'objectivité scientifique n'est possible que si l'on a d'abord rompu avec l'objet immédiat, si l'on a refusé la séduction du premier choix, si l'on a arrêté et contredit les pensées qui naissent de la première observation. Toute objectivité, dûment vérifiée, dément le premier contact avec l'objet. Elle doit d'abord tout critiquer : la sensation, le sens commun, la pratique même la plus constante, l'étymologie enfin, car le verbe, qui est fait pour chanter et séduire, rencontre rarement la pensée. » BACHELARD.

- dérive linéaire des coefficients, caractérisée par une évolution d’amplitude stable et continue dans le temps ;
- multiple changements de sens des coefficients sur la période étudiée.

Le cas d’une rupture est illustré ici par l’exemple de la série IPI 2920, industrie de la fabrication de carrosseries et de remorques (graphique 17)⁴³. Le coefficient constant estimé sur toute la période 1990-2017 (trait plein continu dans le graphique) est éloigné des coefficients estimés sur le début et la fin de séries. La série CVS ainsi obtenue présente des effets de jours ouvrables résiduels. L’estimation des coefficients par sous-période donne de meilleurs résultats et supprime les effets résiduels. Sur la sous-période « le Courant » en particulier, le coefficient estimé (trait discontinu) correspond parfaitement à ceux de la fin de série, qui est de niveau constant.

GRAPHIQUE 17 – Rupture du coefficient jours ouvrables, exemple de la branche 2920 (industrie de la fabrication de carrosseries et de remorques)



L’exemple de la série IPI 2829, fabrication de machines diverses d’usage général (graphique 18), illustre le cas d’une dérive linéaire des coefficients dans le temps. Le constat est identique que précédemment. En effet, les jours ouvrables résiduels sont présents sur la CVS estimée sur la série longue. La correction par sous-période supprime les effets résiduels. Toutefois, dans le cas d’une dérive linéaire par rapport à la situation précédente, les coefficients estimés sur les deux sous-périodes présentent des écarts plus marqués avec ceux attendus pour le début et la fin de série. Ce qui est naturel puisque l’on estime un coefficient constant pour approximer un effet qui évolue, même sur courtes périodes⁴⁴.

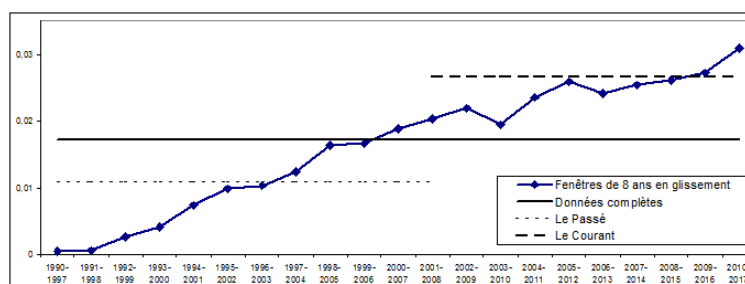
Enfin en cas de fluctuations des coefficients, où les effets de calendrier changent fréquemment de niveau et de sens d’évolution, couper une série en deux sous-périodes ne permet pas nécessairement d’effectuer une meilleure correction. L’exemple de la branche tréfilage à froid (IPI 2434, graphique 19) montre que le coefficient estimé à l’aide des données complètes est proche de ceux estimés sur la série coupée. Entre les deux sous-périodes, le niveau moyen des coefficients sont proches, bien que les coefficients estimés sur les fenêtres glissantes ne sont pas constants dans le temps.

Ces trois exemples montrent que couper une série longue n’améliore réellement la correction des jours ouvrables que dans la mesure où il existe en moyenne une différence de niveau entre les deux sous-périodes. Cela est vrai en particulier dans le cas de rupture des coefficients de calendrier ou de dérive linéaire.

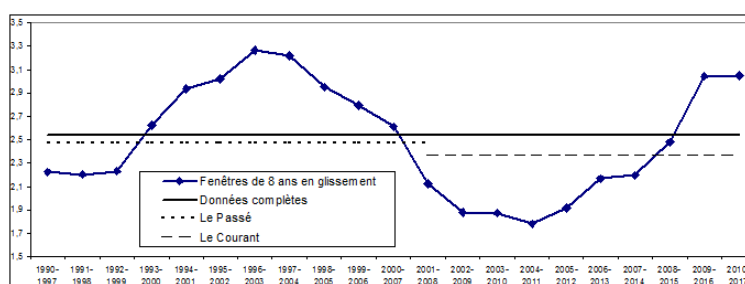
43. Une rupture se produisant brutalement sur une année apparaît nécessairement de manière lissée dans le graphique. Le calcul des coefficients sur des fenêtres glissantes de 8 ans implique que le changement de niveau se répercute de manière progressive dans 8 fenêtres.

44. On peut noter que dans le cas de rupture ou de dérive linéaire, estimer un coefficient par sous-période revient à modéliser une rupture par changement de régime, puisque deux coefficients constants de différents niveaux en sont les produits. Si l’objectif est seulement de prendre en compte le changement de niveau des coefficients, il suffirait d’introduire un régresseur supplémentaire dans la modélisation Reg-ARIMA sans qu’il soit nécessaire de couper la série. La décision de couper une série suppose donc que les autres paramètres du modèle sont également estimés de manière sous-optimale sur les séries longues.

GRAPHIQUE 18 – Dérive linéaire du coefficient jours ouvrables, exemple de la branche 2829 (fabrication de machines diverses d’usage général)



GRAPHIQUE 19 – Multiple changements de sens d’évolution du coefficient jours ouvrables, exemple de la branche 2434 (branche tréfilage à froid)



C.2 La détection des *outliers*

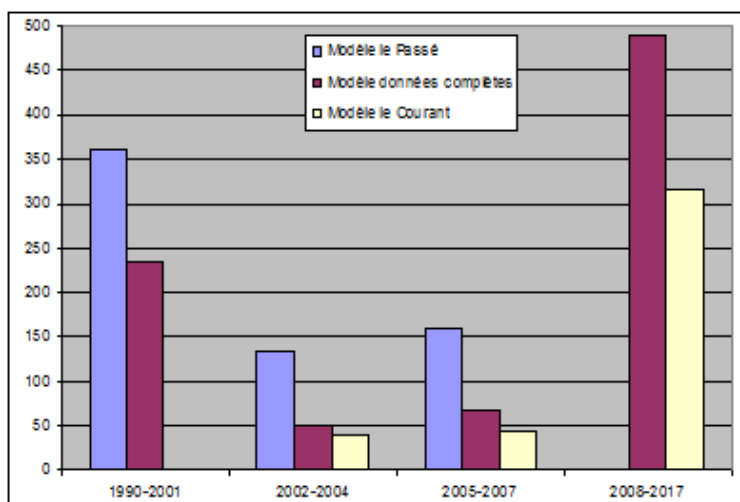
Dans la phase de pré-ajustement la détection et la correction des *outliers* dépendent des autres paramètres du modèle, le schéma de composition, la correction des jours ouvrables, ainsi que le modèle ARIMA. Couper une série induit des changements de paramètres de modèle et donc l’appréciation du caractère atypique d’un point. Pour une même fenêtre de données le nombre d’*outliers* détectés peut varier si le modèle que l’on considère est relatif aux données complètes ou à une sous-période. Ainsi, sur les 205 séries de l’IPI, le nombre total d’*outliers* détectés sur la sous-période « le Passé » passe de 352 à 657 lorsque l’on coupe les séries. En revanche, sur la sous-période « le Courant » le nombre total d’*outliers* diminue de 606 à 397 (graphique 20). Globalement le nombre d’*outliers* varie donc fortement lorsque l’on coupe les séries, et selon la sous-période considérée il peut augmenter ou diminuer ⁴⁵.

Nous produisons ici deux exemples afin d’illustrer l’influence des fenêtres de données et les modèles correspondants sur la détection des *outliers*.

Le premier exemple est la série IPI 1052 (fabrication de glaces et sorbets). Graphiquement cette série semble très régulière (graphique 21), pourtant le nombre d’*outliers* détectés s’élève à 30 lorsqu’on utilise le modèle estimé sur les données complètes. Les 30 *outliers* sont concentrés sur la sous-période 2008-2017 (tableau 3). En effet, en examinant la courbe de la série, on devine un changement de profil à partir de 2009. La série est plus volatile, ce qui fait que par rapport à la régularité des données de 1990 à 2008, de nombreux points depuis 2009 apparaissent comme des *outliers*. En revanche, si la détection des *outliers* s’appuie uniquement sur la sous-période « le Courant », on ne trouve plus que 3 *outliers* entre 2008 et 2017.

45. La détection de *outliers* paraît ici peu robuste dans la mesure où elle produit des résultats qui peuvent varier fortement quand certains paramètres du modèle changent. Toutefois, un point est un *outlier* lorsqu’il s’écarte significativement (i.e. relativement à seuil de probabilité) d’une distribution (i.e. un modèle)[1]. Par conséquent, il n’y a pas de raisons que deux modèles différents donnent les mêmes *outliers*.

GRAPHIQUE 20 – Nombre d'*outliers* détectés par sous-période et selon le modèle



GRAPHIQUE 21 – Série IPI branche 1052 (fabrication de glaces et sorbets)

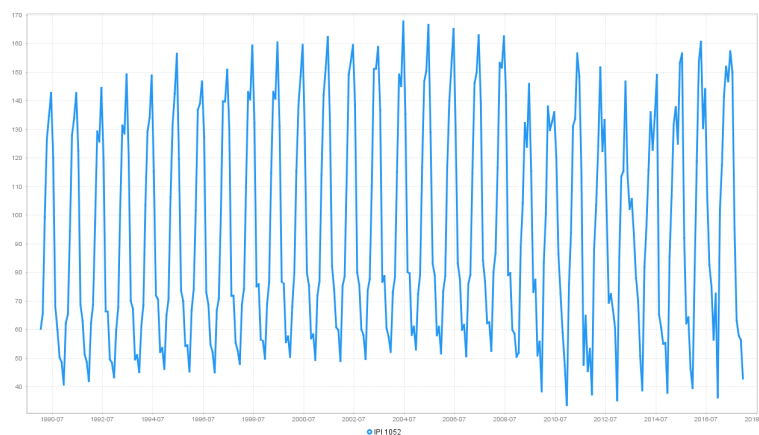


TABLEAU 3 – Nombre de *outliers* de la branche 1052 (fabrication de glaces et sorbets)

	1990-2002	2002-2004	2005-2007	2008-2017	TOT
Données complètes	0	0	0	30	30
Le Passé	0	0	0		0
Le Courant		0	0	3	3

Dans le contexte de notre étude, ce résultat peut être aisément interprété et accepté. En effet, en réduisant la fenêtre de données, le modèle que l'on estime a moins de chance d'être contaminé par un autre mécanisme de génération des données et donc on devrait au total avoir moins de *outliers*. L'exemple suivant va montrer que cette intuition sous-estime la complexité de la modélisation Reg-ARIMA : l'appréciation du caractère atypique d'un point est dépendant de nombreux paramètres⁴⁶.

La série 1412 (fabrication de vêtements de travail) paraît moins lisse que la précédente (graphique 22), cependant le modèle relatif aux données complètes ne détecte que 19 *outliers*, dont 15 sur la sous-période 2008 - 2017 (tableau 4). En coupant la série, la sous-période « le Courant » ne comporte plus aucun *outlier*, tandis que sur la sous-période « le Passé » le nombre d'*outliers* s'accroît de 4 à 29⁴⁷. Ce qui fait qu'au total le nombre d'*outliers* augmente fortement. On remarque également dans cet exemple que sur la période de recouvrement 2002-2007, selon le modèle considéré, le nombre de *outliers* détectés peut être très différent.

GRAPHIQUE 22 – Série IPI branche 1412 (fabrication de vêtements de travail)

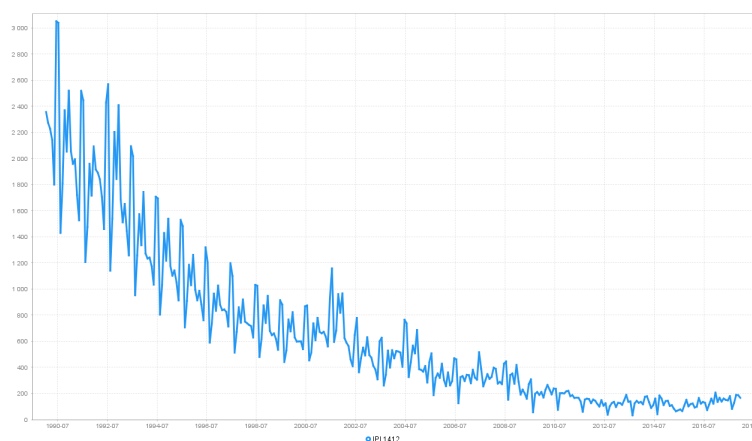


TABLEAU 4 – Nombre de *outliers* de la branche 1412 (fabrication de vêtements de travail)

	1990-2002	2002-2004	2005-2007	2008-2017	TOT
Données complète	0	1	3	15	19
Le Passé	6	9	14		29
Le Courant		0	0	0	0

C.3 L'identification d'un modèle ARIMA

Un modèle ARIMA est une approximation de la structure d'autocorrélation d'une série. On tend à considérer que sur de très longues séries cette approximation finit par s'écarter du processus effectif de génération des données. Il est difficile de formuler une interprétation

46. Les autres paramètres du modèle Reg-ARIMA, mais aussi du seuil dit « critique », qui est le seuil de significativité du test pour décider de classer un point comme *outlier* ou non. Dans JDemetra+ ce seuil critique n'est pas constant, mais dépend de la longueur des données. Plus une série est longue, plus le seuil est haut, i.e. que chaque point aura moins de chance d'être classé comme *outlier*.

47. Dans ce cas précis, le facteur qui influe fortement sur la détection des *outliers* de la sous-période « le Passé » est le modèle ARIMA. Si on prend le même modèle que celui estimé sur les données complètes, le nombre d'*outliers* tombe à 8. En forçant en plus la correction du calendrier le nombre d'*outliers* tombe à 3.

concrète d'un modèle ARIMA et il est encore plus difficile d'expliquer et de pointer les causes de changement d'un tel modèle. Cela d'autant plus que l'on ne peut identifier aisément ce qui relève de l'instabilité des méthodes d'estimation et ce qui est le fait de changements de structure dans les données. En outre, un modèle ARIMA et les autres paramètres de la phase de pré-ajustement (*outliers* et régresseurs jours ouvrables) sont interdépendants⁴⁸, l'identification d'un modèle ARIMA peut donc varier simplement du fait des hypothèses d'estimation⁴⁹, sans qu'il y ait de changement dans les données.

Ainsi, couper une série conduit parfois à des changements de modèle ARIMA sans que l'on puisse identifier précisément de causes plausibles. Prenons par exemple la série IPI 1052 vue précédemment (graphique 21). La série semble très régulière, en particulier entre 1990 et 2009. En faisant varier la longueur des données, on constate que les modèles ARIMA estimés restent très stables jusqu'en 2000⁵⁰, soit le modèle $(1, 0, 0)(0, 1, 0)$ avec un coefficient AR égal à $-0,7$. En ajoutant 2001, puis 2002, on trouve un modèle proche $(1, 0, 1)(0, 1, 1)$, avec l'apparition d'une partie MA qui modifie la représentation des fluctuations ponctuelles⁵¹. Les estimations successives d'un modèle simple et stable dans le temps reflètent bien la courbe de la série. Toutefois, en ajoutant 2003 on passe à un modèle $(2, 0, 1)(0, 1, 1)$ avec le coefficient AR d'ordre 2 à $0,2$ et celui d'ordre 1 à $-0,96$, très proche d'un ordre de différenciation. Puis, avec 2004 le modèle revient à la situation antérieure. Ainsi, sans examen attentif, couper la série en 2003 ou en 2004 conduira à considérer deux différents modèles ARIMA sans que cela soit justifié par les données, d'autant plus qu'aucun des autres paramètres de pré-ajustement ne change.

Cependant, on ne peut conclure pour autant que couper une série n'améliore en rien l'estimation d'un modèle ARIMA, qui serait un facteur trop instable pour refléter des événements factuels. Car malgré des variations injustifiées vues dans l'exemple précédent, il y a de suffisamment de reprises de même modèle sur différentes fenêtres pour nous assurer que l'on dispose d'une bonne approximation de la structure des données.

Considérons la série IPI 1722 (fabrication d'articles en papier à usage sanitaire ou domestique), l'allure de la série montre clairement que le début et la fin de la série n'ont pas la même structure (graphique 23). Le modèle ARIMA estimé sur les données complètes est $(2, 1, 0)(0, 1, 1)$, cependant le coefficient AR d'ordre 2 est instable selon la fenêtre d'estimation et décroît à 0 sur la fin de série (les 10 dernières années). La fin de série ne peut donc être approximée par un AR d'ordre 2. Estimer uniquement sur la sous-période « le Passé » le modèle proposé est également $(2, 1, 0)(0, 1, 1)$, mais avec des coefficients stables dans le temps. En revanche, sur la sous-période « le Courant » le modèle est un *airline* $(0, 1, 1)(0, 1, 1)$. Pour cette série, estimer un modèle ARIMA par sous-période donne de meilleurs résultats, en particulier sur les données les plus récentes.

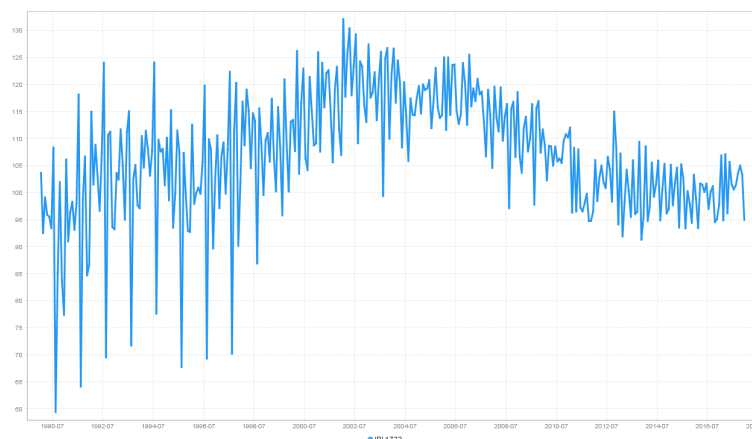
48. Changer les variables de régression modifie la structure des erreurs sur laquelle est estimé un ARIMA

49. Les types d'outliers détectés, les jeux de régresseurs de calendrier choisis, etc.

50. La première fenêtre considérée est 1990-1995, soit 6 ans de données.

51. Les coefficients MA pour la partie régulière et saisonnière sont respectivement de $-0,2$ et $-0,6$, donc réduction de la variance des fluctuations par rapport au coefficient 0.

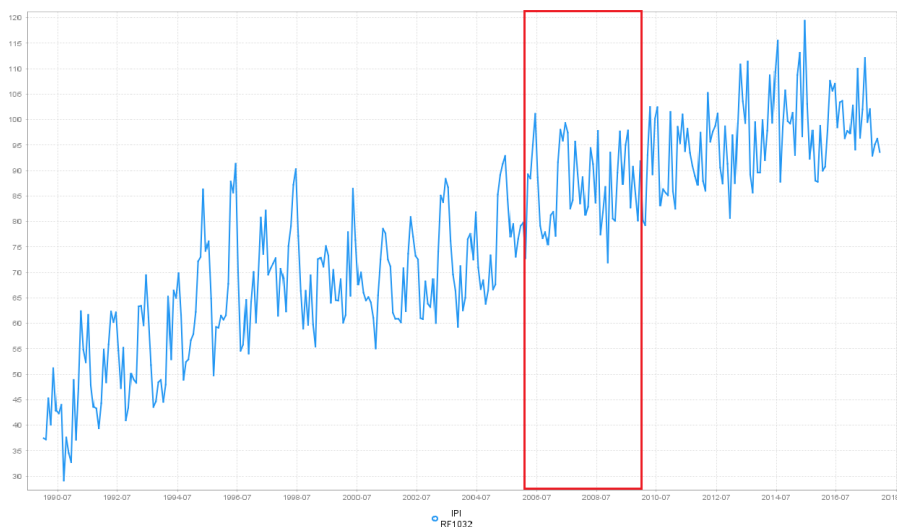
GRAPHIQUE 23 – Série IPI branche 1722 (fabrication d'articles en papier à usage sanitaire ou domestique)



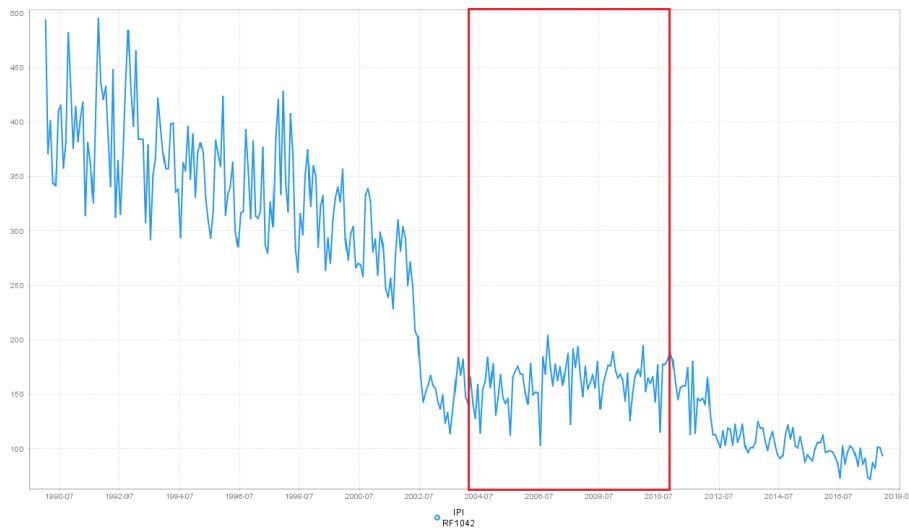
D Exemples de séries IPI avec une plage de fenêtres d'estimation optimale

La plage de fenêtres est représentée dans les graphiques ci-dessous par un rectangle rouge. Elle correspond aux différentes fenêtres de début d'estimation possible pour identifier et estimer les modèles Reg-ARIMA jusqu'à la fin de série. Elle peut être plus ou moins large selon les séries.

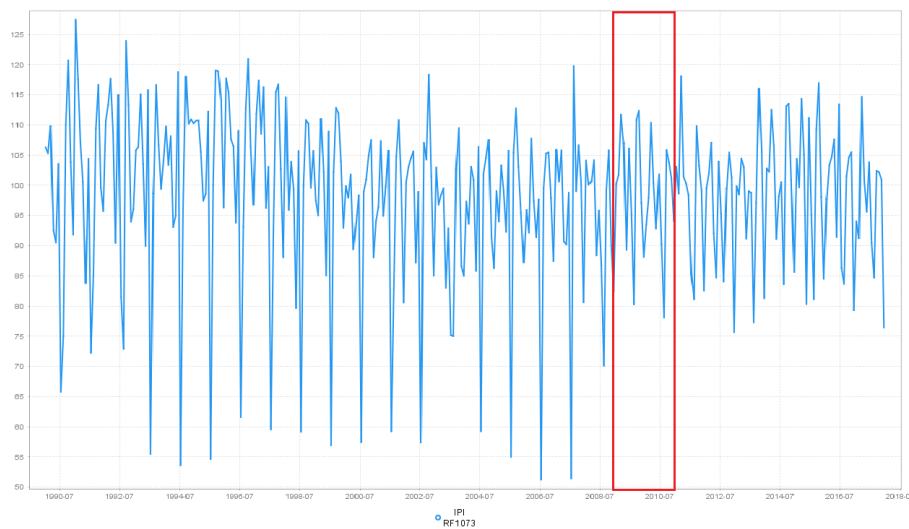
GRAPHIQUE 24 – Série IPI branche 1032



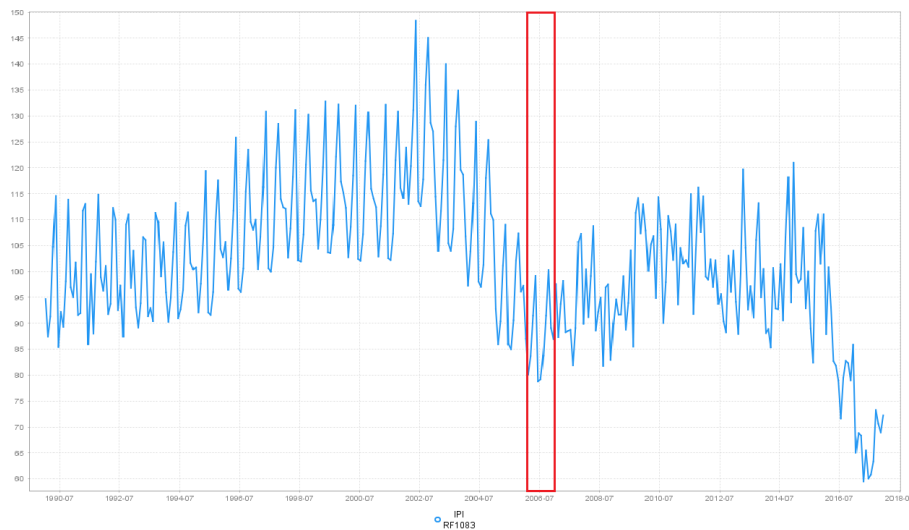
GRAPHIQUE 25 – Série IPI branche 1042



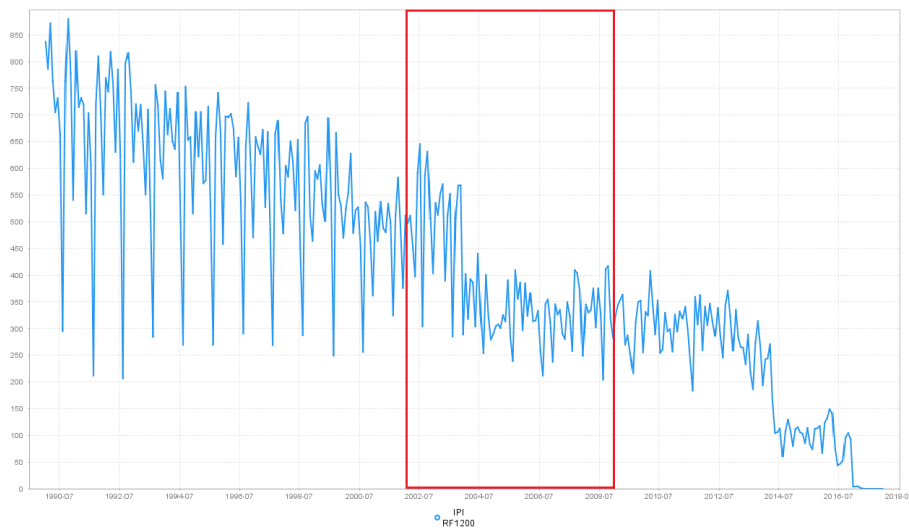
GRAPHIQUE 26 – Série IPI branche 1073



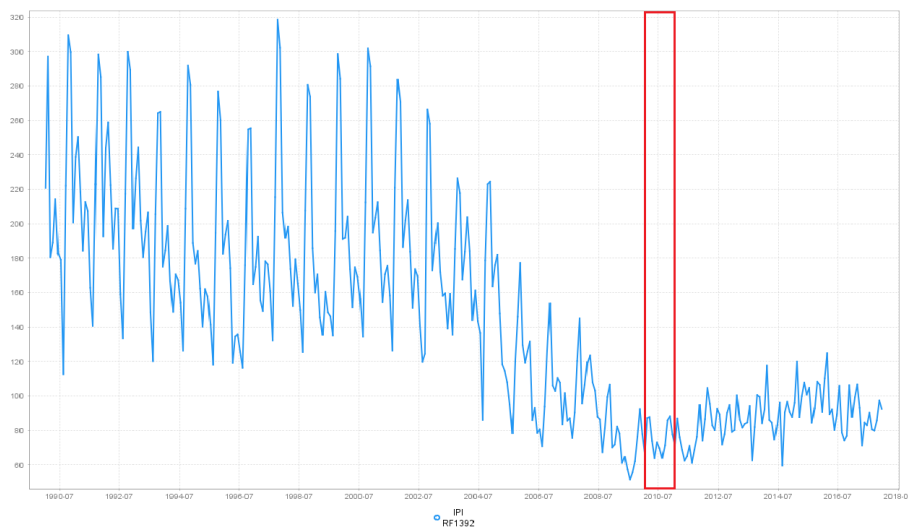
GRAPHIQUE 27 – Série IPI branche 1083



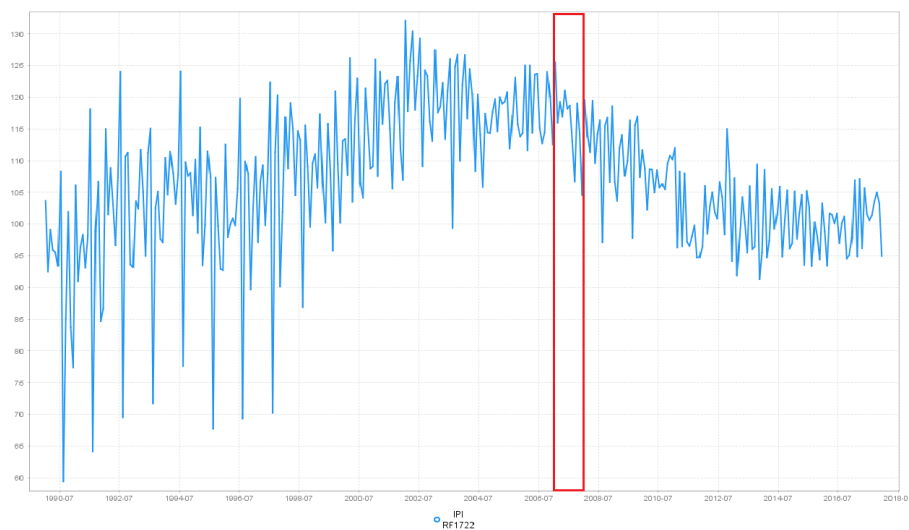
GRAPHIQUE 28 – Série IPI branche 1200



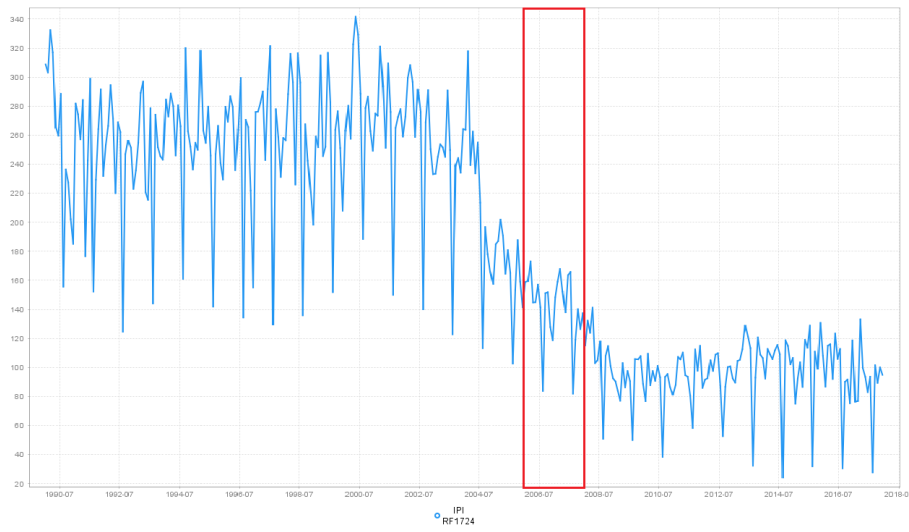
GRAPHIQUE 29 – Série IPI branche 1392



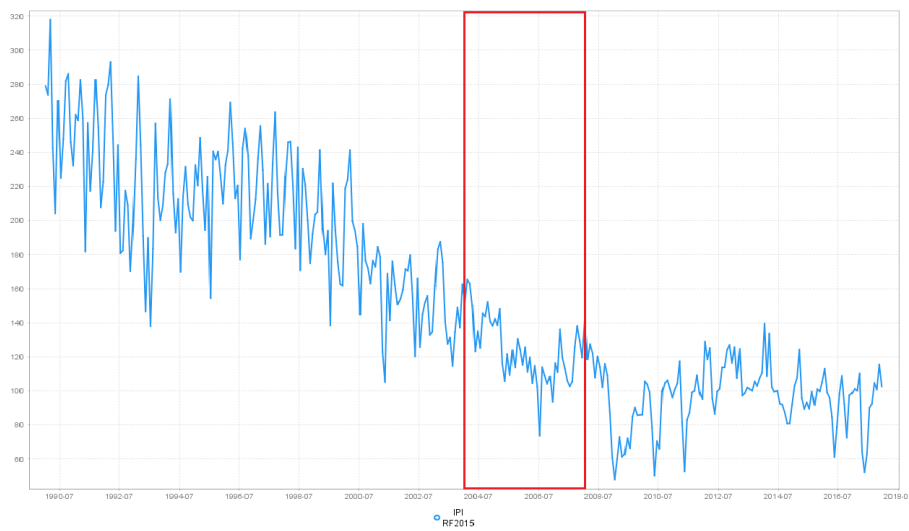
GRAPHIQUE 30 – Série IPI branche 1722



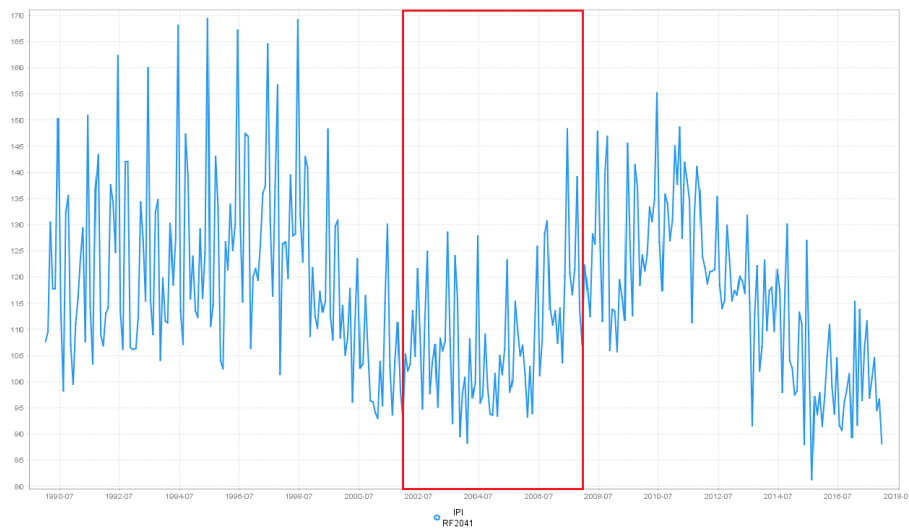
GRAPHIQUE 31 – Série IPI branche 1724



GRAPHIQUE 32 – Série IPI branche 2015



GRAPHIQUE 33 – Série IPI branche 2041



E Description de l'Indice de la production industrielle

Les indices de la production industrielle (IPI) sont calculés par l'Insee à partir des enquêtes mensuelles de branche réalisées auprès d'un échantillon d'entreprises par l'Insee, le Service de la Statistique et de la Prospective (SSP) du ministère de l'Agriculture, le Service de la Donnée et des Études Statistiques (SDES) du ministère de la Transition écologique et solidaire et certains organismes professionnels. Les produits suivis sont situés à tous les niveaux des processus de fabrication.

Le calcul de l'IPI repose sur une décomposition en 520 séries élémentaires, correspondant à des regroupements de produits industriels sur la production desquels un échantillon d'environ 4 800 entreprises est enquêté. Les niveaux supérieurs sont obtenus par agrégation des séries élémentaires en indices de Laspeyres chaînés annuellement, cette méthode de calcul ayant été introduite en 2018 à l'occasion du rebasement 2015 .

Les séries sont corrigées des variations saisonnières et des effets de calendrier au niveau des classes de la NAF rév. 2 (cela correspond à 205 séries). Les indices CVS-CJO des postes des niveaux supérieurs sont obtenus par agrégation des indices CVS-CJO des classes qui les composent.

Depuis 1990, plusieurs événements ont pu modifier le profil saisonnier et les effets de calendrier des séries. C'est en particulier le cas des changements des nomenclatures d'activités et des rebasements. Cela provient notamment de la méthode utilisée pour la rétropolation et du fait que des changements méthodologiques ont été faits en même temps que les rebasements. Les données individuelles des entreprises n'étant pas toujours disponibles sur longues périodes, lorsque les séries sont rétropolées elles le sont avec une approche dite « macro ». C'est-à-dire que les séries rétropolées sont obtenues en contraignant l'égalité (en évolution) avec la série d'origine ou, pour les séries entièrement nouvelles, à la branche à laquelle elles appartiennent. Cette méthode a l'avantage d'être légère car elle repose sur un nombre moins important de données mais a l'inconvénient de reposer sur un certain nombre d'hypothèses (sur l'éclatement ou le regroupement d'indices). C'est la méthode principalement utilisée lors de la modification ou l'ajout de nouvelles séries témoins.

Dans l'histoire des indices de production industrielle on peut recenser les changements méthodologiques suivants :

- 1995 : rebasement 1990 et prise en compte du passage à la nomenclature d'activités française 1993 (NAF, 1993) qui se substitue à la partie « activités » de la nomenclature d'activité et de produits (NAP). En revanche, ce passage n'ayant été achevé qu'en 1999 par la Comptabilité Nationale les pondérations des séries élémentaires des IPI était celles de la NAP et la NAF n'était utilisée que comme nomenclature de diffusion.
- 2000 : rebasement 1995 et première mise à jour des pondérations des séries élémentaires avec la NAF.
- 2003 : rebasement 2000 et prise en compte de la première révision de la NAF (NAF rév.1, 2003).
- 2009 : rebasement 2005 et prise en compte de la seconde révision de la NAF (NAF rév. 2, 2009).
- 2013 : rebasement 2010. Pour se rapprocher des standards internationaux, le nombre de séries de l'IPI observées en quantités a été réduit.
- 2018 : rebasement 2015 avec l'introduction d'indices chaînés annuellement en lieu et place d'un indice à pondérations constantes.

D'autres événements n'apparaissant pas à date fixe peuvent créer des ruptures dans les séries :

- la délocalisation, fermeture ou création d'une grande entreprise dans les branches très concentrées (pharmacie, aéronautique) ;

- la modification, la création ou la suppression de séries témoins. Bien que ces modifications soient généralement faites lors des rebasements, l'historique des nouvelles séries est connu depuis une période plus ou moins longues : la date de début de la rétropolation peut donc varier d'une série à l'autre.

Références

- [1] V. BARNETT & T. LEWIS – *Outliers in statistical data*, Wiley, 1974.
- [2] EUROSTAT – « Ess guidelines on seasonal adjustment », Tech. report, Eurostat Methodologies and Working Papers, European Commission, 2015.
- [3] D. F. FINDLEY, D. P. LYTRAS & T. S. MCELROY – « Detecting seasonality in seasonally adjusted monthly time series », *Statistics* (2017).
- [4] V. GÓMEZ & A. MARAVALL – « Automatic modeling methods for univariate series », Working papers, Banco de España, 1998.
- [5] S. GRUDKOWSKA – « Jdemetra+ reference manual », *Department of Statistics, Narodowy Bank Polski, Warsaw*. (2015).
- [6] D. LADIRAY & B. QUENNEVILLE – « Comprendre la méthode x11 », *Institut National de la Statistique et des Études Économiques* (1999).
- [7] J. LOTHIAN & M. MORRY – « A set of quality control statistics for the x-11-arima seasonal adjustment method », resreport, Seasonal Adjustment and Time Series Staff, Statistics Canada, october 1978.
- [8] J. M. WOOLDRIDGE – *Econometric analysis of cross section and panel data*, MIT press, 2010.