

Chapitre 7 - La correction de la saisonnalité et des effets de jours ouvrables

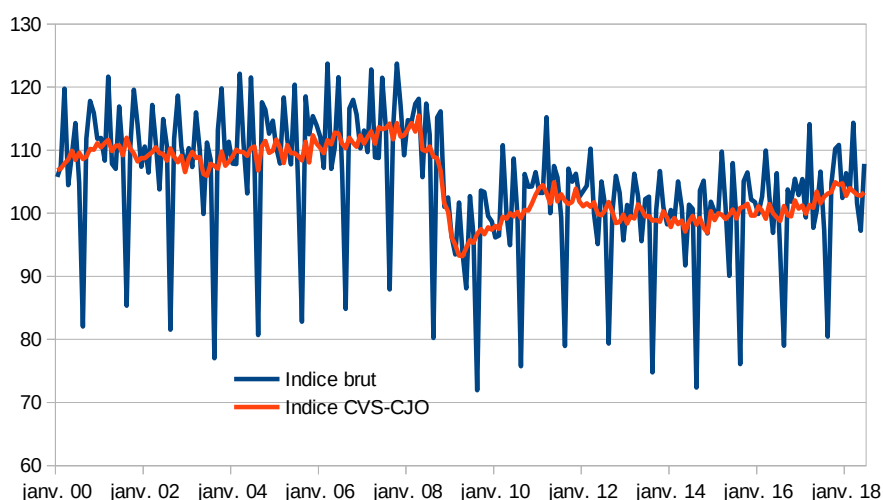
1- La difficulté d'interpréter des indices bruts

L'objectif de l'IPI est de retracer la dynamique de la production en France au mois le mois. Cependant, les variations d'un mois sur l'autre de l'indice « brut » sont difficilement interprétables du fait de la présence, comme dans la plupart des séries économiques, d'effets saisonniers et d'effets dus aux variations du nombre de jours ouvrables (autrement dit au calendrier spécifique et évolutif d'un mois sur l'autre).

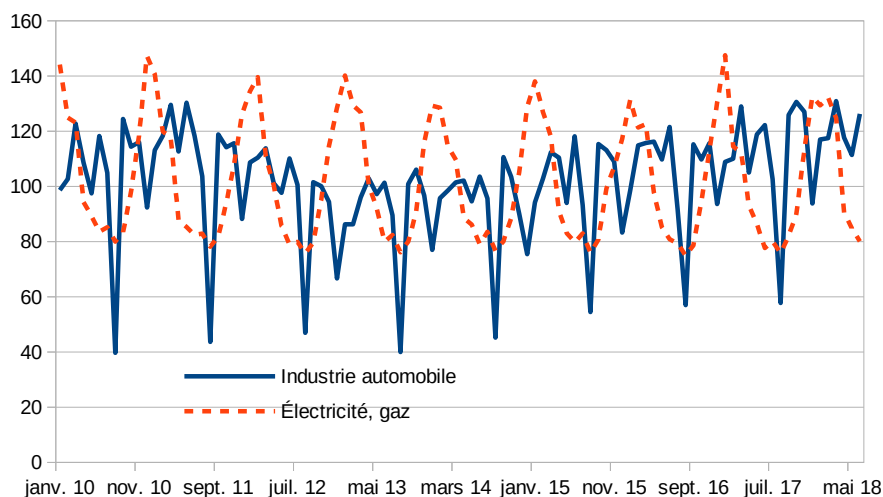
La production industrielle présente en effet une allure saisonnière marquée (*cf.* graphique 1). Chaque année, elle présente en particulier un creux en août car beaucoup d'entreprises réduisent leur activité en raison des congés d'été, rendant difficile l'interprétation des évolutions mensuelles. La forme de la saisonnalité peut de plus être très diverse selon les branches (industrie automobile par exemple ou production d'électricité avec un niveau élevé en hiver et faible en été), ce qui complique encore la comparaison, certaines séries présentant des profils encore largement plus marqués que l'industrie dans son ensemble (*cf.* graphique 2).

Pour pouvoir analyser l'évolution conjoncturelle de la production industrielle, il est donc nécessaire de corriger les séries des variations saisonnières (CVS) et des variations liées au nombre de jours ouvrables mensuel (CJO), de la même façon que de nombreuses autres séries économiques produites par l'Insee. Les méthodes utilisées sont actualisées régulièrement.

Graphique 1 : Indice de la production industrielle (base et référence 100 en 2015)



Graphique 2 : deux types de saisonnalités caractéristiques (indices bruts, base et référence 100 en 2015)



2- Principe de la correction des variations saisonnières et des jours ouvrables

La correction des variations saisonnières et des effets de calendrier (CVS-CJO) vise généralement à décomposer la série brute, notée X_t en plusieurs composantes :

- la tendance-cycle : elle regroupe d'une part la tendance, qui traduit les variations de fond de la série observée sur une longue période de temps, et le cycle, qui est le mouvement lisse et presque périodique autour de la composante tendancielle (il se caractérise par une alternance de périodes d'expansion et de récession dont la longueur varie généralement entre trois et dix ans) ; elle est notée TC_t ;
- la composante saisonnière : elle correspond aux fluctuations saisonnières de la série qui se répètent d'une année sur l'autre et que l'on souhaite neutraliser. Elle est notée S_t ;
- la composante correspondant aux effets de calendrier : elle permet de prendre en compte les compositions journalières différentes des mois ou des trimestres en matière de jours ouvrables ; comme pour la saisonnalité, on cherche à neutraliser ses effets, car elle n'apporte généralement pas non plus d'informations conjoncturelles ; elle est notée WD_t ;
- la composante irrégulière : elle est faite des fluctuations résiduelles et erratiques qui ne peuvent être attribuées à l'une des composantes décrites ci-dessus. Elle est notée I_t .

Deux modèles de décomposition sont utilisés pour le traitement CVS-CJO des séries IPI :

- le modèle additif : $X_t = TC_t + S_t + WD_t + I_t$;
- le modèle multiplicatif : $X_t = TC_t \times S_t \times WD_t \times I_t$; ce modèle peut être transformé en modèle additif en appliquant la fonction logarithme à chaque membre de l'égalité.

Pour désaisonnaliser la série et la corriger des effets de calendrier, il est donc nécessaire d'identifier les deux composantes associées S_t et WD_t et de les retirer de la série étudiée. La série corrigée des variations saisonnières et des effets de calendrier est alors $X_t - S_t - WD_t$ dans le cas du modèle additif et $X_t / (S_t \times WD_t)$ dans le cas du modèle multiplicatif.

Les deux principales méthodes de désaisonnalisation utilisées au sein du système statistique européen sont la méthode Tramo-Seats, méthode paramétrique utilisée pour l'IPI jusqu'à la base 2005, et la méthode non paramétrique X13-ARIMA.

3- Méthode utilisée pour la correction CVS-CJO de l'IPI

Depuis la base 2010, la désaisonnalisation de l'IPI³⁴ repose sur la méthode X13-ARIMA, mise en œuvre à l'aide du logiciel JDemetra+ proposé par Eurostat³⁵ avec l'utilisation de régresseurs CJO spécifiques (*cf. infra*). La méthode X13-ARIMA s'articule autour de deux modules principaux³⁶.

Un premier module (« RegArima ») permet d'effectuer le pré-ajustement de la série : détection des perturbations affectant la série (« *outliers* »), correction des effets « jours ouvrables », prolongement de la série aux bords et fourniture de diagnostics. En effet, l'estimation des coefficients saisonniers sera d'autant plus satisfaisante si la série brute n'est pas trop perturbée par les aléas conjoncturels.

Quatre types de perturbations peuvent être décelées et sont ensuite prises en compte à travers l'ajout de régresseur dans un modèle de type Reg-Arima :

- les points atypiques (« *additive outliers* ») sont des perturbations intervenues un mois donné qui ne peuvent être attribuées à de la saisonnalité. Il s'agit par exemple d'une grève dont on peut supposer qu'elle n'aura pas de répercussion sur la production des mois suivants ;
- les changements transitoires de niveau (« *transitory changes* ») : un incident a été suffisamment important pour affecter le niveau de la production des mois suivants (par exemple une inondation) ;
- les ruptures de niveau (« *level shifts* ») : par exemple l'ouverture d'une nouvelle usine, ou une évolution économique majeure comme la crise de 2008 ;
- les changements de saisonnalité (« *seasonal outlier* ») : ils permettent de prendre en compte un changement brusque dans la saisonnalité avec un impact durable, comme un changement de méthode de comptabilisation de la production.

Pour la correction des effets de jours ouvrables, le module utilise (si une telle correction est nécessaire) des régresseurs reproduisant la composition du calendrier (au travers de la structure des mois en jours ouvrables non fériés). Afin de tenir compte des spécificités du calendrier national (fêtes nationales, jours fériés), l'Insee met en place ses propres variables et les intègre ensuite dans JDemetra+. Les régresseurs sont centrés en retirant les moyennes de long terme pour chaque mois, ce qui permet de retirer du calendrier sa composante saisonnière. Enfin, on applique une procédure de choix automatique entre différentes combinaisons de régresseurs possibles (*cf. infra*).

Après avoir détecté ces effets éventuels, le module RegArima les estime dans le cadre d'un modèle Reg-Arima afin de « linéariser » la série et la prolonger aux bords en vue de sa désaisonnalisation.

Dans un deuxième temps, un second module (X11) procède à la désaisonnalisation proprement dite par lissages itératifs de moyennes mobiles et décompose la série linéarisée dans la première partie en composantes orthogonales : la tendance-cycle, la saisonnalité et la composante irrégulière.

34 Le lecteur pourra se reporter aux « *ESS guidelines on seasonal adjustment* » d'Eurostat (2015) pour une revue des bonnes pratiques en matière de désaisonnalisation.

35 Ce logiciel permet de mettre en œuvre les deux principales méthodes de désaisonnalisation utilisées au sein du système statistique européen : la méthode Tramo-Seats, méthode paramétrique utilisée précédemment dans la base 2005 de l'IPI, et la méthode non paramétrique X-13-ARIMA.

36 Pour une présentation complète du programme, *cf. « X-13ARIMA-SEATS Reference Manual »* (2017), U.S. Census Bureau. Pour une présentation pédagogique du fonctionnement du module X11, on pourra se rapporter au document suivant : « *Comprendre la méthode X11* » (1999), D. Ladiray, B. Quenneville.

La désaisonnalisation est de type indirect : elle est effectuée à un niveau fin, le niveau classe (NACE 4), et les agrégats CVS-CJO sont ensuite calculés directement à partir de ces séries³⁷ à l'aide des méthodes présentées au chapitre 6.

Enfin, une évolution est mise en place à partir de la campagne de mars 2019 en estimant et en appliquant les modèles de correction CVS-CJO sur une sous-période réduite pour les données récentes³⁸, et ce afin de mieux répondre aux recommandations d'Eurostat sur le sujet. En effet, l'application des méthodes de désaisonnalisation sur 30 ans (soit la profondeur temporelle des séries IPI) peut poser des problèmes de robustesse en raison des évolutions économiques ou de comportement. Cette façon de faire permet également de figer les évolutions sur le passé alors que les méthodes de désaisonnalisation induisent par nature des révisions sur l'ensemble de la série à chaque ajout d'un nouveau point (même si ces révisions sont très faibles au-delà de quelques années dans le passé).

4- Les campagnes mensuelles et annuelles

Le processus de correction CVS-CJO des indices IPI consiste à appliquer chaque mois les modèles identifiés série par série. Ces modèles sont construits et/ou révisés lors de campagnes annuelles (*cf. infra*) qui ont lieu généralement sur la période janvier-février, avant la première campagne de l'année (publication des indices de janvier au début du mois de mars). La forme générale des modèles (hors actualisation des paramètres) est ensuite fixée à l'issue de la mise à jour annuelle pour toute l'année à venir.

Les campagnes mensuelles de correction CVS-CJO prennent en compte les nouvelles données disponibles chaque mois, ré-estiment les coefficients des modèles, recherchent les « *outliers* » sur la période récente (12 derniers mois) et mettent à jour les coefficients CVS-CJO (impliquant par la même des révisions sur le passé au fur et à mesure que de nouvelles données arrivent).

Cette méthode consistant à ré-estimer les modèles avec la dernière donnée mensuelle disponible, mais à modèle, jeu de régresseurs et filtres inchangés (appelée « *Partial concurrent last outliers* »), est recommandée par Eurostat.

Déroulement des campagnes annuelles de remise à jour complète des modèles

La démarche globale d'une campagne annuelle consiste à comparer la modélisation actuelle (*i.e.* les modèles utilisés actuellement en production, considérée comme la référence) à une modélisation automatique. Si la modélisation automatique est de meilleure qualité (au regard du bilan qualité construit à l'aide des statistiques et résultats fournis par le programme) que la modélisation actuelle, alors la série est étudiée de manière plus précise pour savoir s'il est nécessaire de changer la modélisation. Plus précisément :

- Pour chaque série, une modélisation automatique est lancée en réinitialisant tous les paramètres (régresseurs de jours ouvrables, modèle ARIMA sélectionné, points atypiques, etc.) ; les combinaisons testées pour les modèles CJO sont présentées dans le tableau 1 ;
- Les modèles actuels (en production lors des campagnes mensuelles) sont mis à jour avec les dernières données en ré-identifiant les *outliers* sur la dernière année (option de rafraîchissement utilisée pendant les campagnes mensuelles) ;
- Pour chaque série examinée, un bilan qualité³⁹ est édité. Ce bilan est utilisé pour comparer la qualité de la modélisation actuelle et de la modélisation automatique. Une note est ainsi calculée pour chaque série

37 Une autre méthode, dite « directe », consiste à effectuer des désaisonnalisations à chaque niveau d'agrégation.

38 Le choix a été fait d'estimer des modèles depuis 2005, et de fixer les données avant 2012.

39 Le bilan qualité reprend des diagnostics issus de JDemetra+ et crée des indicateurs synthétiques sous forme de notes. Les grandes catégories de diagnostics disponibles dans ce fichier sont les suivantes :

- *Qualité globale du traitement* ;
- *Qualité du modèle Reg-ARIMA* ;
- *Qualité de la décomposition* ;
- *Qualité des révisions* (qui porte sur les deux dernières années disponibles).

Une note de qualité globale sur 20 est établie et correspond à la moyenne pondérée des notes de chaque famille de diagnostics. Cette note permet de comparer les modèles (traitement automatique / ancien modèle) et de repérer les séries « problématiques », tout en donnant la priorité aux séries ayant les pondérations les plus importantes.

et ne sont étudiées de façon approfondie que les séries pour lesquelles la note fait apparaître une meilleure qualité de la modélisation automatique par rapport à la modélisation actuelle (de référence) ; à titre d'illustration, il y avait 58 séries dans ce cas lors de la campagne mise en œuvre début 2018 (sur un peu plus de 200) ;

- Pour prioriser le traitement des séries à étudier, la note tient également compte de la pondération (et donc du poids économique) de la série. Ainsi, dans le cas où il n'est pas possible d'étudier l'ensemble des séries problématiques, cette note « pondérée » permet de détecter en priorité celles dont la mauvaise qualité est la plus susceptible de dégrader la qualité des séries agrégées ;
- Une fois le modèle sélectionné, les nouvelles séries CVS-CJO obtenues sont comparées avec les anciennes.

Tableau 1 : différentes spécifications pour la CJO

| Jeu de régresseurs | Caractéristiques |
|--|--|
| Sans objet | Pas d'effets CJO (=pas de régresseurs CJO) |
| Lpyear | Seul un effet « année bissextile » et pas d'effets selon le type de jours (un seul régresseur) |
| Regcjo1 (+lpy), jours différenciés (y compris samedi) | 6 régresseurs (lundis, mardis, ..., samedis ouvrables) en contraste aux dimanches et jours fériés (+ éventuellement un effet « année bissextile ») |
| Regcjo2 (+lpy), effet semaine / week-end | 1 régresseur (jours de la semaine, du lundi au vendredi, ouvrables) en contraste aux week-ends et jours fériés (+ éventuellement un effet « année bissextile ») |
| Regcjo3 (+lpy), jours différenciés par rapport au week-end | 5 régresseurs (lundis, mardis, ..., vendredis ouvrables) en contraste aux week-ends et jours fériés (+ éventuellement un effet « année bissextile ») |

Déroulement des campagnes mensuelles

Les campagnes mensuelles ré-estiment les modèles en prenant en compte les nouvelles données disponibles :

- Actualisation du fichier de données / lancement des traitements dans JDemetra+ ;
- Contrôle de la qualité des traitements dans JDemetra+ ;
- Changement interactif des spécifications si nécessaire (*outliers* notamment) ;
- Export et archivage des traitements mensuels.

5- Les principales évolutions du traitement CVS depuis la base 2010

Comme évoqué plus haut, depuis la base 2010, la désaisonnalisation de l'IPI repose désormais sur la méthode X13-ARIMA ou ses versions ultérieures. Auparavant, la correction CVS-CJO reposait sur la méthode Tramo-Seats, autre méthode recommandée par Eurostat, particulièrement dans le cas des statistiques de court terme.

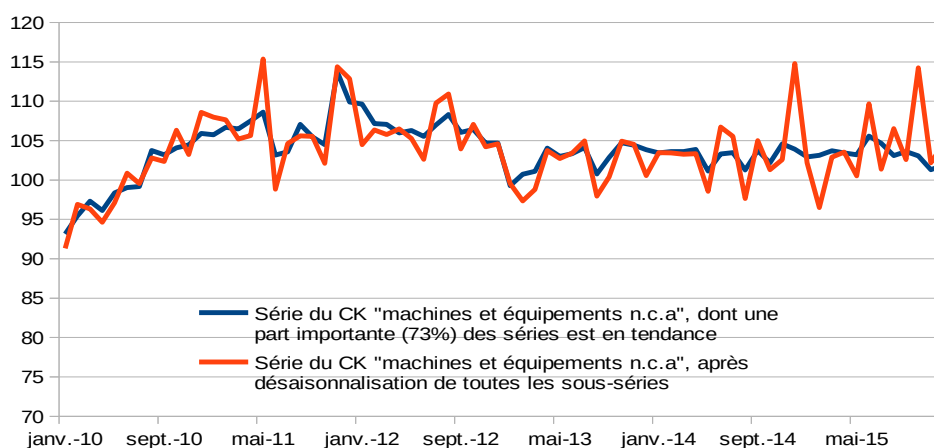
Le changement de méthode de désaisonnalisation⁴⁰ pour la base 2010 a été décidé afin d'améliorer la cohérence avec la méthode de construction des comptes nationaux trimestriels, qui s'appuient pour partie sur les séries de l'IPI. En effet, les comptes trimestriels utilisent historiquement le programme X13-ARIMA. Utiliser des méthodes différentes pour élaborer l'IPI et les comptes trimestriels pouvait induire des écarts parfois difficiles à interpréter. Ce changement de méthode a entraîné des révisions sur le passé.

En 2015, il a par ailleurs été décidé de réaliser l'ajustement saisonnier et la correction des effets « jours ouvrables » sur les séries au niveau NACE 4 (niveau classe) et non plus sur les séries de niveau NACE 3, pour des raisons pratiques et de cohérence avec les séries publiées sur le site d'Eurostat. De nouveaux modèles pour le niveau classe ont donc dû être mis au point. Bien que pertinente pour les raisons évoquées ici, cette évolution peut rendre plus ardue l'identification de modèles robustes au niveau fin NACE 4, du fait d'aléas relativement importants à ce niveau de détail, alors que ces aléas ont tendance à se neutraliser à un niveau plus agrégé.

Enfin, début 2016, il a été mis fin au suivi de certaines séries en tendance. Jusqu'en décembre 2015, le calcul de l'indice CVS-CJO de la production industrielle était basé, dans certaines branches, sur des séries en tendance, et non sur des séries corrigées des variations saisonnières et des effets « jours ouvrables ». Ces séries, au nombre de 18 (sur 203), représentaient environ 12,5 % de la production manufacturière et étaient principalement localisées dans les « machines et équipements n.c.a. (CK) » (dont elles représentaient 73 % de la branche) et dans « l'industrie pharmaceutique (CF) » (dont elles représentaient la totalité de la branche).

Ce changement de méthode visait à se mettre en conformité avec les lignes directrices d'Eurostat (*cf. supra*). La volatilité de ces séries s'est accrue mais reste de l'ordre de celle des séries de l'industrie automobile (CL1) ou encore des autres matériels de transport (CL2), dont les poids sont comparables. Le graphique 3 permet d'illustrer l'impact de ce changement en base 2010 sur les indices de la branche CK.

Graphique 3 : Indices de la branche CK - "machines et équipements n.c.a." avant et après l'abandon des séries en tendance (base et référence 100 en 2010)



40 En pratique, il y a eu peu d'impact sur les principaux agrégats de l'IPI, mais des divergences parfois importantes ont été constatées sur les séries plus fines. Pour environ un quart des séries, les séries CVS-CJO issues de X13-ARIMA étaient pratiquement confondues avec les séries produites avec Tramo-Seats. Au contraire, pour environ un tiers des séries, les résultats différaient assez fortement. Ces divergences ne sont pas surprenantes dans la mesure où Tramo-Seats et X13-ARIMA reposent sur des principes de décomposition de la série très différents. Cependant, certaines divergences provenaient de l'étape dite de « pré-ajustement » : modélisation ARIMA, détection et correction des points extrêmes et des ruptures, transformation éventuelle en utilisant la fonction logarithme. Quant à la correction des effets « jours ouvrables », elle n'a pas paru être une cause de divergence : les séries CJO obtenues avec les deux méthodes étaient très proches dès lors évidemment que les mêmes régresseurs CJO étaient utilisés.