

# L'APPARITION DU REGIME « AUTO-ENTREPRENEUR » : CONSEQUENCES SUR LA DESAISONNALISATION DES CREATIONS D'ENTREPRISES

*Thomas Balcone<sup>1</sup> (\*)*

*(\*) Insee, Direction de la méthodologie et de la coordination statistique et internationale*

## **Résumé**

L'apparition du régime des « auto-entrepreneurs » en janvier 2009 a entraîné un choc en niveau et en saisonnalité sur les séries de créations d'entreprises. La distinction entre les créations correspondant au nouveau statut et les créations « hors auto-entrepreneurs » a ainsi été nécessaire pour assurer une désaisonnalisation de qualité satisfaisante. Cependant, les séries des « auto-entrepreneurs » étant courtes (moins de 7 ans de données disponibles), la désaisonnalisation de ces séries s'est avérée délicate. Des résultats de qualité satisfaisante ont toutefois été obtenus en août 2013.

## **Abstract**

In January 2009, the establishment of the « auto-entrepreneurs » status impacted strongly the level and the seasonality of the business start-ups series. Thus, the distinction between the « auto-entrepreneurs » business start-ups and the other ones was necessary to ensure a seasonal adjustment of good quality. However, as the « auto-entrepreneurs » business start-ups series are short (less than 7 years of data available), the seasonal adjustment of these series has proved difficult. Nevertheless, satisfactory results were obtained in August 2013.

## **Mots-clés**

Séries temporelles, désaisonnalisation, X12-ARIMA, séries courtes

---

<sup>1</sup> Thomas.balcone@insee.fr

## Introduction

Les séries de créations d'entreprises sont des indicateurs importants dans l'appréciation de la situation économique conjoncturelle nationale. Ces séries présentent une saisonnalité qu'il est nécessaire de corriger afin de permettre une meilleure analyse de la situation économique. Cependant, la désaisonnalisation de ces séries peut s'avérer complexe. En effet, le nombre de créations d'entreprises est très sensible à certains phénomènes d'origine administrative, comme la mise en place d'un nouveau régime par exemple. C'est ce qu'il s'est produit en janvier 2009 avec l'apparition du régime des « auto-entrepreneurs » : les séries de créations d'entreprises ont alors subi un choc en niveau et en saisonnalité.

L'apparition de ce nouveau régime a, par conséquent, nécessité une nouvelle expertise de la méthode de désaisonnalisation mise au point en 2008 par la Division « infrastructures et répertoire statistiques » (IRS). Cette expertise a été menée au second semestre 2012 par le département des méthodes statistiques (DMS) pour déterminer la meilleure stratégie de désaisonnalisation à adopter. Jusqu'à cette date, chaque série de créations d'entreprises CVS<sup>2</sup>-CJO<sup>3</sup> était alors obtenue comme la somme de la série des créations d'entreprises hors auto-entrepreneurs CVS-CJO et de la série brute des auto-entrepreneurs. Une seconde expertise a ensuite été menée un an plus tard (août 2013), afin de savoir s'il fallait reconsidérer la stratégie de désaisonnalisation adoptée un an plus tôt.

Après une présentation générale des données, nous détaillons, tout d'abord, la désaisonnalisation des séries de créations d'entreprises avant l'apparition du nouveau régime. Puis, nous présentons les conséquences de l'apparition du nouveau statut sur la qualité des désaisonnalisations, ainsi que la stratégie de désaisonnalisation retenue à l'issue de la première expertise menée. Enfin, nous montrons que la désaisonnalisation des séries de créations d'entreprise est de qualité satisfaisante en août 2013 et que, par conséquent, la stratégie de désaisonnalisation a été revue suite à cette seconde expertise.

---

<sup>2</sup> corrigée des variations saisonnières

<sup>3</sup> corrigée des effets « jours ouvrables »

# 1. Présentation

## 1.1. Les données utilisées et publiées

Dans le cadre de cette étude, nous utilisons les données mensuelles relatives aux créations d'entreprises couvrant **la période allant du 1<sup>er</sup> janvier 2000 au 1<sup>er</sup> décembre 2014**. Les séries des créations d'entreprises sont constituées à partir des informations du répertoire Sirene<sup>4</sup>. Toute entreprise, personne physique ou personne morale, est inscrite au répertoire lors sa première déclaration de démarrage d'activité non salariée et reçoit un numéro d'identification unique.

Depuis le 1<sup>er</sup> janvier 2007, la notion de création d'entreprise retenue par l'Insee s'appuie sur le concept défini par Eurostat afin d'harmoniser et de pouvoir comparer les données européennes. Une création d'entreprise au sens d'Eurostat correspond à l'apparition d'une unité légale exploitante n'ayant pas de prédécesseur. Ce concept de création se base sur la mise en œuvre de nouveaux moyens de production. Ainsi, par rapport aux immatriculations dans Sirene, on retiendra comme création :

- les créations d'entreprises correspondant à la création de nouveaux moyens de production (cela se traduit par une nouvelle immatriculation dans Sirene)
- les cas où l'entrepreneur (en général, un entrepreneur individuel) reprend une activité après une interruption de plus d'un an (il n'y a pas de nouvelle immatriculation dans Sirene mais reprise de l'ancien numéro Siren)
- les cas où l'entrepreneur reprend une activité après une interruption de moins d'un an mais avec changement d'activité
- les reprises par une nouvelle entreprise de tout ou partie des activités et moyens de production d'une autre entreprise (cela se traduit par une nouvelle immatriculation dans Sirene) s'il n'y a pas continuité de l'entreprise reprise, i.e. si au moins deux des trois éléments suivants concernant le siège de l'entreprise sont modifiés lors de la reprise : l'unité légale contrôlant l'entreprise, l'activité économique et la localisation.

A partir de la publication de février 2009 relatives aux données de janvier 2009, les créations d'entreprises sont diffusées dans **la nomenclature agrégée NA fondée sur la NAF rév.2**, entrée en vigueur le 1<sup>er</sup> janvier 2008. Pour permettre des analyses sur longue période, la série des créations d'entreprises a été recalculée en NA depuis l'année 2000. Les séries de créations d'entreprises « France entière<sup>5</sup> » sont ainsi diffusées pour les 9 regroupements de sections (niveau A10 de la nomenclature NA, 2008) et les 4 sections (niveau A21) suivants :

---

<sup>4</sup> Système Informatique pour le Répertoire des Entreprises et de leurs Établissements

<sup>5</sup> France métropolitaine et les 5 Départements d'Outre-mer (Guadeloupe, Martinique, Guyane, La Réunion et Mayotte depuis avril 2011)

**Tableau 1.1. : Nomenclature**

Niveau d'agrégation de la NA		Intitulé court (tel qu'il apparaît sur les Informations rapides)	Intitulé complet (tel qu'il apparaît dans la nomenclature NA, 2008)
A10	A21		
Ensemble		Ensemble des créations d'entreprises France entière	
BE		Industrie	Industrie manufacturière, industries extractives et autres
	C	<i>dont industrie manufacturière</i>	<i>Industrie manufacturière</i>
FZ		Construction	Construction
GI		Commerce, transports, hébergement et restauration	Commerce de gros et de détail, transports, hébergement et restauration
	G	<i>dont Commerce ; réparation d'autos et de motos</i>	<i>Commerce; réparation d'automobiles et de motocycles</i>
	H	<i>dont Transports et entreposage</i>	<i>Transports et entreposage</i>
	I	<i>dont hébergement et restauration</i>	<i>Hébergement et restauration</i>
JZ		Information et communication	Information et communication
KZ		Activités financières et d'assurance	Activités financières et d'assurance
LZ		Activités immobilières	Activités immobilières
MN		Soutien aux entreprises	Activités spécialisées, scientifiques et techniques et activités de services administratifs et de soutien
OQ		Enseignement, santé, action sociale	Administration publique, enseignement, santé humaine et action sociale
RU		Services aux ménages	Autres activités de services

Dans l' « Informations rapides » publiée mensuellement, les 13 séries mensuelles « France entière » CVS<sup>6</sup>-CJO<sup>7</sup> correspondant aux différents niveaux d'agrégation A10 et A21 mentionnés ci-dessus sont diffusés. Une série CVS-CJO correspondant à l'ensemble des créations d'entreprises « France entière » (série « Ensemble ») est également publiée. Cette dernière est obtenue en utilisant la méthode indirecte, i.e. en sommant les 9 séries CVS-CJO correspondant aux 9 niveaux d'agrégation A10.

## 1.2. La méthode de désaisonnalisation et de correction des effets « jours ouvrables » utilisée

Jusqu'à la publication de janvier 2007 portant sur les données de créations d'entreprises de décembre 2006, les séries CVS-CJO étaient obtenues de la manière suivante :

- la correction des effets « jours ouvrables » était réalisée en appliquant un coefficient de correction de jours ouvrables à la série brute. On obtenait ainsi une série CJO de la manière suivante :

$$\text{serie}^{\text{CJO}}(m/A) = \text{serie}^{\text{brute}}(m/A) * \frac{\text{nombre\_jours\_ouvrables}(m/A)}{\frac{1}{12} \sum_{l=1}^{12} \text{nombre\_jours\_ouvrables}(l/A)}$$

où m est le mois et A l'année

- puis, on utilisait la procédure X11 du logiciel SAS pour désaisonnaliser les séries corrigées des effets « jours ouvrables » et ainsi obtenir des séries CVS-CJO

<sup>6</sup> corrigées des variations saisonnières

<sup>7</sup> corrigées des effets « jours ouvrables »

La série CVS-CJO correspondant à l'ensemble des créations d'entreprises était obtenue par la méthode directe. On obtenait ainsi directement cette série CVS-CJO à partir de la série « Ensemble » brute.

Puis, à partir de la publication de février 2007, portant sur les données de janvier 2007, la méthode de correction des effets « jours ouvrables » et de désaisonnalisation utilisée a changé. En effet, à partir de cette date, les données CVS-CJO sont obtenues en utilisant la procédure X12-ARIMA du logiciel SAS. La méthode X12-ARIMA, mise en œuvre à travers cette procédure, permet de réaliser la correction des effets « jours ouvrables » **et** la correction des variations saisonnières (cf. infra pour une présentation plus détaillée de cette méthode). Ce changement de méthode s'est naturellement accompagné du calcul de séries CVS-CJO rétropolées.

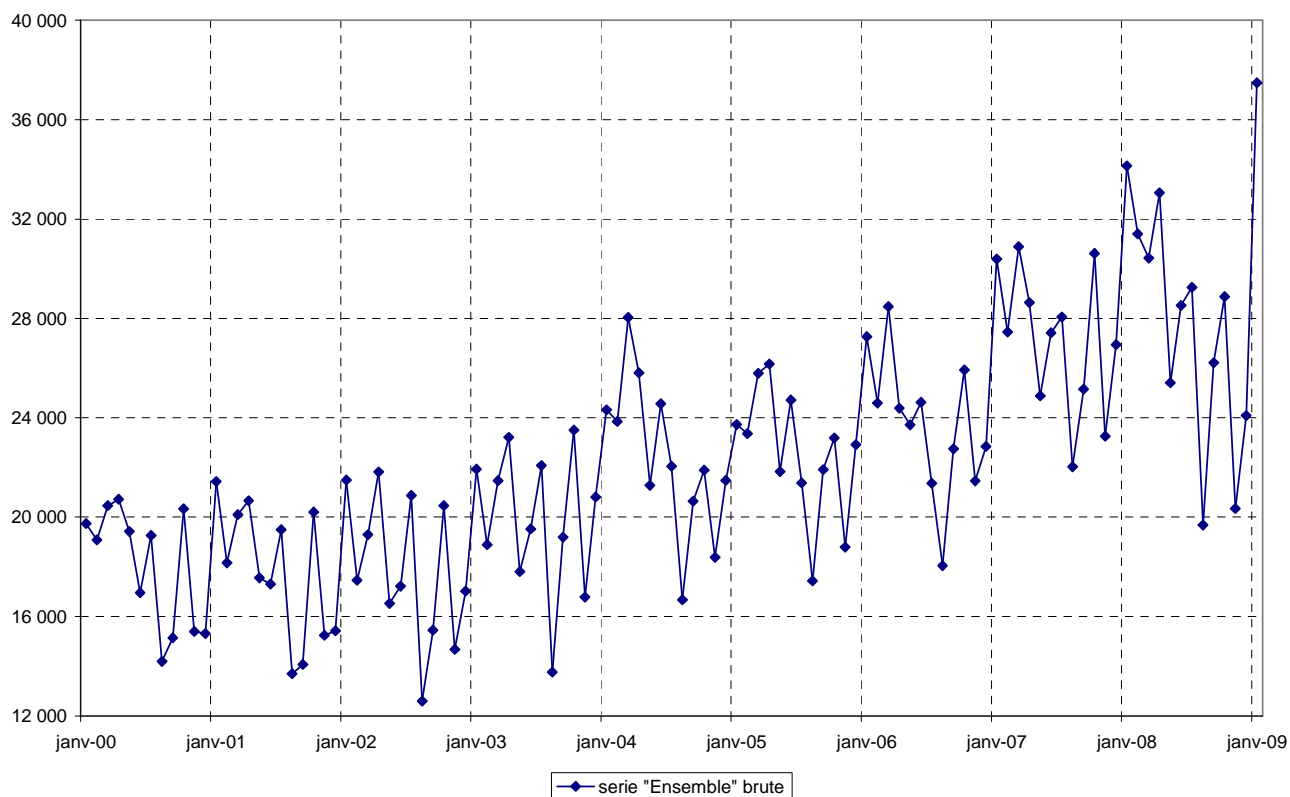
De plus, à partir de cette date, la série « Ensemble » CVS-CJO est obtenue de manière indirecte en sommant les séries CVS-CJO des différents niveaux d'agrégation (méthode indirecte).

### 1.3. L'apparition du régime « auto-entrepreneur » en janvier 2009

En janvier 2009, le régime « auto-entrepreneur »<sup>8</sup> est apparu. Ainsi, à partir de cette date, les statistiques de créations d'entreprises incluent les demandes de créations enregistrées dans Sirene au titre de ce régime. Ce dénombrement n'inclut pas les entrepreneurs déjà en activité avant le 1<sup>er</sup> janvier 2009 et qui avaient demandé, à titre dérogatoire au plus tard le 31 mars 2009, à bénéficier du régime micro-social et du versement fiscal libératoire en 2009.

L'apparition de ce régime s'est traduite par une forte hausse pour la série d'ensemble et pour tous les niveaux d'agrégation concernés par ce régime (cf. graphique 1.1. et tableau 1.2.).

**Graphique 1.1. : la série « Ensemble » brute (01/2000 - 01/2009)**



<sup>8</sup> depuis le 19 décembre 2014, suite à la loi Pinel du 18 juin 2014, le terme de micro-entrepreneur se substitue à celui d'auto-entrepreneur.

Tableau 1.2. : l'impact de l'apparition du régime « auto-entrepreneur » en janvier 2009

Niveau d'agrégation de la NA		Intitulé court (tel qu'il apparaît sur les Informations rapides)	Série brute	
A10	A21		Taux de croissance en glissement mensuel entre 12/2008 et 01/2009	Taux de croissance en glissement mensuel moyen entre 12/A et 01/(A+1), A=2000,...,2007 (série brute)
Ensemble		Ensemble des créations d'entreprises France entière	55,57%	26,78%
BE		Industrie	70,81%	32,83%
	C	<i>dont industrie manufacturière</i>	93,82%	36,67%
FZ		Construction	52,59%	45,70%
GI		Commerce, transports, hébergement et restauration	60,44%	24,86%
	G	<i>dont Commerce ; réparation d'autos et de motos</i>	79,32%	26,53%
	H	<i>dont Transports et entreposage</i>	2,66%	18,78%
	I	<i>dont hébergement et restauration</i>	21,00%	21,56%
JZ		Information et communication	133,37%	18,53%
KZ		Activités financières et d'assurance	-18,66%	-18,77%
LZ		Activités immobilières	-5,82%	7,72%
MN		Soutien aux entreprises	61,14%	32,35%
OQ		Enseignement, santé, action sociale	29,01%	35,19%
RU		Services aux ménages	119,51%	22,33%

Il est à noter que pour les secteurs « Transports et entreposage », « Hébergement et restauration », « Activités financières et d'assurances », « Activités immobilières » et « Enseignement, santé, action sociale », l'apparition du régime d'« auto-entrepreneur » en janvier 2009 n'a qu'un faible impact sur le nombre de créations d'entreprises enregistrées. Pour les autres secteurs et pour l'« ensemble », cela se traduit par une hausse beaucoup plus forte que celle observée en moyenne sur la période 2001 - 2008. Le secteur « Services aux ménages » voit même son nombre de créations plus que doublé (+120%) entre décembre 2008 et janvier 2009.

Cette forte hausse observée sur les séries brutes a eu des conséquences sur le calcul des séries CVS-CJO. En effet, l'apparition du régime d'« auto-entrepreneur » a contraint le service producteur à adopter une nouvelle stratégie d'agrégation à partir de la publication d'octobre 2009 portant sur les données de septembre 2009. Cette stratégie consistait à calculer la série CVS-CJO d'un secteur donné comme la somme de la série « hors auto-entrepreneurs » CVS-CJO et la série « auto-entrepreneurs » brute afin de traiter correctement le choc en niveau, en saisonnalité et en effets « jours ouvrables » subi par les séries brutes (cf. infra). Cependant, la méthode de correction des effets « jours ouvrables » et des variations saisonnières utilisée était toujours X12-ARIMA.

Nous allons tout d'abord nous intéresser à l'avant « auto-entrepreneurs », i.e. à la période allant de janvier 2000 à décembre 2008, soit 9 années complètes.

## 2. L'avant « auto-entrepreneurs » (2000-2008)

Afin de pouvoir étudier l'impact de l'apparition du régime « auto-entrepreneurs » en janvier 2009 sur la qualité des traitements, il est important de :

- connaître les principales caractéristiques des dernières désaisonnalisations réalisées avant l'apparition du régime
- et d'évaluer la qualité des ajustements réalisées

Dans un premier temps, nous réalisons ainsi la désaisonnalisation des 13<sup>9</sup> séries présentes dans le tableau 1.1 en considérant l'avant « auto-entrepreneurs » (i.e. la période 2000-2008). La série « Ensemble » CVS-CJO est obtenue de manière indirecte. Ce traitement est réalisé en utilisant le logiciel Jdemetra<sup>10</sup> et en mettant en œuvre la méthode X12-ARIMA.

### 2.1. Bref rappel sur la méthode X12-ARIMA

X12-ARIMA est l'une des deux méthodes de désaisonnalisation dont l'usage est préconisé dans les codes de bonnes pratiques en matière de désaisonnalisation établis en 2009 par Eurostat<sup>11</sup>, l'autre méthode étant TRAMO-SEATS. X12-ARIMA est une méthode non paramétrique basée sur l'usage de moyennes mobiles et qui est composée de deux étapes :

- la phase de pré-ajustement<sup>12</sup> :  
Cette phase de pré-ajustement consiste en la modélisation de la série brute par **un modèle Reg-ARIMA**, i.e. un modèle de régression avec des résidus suivant un modèle ARIMA. La partie « modèle de régression » permet principalement de corriger la série brute des non linéarités (points atypiques<sup>13</sup>(AO, TC et LS) et effets « jours ouvrables »). Quant à la partie « modèle ARIMA », elle permet notamment de prolonger la série brute aux extrémités afin d'éviter l'utilisation de moyennes mobiles asymétriques dans l'algorithme X11.
- la désaisonnalisation (X11) :  
La méthode X11 un algorithme<sup>14</sup> reposant sur l'utilisation de moyennes mobiles judicieusement choisies afin d'extraire la composante saisonnière de la série linéarisée issue de la phase de pré-ajustement.

Dans le cadre de cette étude, nous autorisons l'utilisation des quatre jeux de régresseurs CJO suivants :

<sup>9</sup> La série « Ensemble » est désaisonnalisée de manière indirecte

<sup>10</sup> Jdemetra+ est l'outil privilégié recommandé par Eurostat et la Banque Centrale Européenne pour réaliser les ajustements saisonniers des statistiques publiques depuis décembre 2014

<sup>11</sup> <http://www.cmfb.org/pdf/2009-12-01%20ESS%20Guidelines%20on%20seasonal%20adjustment.pdf>

<sup>12</sup> cf. annexe 1 pour une description plus détaillée

<sup>13</sup> cf. annexe 2 pour une présentation détaillée des trois grands types de points atypiques

<sup>14</sup> cf. annexe 3 pour une présentation détaillée de l'algorithme

**Tableau 2.1 : les différents jeux de régresseurs CJO**

Jeu de régresseurs CJO	Composition
Leap_year	1 seul régresseur correspondant à l'effet « année bissextile » (leap year)
Regcjo_2	2 régresseurs : 1 régresseur (jours de la semaine, du lundi au vendredi, ouvrables) en contraste aux week-ends et jours fériés + effet « année bissextile » (leap year)
Regcjo_7	6 régresseurs : 5 régresseurs (lundis, mardis, mercredis, jeudis, vendredis ouvrables) en contraste aux week-ends et jours fériés + effet « année bissextile » (leap year).
Regcjo_1	7 régresseurs : 6 régresseurs (lundis, mardis, mercredis, jeudis, vendredis, samedis ouvrables) en contraste aux dimanches et jours fériés + effet « année bissextile » (leap year)

Ces différents jeux de régresseurs CJO ont été construits selon la méthodologie décrite dans le document « Régresseurs pour effets de calendrier : comment les construire, comment les choisir ? » (Attal-Toubert, 2012). Une description sommaire de la construction de ces jeux de régresseurs est donnée pour rappel dans l'Annexe 4.

De plus, dans le cadre de cette étude, pour déterminer le meilleur jeu de régresseurs « CJO » pour chaque série, nous utiliserons l'algorithme de sélection présentée dans le document d'Attal-Toubert et rappelée dans l'annexe 5.

## 2.2. La saisonnalité des séries

L'ensemble des séries présente une saisonnalité identifiable par la méthode X12-ARIMA. En effet, le test combiné de « présence d'une saisonnalité identifiable » fourni par X11 donne pour résultat la modalité « Facteur saisonnier identifiable » pour l'ensemble des séries. Ce résultat est confirmé par la valeur prise par la statistique M7<sup>15</sup> pour chaque série ( $M7 < 0,5$  - cf. tableau 2.2.) : cela traduit le fait que la saisonnalité stable est relativement plus importante que la saisonnalité mobile et ainsi la saisonnalité est identifiable par X11.

Il est ainsi justifié de réaliser la désaisonnalisation des séries considérées.

**Tableau 2.2. : la présence d'une saisonnalité identifiable par X11**

Niveau d'agrégation de la NA		Test combiné de saisonnalité identifiable	M7(12/2008)
A10	A21		
BE		Present	0,30
	C	Present	0,31
FZ		Present	0,23
GI		Present	0,21
	G	Present	0,19
	H	Present	0,39
	I	Present	0,14
JZ		Present	0,31
KZ		Present	0,22
LZ		Present	0,22
MN		Present	0,20
OQ		Present	0,31
RU		Present	0,44

<sup>15</sup> Cette statistique fournie par X11 compare, à partir des statistiques de test correspondantes, la part relative des saisonnalités stable et mobile : plus cette statistique sera petite, plus la part de la saisonnalité stable sera importante, et ainsi plus la saisonnalité sera identifiable par X11



Dans un premier temps, nous avons réalisé un traitement « automatique », puis, à la lecture des notes du bilan qualité<sup>16</sup> et des diagnostics disponibles dans « Jdemetra+ », nous avons changé certaines spécifications afin d'obtenir des traitements de qualité satisfaisante.

### 2.3. La phase de pré-ajustement

Nous résumons dans le tableau ci-dessous l'ensemble des caractéristiques de la phase de pré-ajustement pour chacune des séries :

Tableau 2.3. : Les caractéristiques des pré-ajustements

Niveau d'agrégation de la NA		Modèle ARIMA (p,d,q) (P,D,Q)	Schéma de décomposition (A : additif ; M : multiplicatif)	Points atypiques	Jeu de régresseurs « CJO »
A10	A21				
BE		(2 1 0)(0 1 1)	M	LS (4-2008)	Regcjo_2
	C	(2 1 0)(0 1 1)	M	LS (4-2008)	Regcjo_7
FZ		(2 1 0)(0 1 0)	M		Regcjo_2
GI		(1 1 0)(0 1 0)	M		Regcjo_2
	G	(1 1 0)(0 1 1)	M		Regcjo_2
	H	(0 1 1)(0 1 1)	A		Regcjo_2
	I	(2 1 1)(0 1 1)	M		Regcjo_2
JZ		(0 1 1)(0 1 1)	M		Regcjo_2
KZ		(1 1 0)(0 1 1)	M	AO (12-2004) AO (1-2005)	Regcjo_2
LZ		(0 1 1)(0 1 1)	M		Regcjo_2
MN		(2 1 0)(0 1 1)	M	TC (12-2007)	Regcjo_2
OQ		(1 1 1)(0 1 1)	M		Regcjo_2
RU		(0 1 1)(0 1 1)	M	LS (8-2007)	Regcjo_2

Il est à noter que l'introduction d'un jeu de régresseurs « CJO » a été nécessaire pour corriger les effets « jours ouvrables » pour toutes les séries. De plus, excepté pour la série « H : Transports et entreposage », le schéma de décomposition retenu est multiplicatif.

Le pré-ajustement est de très bonne qualité (toutes les notes portant sur le pré-ajustement sont égales à 20) pour la grande majorité des séries. En effet, seules 2 d'entre elles présentent une note globale inférieure à 20 : la série « KZ : Activités financières et d'assurance » et la série « I : hébergement et restauration ». Pour la seconde, la note globale est tout de même de bonne qualité (18,57), le test d'homoscédasticité des résidus étant rejeté au niveau 5% mais accepté au niveau 1%. Quant à la série « KZ : Activités financières et d'assurance », la note globale est plus faible (14,29). Cela est dû principalement à deux faits :

- la présence de deux Additive Outliers consécutifs en décembre 2004 et en janvier 2005 : cependant, la présence de ces outliers est justifiée, les coefficients associés étant très significatifs (la t-stat étant égale à -15,33 pour le premier et à -8,68 pour le second). De plus, ce choix est confirmé par une lecture graphique.
- la rejet de l'hypothèse de normalité des résidus au niveau 1% (p-value = 0,5%). Cependant, dans la méthode X12-ARIMA, l'hypothèse de normalité des résidus n'est pas « cruciale » pour assurer une désaisonnalisation de bonne qualité. Elle permet simplement de pouvoir utiliser les tests de significativité des coefficients du modèle Reg-ARIMA.

Ainsi, on peut considérer que le pré-ajustement est de qualité satisfaisante pour l'ensemble des séries.

<sup>16</sup> cf. le document « Comment juger de la qualité d'une désaisonnalisation ? » (Himpens, 2012) pour une description détaillée des notes disponibles dans ce bilan. Dans le cadre de cette étude, toutes les notes sont ramenées à 20 (cf. annexe 6 pour plus de détails)

## 2.4. La désaisonnalisation (X11)

Nous récapitulons dans le tableau ci-dessous les principales caractéristiques de l'algorithme X11 utilisé pour réaliser la désaisonnalisation, i.e. la moyenne mobile d'Henderson utilisée pour extraire la tendance et la moyenne mobile utilisée pour extraire la composante saisonnière :

**Tableau 2.4. : Les caractéristiques de la désaisonnalisation (X11)**

Niveau d'agrégation de la NA		Nombre de termes de la moyenne mobiles de Henderson utilisée finale	Moyenne mobile finale utilisée pour extraire la composante saisonnière
A10	A21		
BE		13	3x3
	C	23	3x3
FZ		13	3x3
GI		13	3x5
	G	13	3x5
	H	23	3x5
	I	13	3x5
JZ		13	3x5
KZ		13	3x5
LZ		13	3x5
MN		13	3x5
OQ		13	3x5
RU		13	3x5

On constate que pour la grande majorité des séries, X11 a réussi à extraire la tendance en utilisant la moyenne mobile d'Henderson usuelle (moyenne mobile d'Henderson sur 13 termes). Cependant, pour 2 séries, « C : Industrie manufacturière » et « H : Transports et entreposage », l'utilisation d'une moyenne mobile plus longue ayant un pouvoir lissant plus important s'est avéré nécessaire (la moyenne mobile d'Henderson sur 23 termes). Quant à la moyenne mobile utilisée pour extraire la composante saisonnière, il s'agit des moyennes mobiles utilisées usuellement dans l'algorithme X11 (« 3x3 » ou « 3x5 »).

On s'intéresse dans un second temps à la qualité des désaisonnalisations réalisées. Une désaisonnalisation est, par définition, de qualité satisfaisante si la série CVS-CJO obtenue à l'issue du traitement ne présente pas d'effet saisonnier résiduel. C'est le cas pour toutes les séries du niveau A10 et A21 de la nomenclature NA. L'algorithme X11 a ainsi réussi à extraire correctement la composante saisonnière de toutes les séries brutes.

Il reste ensuite à s'assurer que le traitement réalisé permet de corriger correctement les séries brutes des effets « jours ouvrables ». La grande majorité des séries CVS-CJO obtenues ne présente pas d'effets « jours ouvrables » résiduels. En effet, seules deux séries, « KZ : Activités financières et d'assurance » et « BE : Industrie » semblent présenter un effet « jours ouvrables » résiduel à la lecture des résultats du test correspondant (pour ces deux séries, le test d'effets « jours ouvrables » résiduel est rejeté au niveau 5%). Cependant, les spectres associés aux séries CVS-CJO ne présentant pas de pic significatif à la fréquence « jours ouvrables », on peut considérer ces effets « jours ouvrables » résiduels comme négligeables.

D'autre part, la série « Ensemble » désaisonnalisée de manière indirecte ne présente ni de saisonnalité résiduelle, ni d'effets « jours ouvrables » résiduels. Ainsi, la désaisonnalisation et la correction des effets « jours ouvrables » est de qualité acceptable pour l'ensemble des séries.

## 2.5. Les révisions

Il n'existe pas de critère bien établi permettant de déterminer ce qu'est une série fortement révisée. En effet, une série instable sera par nature plus sujette à des révisions qu'une série présentant un profil plus stable au cours du temps. Nous quantifierons les révisions au moment de comparer la qualité des traitements avant et après l'apparition du régime des « auto-entrepreneurs ».

### 3. L'après « auto-entrepreneurs » (2000- 06/2012)

#### 3.1. Les séries « créations d'entreprises y compris auto-entrepreneurs »

##### 3.1.1. La saisonnalité des séries

L'ensemble des séries présente une saisonnalité identifiable par la méthode X12-ARIMA, en dépit de l'apparition du régime d'« auto-entrepreneurs ». Cependant, la note M7 se détériore sérieusement pour les séries relatives aux créations dans l'industrie (« BE : industrie » et « C : industrie manufacturière ») : une hausse de 0,26 pour la série « BE : industrie » et de 0,15 pour la série « C : industrie manufacturière » (cf. tableau 3.1). Ainsi, la part relative de la saisonnalité évolutive augmente pour ces deux séries mais reste cependant suffisamment faible pour laisser la saisonnalité identifiable par X11.

On peut donc a priori envisager de désaisonnaliser toutes les séries en utilisant la méthode X12-ARIMA.

Tableau 3.1. : la présence d'une saisonnalité identifiable par X11

Niveau d'agrégation de la NA		Test combiné de saisonnalité identifiable	M7(06/2012)	M7(06/2012)-M7(12/2008)
A10	A21			
BE		Present	0,56	<b>0,26</b>
	C	<i>Present</i>	<i>0,47</i>	<b>0,15</b>
FZ		Present	0,26	0,03
GI		Present	0,21	0,00
	G	<i>Present</i>	<i>0,21</i>	<i>0,01</i>
	H	<i>Present</i>	<i>0,34</i>	<i>-0,04</i>
	I	<i>Present</i>	<i>0,15</i>	<i>0,01</i>
JZ		Present	0,36	0,05
KZ		Present	0,16	-0,07
LZ		Present	0,20	-0,02
MN		Present	0,22	0,01
OQ		Present	0,29	-0,02
RU		Present	0,45	0,02

##### 3.1.2. La phase de pré-ajustement

Pour toutes les séries très concernées par l'auto-entrepreneuriat (i.e. toutes les séries exceptées « H : transports et entreposage », « KZ : activités financières et d'assurance » et « LZ : Activités immobilières »), l'apparition du nouveau régime est pris en compte au travers d'au moins un Level Shift (LS) portant sur le début de l'année 2009 (cf. tableau 3.2). Les séries pour lesquelles le choc en niveau est le plus violent, le « saut » est modélisé soit :

- par un LS et un TC en janvier 2009 : il s'agit des séries « GI : Commerce, transports, hébergement et restauration », « G : Commerce ; réparation d'autos et de motos » et « MN : soutien aux entreprises »
- par un LS et un AO en janvier 2009 : c'est le cas de la série « RU : services aux ménages »

Tableau 3.2. : Les caractéristiques des pré-ajustements

Niveau d'agrégation de la NA		Modèle ARIMA (p,d,q) (P,D,Q)	Schéma de décomposition (A : additif ; M : multiplicatif)	Points atypiques	Jeu de régresseurs « CJO »
A10	A21				
BE		(3 1 1)(0 1 1)	M	LS (1-2009) AO (12-2009)	Regcjo_7
	C	(0 1 1)(0 1 1)	M	LS (1-2009) LS (6-2010) LS (4-2008) LS (3-2007)	Regcjo_1
FZ		(3 1 1)(0 1 1)	M	TC (11-2008) LS (2-2009) LS (12-2010) TC (6-2010)	Regcjo_2
GI		(1 1 0)(0 1 1)	M	LS (1-2009) TC (1-2009)	Regcjo_7
	G	(1 1 0)(0 1 1)	M	LS (1-2009) TC (1-2009)	Regcjo_7
	H	(0 1 1)(0 1 1)	A		Regcjo_2
	I	(0 1 1)(0 1 1)	M	LS (2-2009)	Regcjo_7
JZ		(0 1 1)(0 1 1)	M	LS (1-2009)	Regcjo_2
KZ		(0 1 1)(0 1 1)	M	TC (12-2004) LS (12-2007) TC (1-2012) LS (2-2005)	Regcjo_7
LZ		(0 1 1)(0 1 1)	M	AO (11-2011)	Regcjo_2
MN		(0 1 1)(0 1 1)	M	TC (12-2007) LS (1-2009) TC (1-2009) LS (7-2009) LS (12-2010)	Regcjo_2
OQ		(2 0 1)(0 1 1)	M	LS (3-2009)	Regcjo_2
RU		(1 1 1)(0 1 1)	M	LS (1-2009) AO (1-2009)	Regcjo_2

Ceci a pour conséquence une forte détérioration de la qualité de la modélisation. En effet, pour l'ensemble des séries, on observe une hausse substantielle du AIC (entre +42% et +56% - cf. tableau 3.3).

Tableau 3.3. : Les AIC

Niveau d'agrégation de la NA		AIC (06/2012)	Taux de croissance du AIC entre 06/2012 et 12/2008
A10	A21		
BE		1 689,35	55,91%
	C	1 620,91	51,64%
FZ		1 937,97	49,88%
GI		2 049,00	50,13%
	G	1 989,75	51,12%
	H	1 498,54	44,89%
	I	1 627,15	48,10%
JZ		1 677,91	54,38%
KZ		1 494,62	42,04%
LZ		1 598,43	45,95%
MN		1 986,13	51,88%
OQ		1 870,24	51,73%
RU		1 859,03	54,74%

La qualité de la phase de pré-ajustement s'est ainsi fortement détériorée suite à l'apparition du régime d'« auto-entrepreneurs ».

### 3.1.3. La désaisonnalisation (X11)

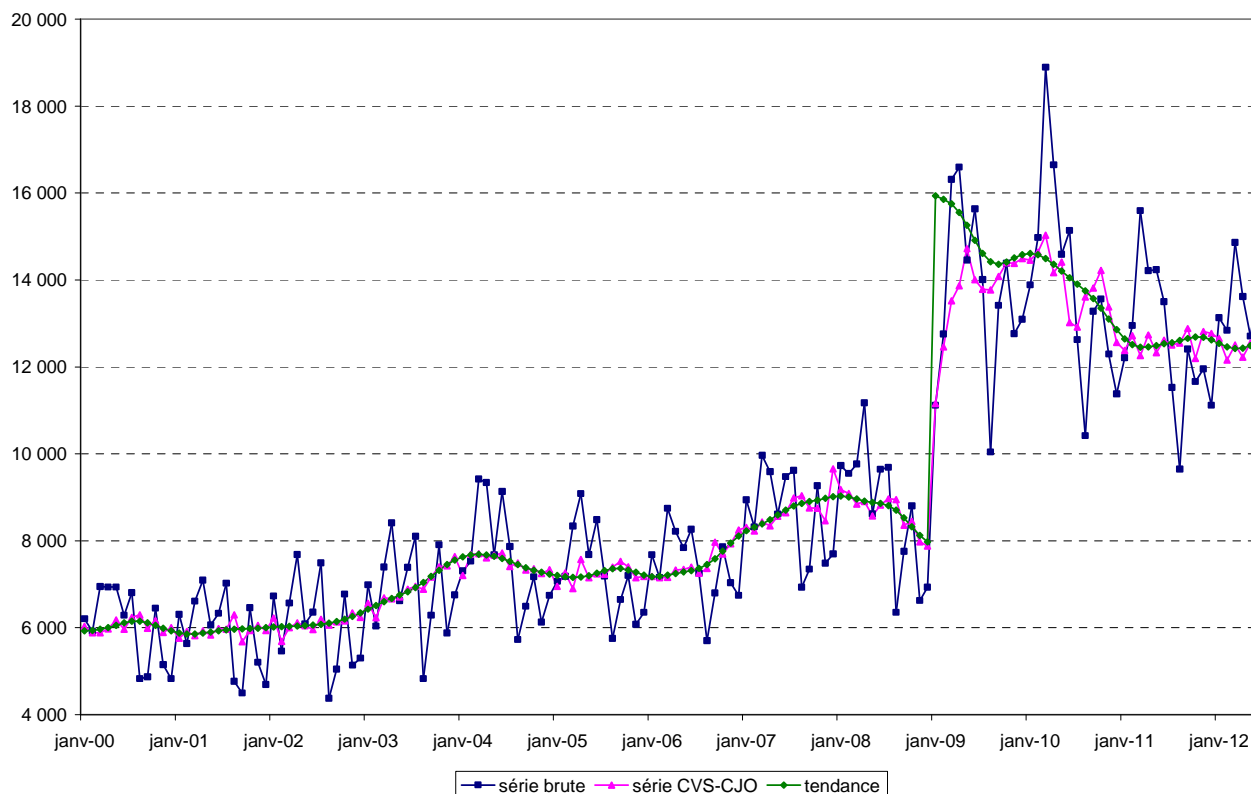
Pour l'ensemble des séries, l'algorithme X11 utilise une moyenne mobile de Henderson sur 13 termes pour extraire la tendance de la série brute. Quant à la moyenne mobile utilisée pour extraire la composante saisonnière, il s'agit des mêmes moyennes mobiles que celles utilisées avant l'apparition du nouveau régime.

Cependant, la qualité de la désaisonnalisation s'est sensiblement détériorée pour 6 séries. Pour les 3 séries, « **C : industrie manufacturière** », « **KZ : activités financières et d'assurance** », « **OQ : enseignement, santé, action sociale** », le test de saisonnalité résiduelle conclut à la présence d'effets saisonniers résiduels significatifs sur la série CVS-CJO lorsqu'on considère l'ensemble de la période. Ce résultat est confirmé par la présence de pics significatifs aux fréquences saisonnières sur le spectre de la série CVS-CJO ou de l'irrégulier. Pour la série « **LZ : activités immobilières** », la désaisonnalisation est encore de plus mauvaise qualité car le test de saisonnalité résiduelle conclut à la présence d'effets saisonniers résiduels significatifs sur la série CVS-CJO lorsqu'on considère la fin de période (i.e. les 3 dernières années).

Pour les séries « **FZ : construction** » et « **RU : services aux ménages** », seule l'analyse spectrale signale la présence d'effets saisonniers résiduels.

De plus, pour les 3 séries pour lesquelles le « saut » de janvier 2009 a été modélisé par un LS et un TC, une lecture graphique permet cette modélisation a pour conséquence la création d'un écart conséquent entre la tendance et la série CVS-CJO. Ainsi, la tendance ne « passe » plus à travers la courbe CVS-CJO (cf. graphique 3.1 pour la série « GI : Commerce, transports, hébergement et restauration »). Ainsi, la désaisonnalisation n'est pas de qualité satisfaisante pour ces séries.

**Graphique 3.1. : la série des créations d'entreprises (les « auto-entrepreneurs » inclus) pour « GI : Commerce, transports, hébergement et restauration » (2000-06/2012)**



### 3.1.4. Les révisions

Nous quantifions dans cette étude l'importance des révisions de court terme et des révisions de long terme. On appelle :

- révisions de court terme : la moyenne des révisions des taux de croissance en glissement mensuel de la série CVS-CJO 1 mois, 2 mois et 3 mois après la première publication
- révisions de long terme : les révisions des taux de croissance en glissement mensuel de la série CVS-CJO 12 mois après la première publication

Pour mesurer l'importance des révisions, on se concentre sur les deux dernières années et on calcule les RMSE entre les taux de croissance (exprimés en %) suivants :

$$\forall k \in \{1, 2, 3, 12\},$$

$$\text{RMSE}_{\text{rev}(k_{\text{mois}})} = \sqrt{\frac{1}{25 - k} \sum_{t=\text{date\_finale}-25}^{\text{date\_finale}-k} [\text{tx}_{\text{gliss\_mens}}_t(t+k) - \text{tx}_{\text{gliss\_mens}}_t(t)]^2}$$

où :

- $\text{date\_finale}$  est la date de fin de la série considérée
- $\text{tx}_{\text{gliss\_mens}}_t(t+k)$  est le taux de croissance en glissement mensuel pour la date  $t$  calculée sur la série CVS-CJO obtenue à la date  $t+k$

Les révisions de court terme ont dans l'ensemble augmenté entre 2008 et juin 2012. Elles ont même explosé pour les séries « FZ : Construction » (+159%) et « LZ : activités immobilières » (+131%) (cf. tableau 3.4). Quant aux révisions de long terme, elles ont également dans l'ensemble augmenté (+49% notamment pour la série « G : Commerce, réparation d'autos et de motos »).

**Tableau 3.4. : Les taux de croissance des révisions entre 2008 et juin 2012**

Niveau d'agrégation de la NA		Révisions de court terme	Taux de croissance	Révisions de long terme	Taux de croissance
A10	A21				
BE		2,22%	38%	4,42%	38%
	C	2,25%	-9%	2,75%	-42%
FZ		1,68%	<b>159%</b>	2,99%	12%
GI		1,10%	61%	2,04%	37%
	G	0,64%	45%	2,78%	<b>49%</b>
	H	0,34%	-20%	1,97%	-37%
	I	0,68%	-12%	2,56%	21%
JZ		0,87%	-17%	4,22%	23%
KZ		1,31%	35%	2,02%	-48%
LZ		2,64%	<b>131%</b>	2,65%	6%
MN		1,15%	54%	2,58%	27%
OQ		0,76%	-4%	3,53%	19%
RU		0,88%	-43%	2,54%	-25%

A la lumière de ces résultats, il a été décidé de ne pas désaisonnaliser directement les séries des créations d'entreprises en 2012. Cependant, étant donné que le nombre de données disponibles pour les « auto-entrepreneurs » commençait à devenir suffisant<sup>17</sup> pour envisager de réaliser une désaisonnalisation et une correction de jours ouvrables sur ces séries, la stratégie de désaisonnalisation adoptée depuis la publication de septembre 2009 pouvait être remise en question. Cette stratégie consistait à définir, pour chaque secteur, la série CVS-CJO des créations d'entreprises comme la somme de la série CVS-CJO des créations d'entreprises hors auto-entrepreneurs et de la série brute des créations correspondant à des « auto-entrepreneurs ».

### **3.2. Les séries des créations d'entreprises correspondant à des auto-entrepreneurs**

#### **3.2.1. La saisonnalité des séries**

Seules 2 séries d'auto-entrepreneurs présentent une saisonnalité identifiable la méthode X12-ARIMA (cf. tableau 3.5). Il s'agit des séries « BE : Industrie » et « G : commerce ; réparation d'autos et de motos ». Ces deux secteurs sont concernés par l'apparition du régime d'« auto-entrepreneurs » mais la saisonnalité de ces séries reste assez stable a priori pour pouvoir être identifiée par la méthode X12-ARIMA.

<sup>17</sup> Les codes de bonnes pratiques en ajustement saisonnier considère qu'une série peut être désaisonnalisée à partir du moment où on dispose d'au moins 3 années de données

**Tableau 3.5. : la présence d'une saisonnalité identifiable par X11 pour les séries d'auto-entrepreneurs (2009 - juin 2012)**

Niveau d'agrégation de la NA		Test combiné de saisonnalité identifiable
A10	A21	
BE		Present
	C	<i>Probablement pas</i>
FZ		Probablement pas
GI		Probablement pas
	G	<i>Present</i>
	H	<i>Probablement pas</i>
	I	<i>Probablement pas</i>
JZ		Probablement pas
KZ		Non présent
LZ		Non présent
MN		Non présent
OQ		Probablement pas
RU		Probablement pas

### 3.2.2. La phase de pré-ajustement

Pour les séries courtes, la phase de pré-ajustement revêt une importance beaucoup plus grande. En effet, en raison de la longueur de la série (3 ans et demi), la désaisonnalisation réalisée par X12-ARIMA repose davantage sur les prévisions effectuées à l'aide du modèle Reg-ARIMA aux extrémités de la série. Cependant, étant donné le faible nombre de points disponibles, la mise au point d'une modélisation Reg-ARIMA de qualité satisfaisante peut s'avérer délicate.

Pour 7 séries (« FZ », « H », « JZ », « LZ », « MN », « OQ » et « RU »), le modèle Reg-ARIMA identifié automatiquement est (1 0 0) (1 0 0) et (1 0 0) (0 0 0) pour la série « KZ » (cf. tableau 3.6) : la modélisation de la partie saisonnière en (1 0 0) ou en (0 0 0) confirme la présence d'une saisonnalité non encore parfaitement établie pour ces séries. De plus, le modèle est potentiellement instable pour la série « C » : en effet, le coefficient MA saisonnier pour la série « C » étant très proche de 1 (0,998), une simplification est possible avec l'ordre de différentiation saisonnier.

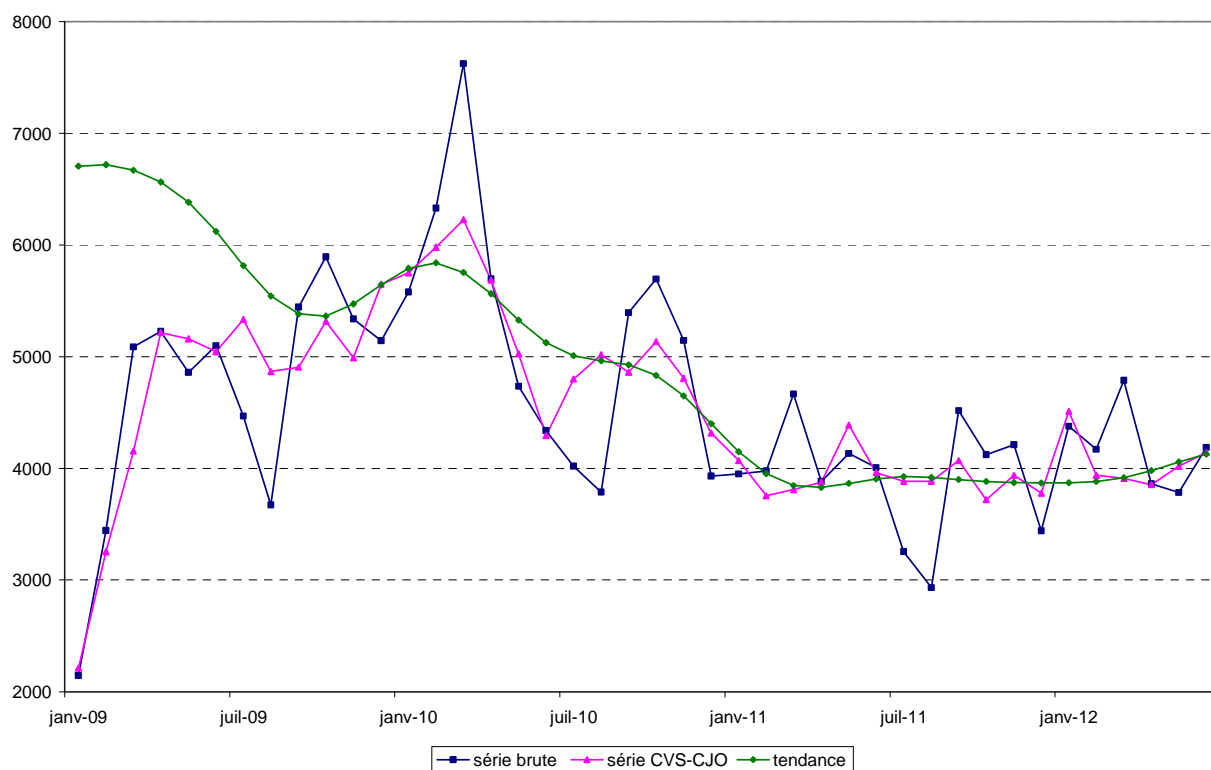
D'autre part, pour les séries « H », « MN », « OQ » et « RU », la mise en place du régime est modélisée à l'aide d'un TC en janvier 2009 pour les quatre séries « H », « MN », « OQ » et « RU ». Cet outlier a un fort impact indésirable sur la tendance dans la mesure où elle ne « traverse » pas la série brute au début de l'année 2009 (cf. graphique 3.2. pour la série « RU »).



Tableau 3.6. : Les caractéristiques des pré-ajustements (2009 - juin 2012)

Niveau d'agrégation de la NA		Modèle ARIMA (p,d,q) (P,D,Q)	Schéma de décomposition (A : additif ; M : multiplicatif)	Points atypiques	Jeu de régresseurs « CJO »
A10	A21				
BE		(0 1 0)(0 1 1)		M	Regcjo_2
	C	(0 1 1)(0 1 1)		M	
FZ		(1 0 0)(1 0 0)		A	
GI		(0 1 0)(0 1 1)		M	
	G	(0 1 0)(0 1 1)		M	
	H	(1 0 0)(1 0 0)		M	TC (1-2009)
	I	(1 0 1)(1 1 0)		A	LS (11-2011) AO (4-2009)
JZ		(1 0 0)(1 0 0)		M	Regcjo_1
KZ		(1 0 0)(0 0 0)		A	LS (3-2009)
LZ		(1 0 0)(1 0 0)		M	AO (1-2009)
MN		(1 0 0)(1 0 0)		A	TC (1-2009)
OQ		(1 0 0)(1 0 0)		A	TC (1-2009)
RU		(1 0 0)(1 0 0)		M	TC (1-2009)

Graphique 3.2. : la série des créations d'entreprises correspondant aux « auto-entrepreneurs » pour « RU : Services aux ménages » (2009-06/2012)



### 3.2.3. La désaisonnalisation (X11)

Pour quatre séries, X11 n'arrive pas à réaliser une décomposition de bonne qualité ( $Q \geq 1$  ou  $Q-M2 \geq 1$  - cf. tableau 3.7.). Il s'agit des séries « H », « KZ », « LZ » et « MN ».

**Tableau 3.7. : Les séries d'« auto-entrepreneurs » pour lesquelles la qualité de décomposition est mauvaise ( $Q \geq 1$  ou  $Q-M2 \geq 1$ ) (2009 - juin 2012)**

Niveau d'agrégation de la NA		Q	Q-M2
A10	A21		
	H	1,68	1,42
KZ		1,64	1,62
LZ		1,73	1,55
MN		1,25	0,90

Cependant, X11 arrive à estimer une composante saisonnière d'assez bonne qualité pour qu'il n'y ait pas d'effets saisonniers résiduels sur la série CVS-CJO ou sur l'irrégulier, pour toutes les séries. En effet, le test de saisonnalité résiduelle ne détecte pas la présence de tels effets résiduels et il n'y a pas de pic spectral à la fréquence saisonnière sur le spectre de la série CVS-CJO ou de l'irrégulier.

Il a été finalement décidé en 2012 de garder la stratégie mise en place, à savoir ne pas désaisonnaliser les séries de créations d'« auto-entrepreneurs », les désaisonnalisations de ces séries n'étant pas jugées de qualité satisfaisante. En septembre 2013, 14 points supplémentaires (de juillet 2012 à août 2013) étaient disponible pour les séries d'« auto-entrepreneurs ». Il a ainsi été décidé de mener une nouvelle étude afin de déterminer si cette stratégie pouvait être remise en question

## **4. La première désaisonnalisation de qualité satisfaisante pour les auto-entrepreneurs (août 2013)**

### **4.1. Une saisonnalité plus identifiable**

L'ajout de 14 points supplémentaires a permis de rendre la saisonnalité des séries de créations d'« auto-entrepreneurs » identifiable par X11 pour un grand nombre de secteurs (cf. les résultats du test combiné de saisonnalité identifiable dans le tableau 4.1.). En effet, pour seulement 4 secteurs (« H », « KZ », « LZ » et « MN »), le test n'aboutit pas à la présence d'une saisonnalité identifiable. Pour les secteurs « H : Transports et entreposage », « KZ : Activités financières et d'assurance » et « LZ : Activités immobilières », ce résultat s'explique par le fait qu'ils soient très peu concernés par l'auto-entreprenariat. Il a ainsi été décidé de ne pas désaisonnaliser les séries d'« auto-entrepreneurs » relatives à ces trois secteurs. La série des auto-entrepreneurs relative au secteur « MN : Soutien aux entreprises » a été quant à elle désaisonnalisée : l'analyse du profil saisonnier de la série a révélé une saisonnalité en août.

Tableau 4.1. : la présence d'une saisonnalité identifiable par X11 pour les séries d'auto-entrepreneurs (2009 - août 2013)

Niveau d'agrégation de la NA		Test combiné de saisonnalité identifiable
A10	A21	
BE		Present
	C	Present
FZ		Present
GI		Present
	G	Present
	H	<b>Probablement pas</b>
	I	Present
JZ		Present
KZ		<b>Non présent</b>
LZ		<b>Non présent</b>
MN		<b>Non présent</b>
OQ		Present
RU		Present

## 4.2. La phase de pré-ajustement

Les modèles ARIMA identifiés automatiquement au cours de la phase de pré-ajustement confirment la présence d'une saisonnalité davantage établie pour toutes les séries, excepté la série « MN : soutien aux entreprises » (cf. tableau 4.2.).

D'autre part, l'allongement des séries permet à l'algorithme de sélection du meilleur jeu de régresseurs CJO de détecter un plus grand nombre d'effets « jours ouvrables ». En effet, il conduit à introduire un jeu de régresseurs « CJO » dans le modèle pour 8 séries, alors que ce n'était le cas que pour 2 séries en juin 2012 : à cette date, en raison de la faible longueur des séries (moins de 4 ans), l'hypothèse nulle d'absence d'effets « CJO » était rarement rejetée.

Enfin, pour les 2 séries « MN » et « RU », la mise en place du régime est toujours modélisée à l'aide d'un TC en janvier 2009 (ce n'est plus le cas pour la série « OQ »). Cet outlier a toujours le même impact sur la tendance pour ces séries.

Tableau 4.2. : Les caractéristiques des pré-ajustements (2009 - août 2013)

Niveau d'agrégation de la NA		Modèle ARIMA (p,d,q) (P,D,Q)	Schéma de décomposition (A : additif ; M : multiplicatif)	Points atypiques	Jeu de régresseurs « CJO »
A10	A21				
BE		(1 0 2)(1 1 0)	M		Regcjo_7
	C	(1 0 2)(0 1 1)	M		Regcjo_7
FZ		(2 0 0)(0 1 1)	A	AO (6-2010) AO (3-2010) LS (12-2009)	Regcjo_2
GI		(1 0 0)(1 1 0)	M		Regcjo_2
	G	(0 1 0)(0 1 1)	M		Regcjo_2
	I	(0 0 2)(1 1 0)	A		Regcjo_2
JZ		(0 1 0)(0 1 1)	M		Regcjo_7
MN		(1 0 0)(1 0 0)	A	TC (1-2009)	
OQ		(0 1 0)(0 1 1)	A		
RU		(0 1 0)(1 1 0)	M	TC (1-2009)	Regcjo_2

Pour toutes les séries, excepté la série « FZ : Construction », le pré-ajustement est de très bonne qualité : toutes les notes du bilan qualité portant sur le pré-ajustement sont égales à 20). Le bilan qualité nous alerte sur la concentration sur l'année 2010 de deux outliers. Après analyse, il s'avère que seule l'AO de juin 2010 est justifié.

### 4.3. La désaisonnalisation (X11)

La qualité de la décomposition réalisée par X11 est maintenant de qualité satisfaisante pour toutes les séries ( $Q < 1$  et  $Q-M2 < 1$ ), excepté la série « MN ». Cependant, Q est très proche de 1 pour cette série ( $Q=1,02$ ) et Q-M2 est strictement inférieur à 1 ( $Q-M2=0,62$ ).

De plus, pour l'ensemble des séries, l'analyse spectrale dans Jdemetra+ et les tests réalisés dans le bilan qualité ne détectent pas la présence d'effets saisonniers résiduels ou d'effets « jours ouvrables » résiduels sur la série CVS-CJO pour tous les secteurs, excepté les secteurs « OQ » et « FZ ». La série CVS-CJO du secteur « OQ : Enseignement, santé, action sociale » présente des effets jours ouvrables résiduels d'après le bilan qualité (la note associée au test étant égale à 0). Ces effets résiduels disparaissent avec l'introduction du jeu de régresseurs regcjo\_1 dans la modélisation Reg-ARIMA.

Quant à la série CVS-CJO relative au secteur « FZ : Construction », elle présente seulement des effets saisonniers résiduels quand on considère l'ensemble de la période (2009 - août 2013, soit 4 ans et 8 mois). En ne retenant que les 3 dernières années, il n'y en a en effet pas d'effets saisonniers résiduels (la note associée dans le bilan qualité étant égale à 15). La suppression du LS détecté en décembre 2009 a notamment permis de résoudre ce problème.

Après une expertise plus approfondie, il a été décidé de modifier légèrement les spécifications automatiques déterminées par X12-ARIMA de la manière suivante :

- à la vue de la représentation graphique des séries, un schéma de décomposition multiplicatif a été retenu pour l'ensemble des séries
- la mise en place du régime a été modélisée par deux AO en janvier 2009 et en février 2009 pour tous les secteurs, exceptés les secteurs « BE : industrie », « C : industrie manufacturière » et « FZ : Construction »<sup>18</sup>

### 4.4. Vers une nouvelle stratégie de désaisonnalisation

Ainsi, suite à la campagne annuelle de désaisonnalisation réalisée en septembre 2013, la stratégie de désaisonnalisation des séries de créations d'entreprises a été modifiée. En effet, il a été décidé, pour chaque secteur s, excepté les secteurs « H », « KZ » et « LZ », de calculer la série CVS-CJO correspondant à l'ensemble des créations d'entreprises, notée  $creat\_ent_{m,A}^{CVS-CJO}(s)$ , en utilisant la relation suivante :

$$\forall s \in S \setminus \{H, LZ, KZ\}, \quad creat\_ent_{m,A}^{CVS-CJO}(s) = creat\_auto\_ent_{m,A}^{CVS-CJO}(s) + creat\_hauto\_ent_{m,A}^{CVS-CJO}(s)$$

où :

- S est l'ensemble des secteurs considérés
- $creat\_auto\_ent_{m,A}^{CVS-CJO}(s)$  est la série CVS-CJO de l'ensemble des créations d'entreprises correspondant à des auto-entrepreneurs
- $creat\_hauto\_ent_{m,A}^{CVS-CJO}(s)$  est la série CVS-CJO de l'ensemble des créations d'entreprises hors auto-entrepreneurs

<sup>18</sup> Pour le secteur « FZ », seul un AO en janvier 2009 a été introduit dans le modèle

Pour les 3 secteurs « H », « KZ » et « LZ », on considère la série brute de l'ensemble des créations d'entreprises (pour les raisons évoquées au §4.1.) :

$$\forall s \in \{H, LZ, KZ\}, \text{creat\_ent}_{m,A}^{\text{CVS-CJO}}(s) = \text{creat\_auto\_ent}_{m,A}^{\text{brute}}(s) + \text{creat\_hauto\_ent}_{m,A}^{\text{CVS-CJO}}(s)$$

Pour rappel, la série CVS-CJO correspondant à l'ensemble des créations d'entreprises, tous secteurs confondus, est obtenue en utilisant la méthode indirecte :

$$\text{creat\_ent}_{m,A}^{\text{CVS-CJO}} = \sum_{s \in A10 \setminus \{AZ\}} \text{creat\_ent}_{m,A}^{\text{CVS-CJO}}(s)$$

Cette série CVS-CJO ne présente pas d'effets saisonniers résiduels (le test combiné de saisonnalité identifiable aboutissant à la modalité « non présente »), ni d'effets « jours ouvrables » résiduels (le test d'effets « jours ouvrables » résiduels n'est pas rejeté au niveau 5%). Il en est de même pour les séries des créations d'entreprises y compris auto-entrepreneurs pour tous les secteurs.

Ces résultats valident la nouvelle stratégie de désaisonnalisation adoptée fin 2013.

## Conclusion

Suite à l'apparition du régime des « auto-entrepreneurs » en janvier 2009, il s'est avéré nécessaire de distinguer la séries des créations correspondant à ce nouveau statut et la séries des créations « hors auto-entrepreneurs » pour chaque secteur au moment de la désaisonnalisation des séries. Cependant, les séries des « auto-entrepreneurs » étant courtes, la qualité de leur désaisonnalisation n'a été acceptable qu'à partir d'août 2013. Ainsi, avant cette date, on considérait les séries brutes correspondant au nouveau statut au moment du calcul de la série d'ensemble CVS-CJO.

Cependant, une série étant considérée comme courte jusqu'au moment où 7 années de données sont disponibles, il faut s'attendre à observer des changements dans les spécifications au moment des campagnes annuelles des prochaines années et observer avec attention les révisions.

## Annexes

### Annexe 1 : la phase de pré-ajustement de la méthode X12-ARIMA

Le modèle Reg-ARIMA se présente sous la forme suivante dans le cas d'un schéma additif :

$$X_t = \sum_i \xi_i Z_{it} + \varepsilon_t$$

où :

- $(X_t)_t$  est la série brute
- $\varepsilon_t \rightarrow \text{ARIMA}(p, d, q)(P, D, Q)$  et suit une loi normale
- $\forall i, (Z_{i,t})_t$  sont les régresseurs (i.e. les points atypiques et les régresseurs « jours ouvrables ») et  $\xi_i$  les coefficients associés

Soit :

$$\boxed{\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D \left( X_t - \sum_i \xi_i Z_{it} \right) = \theta(B)\Theta(B^s)v_t}$$

avec :

- $s$  : la périodicité de la série (ici  $s=12$  car les séries de créations d'entreprises nationales sont mensuelles)
- $d$  et  $D$  : les ordres de différenciation simple et saisonnier du modèle ARIMA
- $B$  : l'opérateur retard ( $BX_t = X_{t-1}$ )
- $\phi$  et  $\Phi$  : les polynômes autorégressifs (AR) du modèle ARIMA :
  - o  $\phi(B) = 1 - \sum_{k=1}^p \varphi_k B^k$  avec  $\varphi_p \neq 0$
  - o  $\Phi(B^s) = 1 - \sum_{k=1}^P \Psi_k (B^s)^k$  avec  $\Psi_P \neq 0$
- $\theta$  et  $\Theta$  : les polynômes moyennes mobiles (MA) du modèle ARIMA :
  - o  $\theta(B) = 1 - \sum_{k=1}^q \delta_k B^k$  avec  $\delta_q \neq 0$
  - o  $\Theta(B^s) = 1 - \sum_{k=1}^Q \Delta_k (B^s)^k$  avec  $\Delta_Q \neq 0$
- $(v_t)_t$  est un bruit blanc faible de variance  $\sigma^2$  :  $v_t \sim \text{BB}(0; \sigma^2)$

Dans le cas d'un schéma multiplicatif, une transformation de la série brute par le logarithme est préalablement effectuée. Le modèle Reg-ARIMA se présente alors sous la forme suivante :

$$\boxed{\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D \left[ \ln(X_t) - \sum_i \xi_i Z_{it} \right] = \theta(B)\Theta(B^s)v_t}$$

## Annexe 2 : les différents types de points atypiques

Dans le cadre de cette étude, on autorise la recherche des trois principaux types de points atypiques :

- l'Additive Outlier (AO) ou « point aberrant » (cf. graphique 2.1) : le régresseur correspondant est une indicatrice qui prendra la valeur 1 à la date correspondant au « point aberrant ». Par exemple, si un Additive Outlier est détecté au mois  $m$  de l'année  $A$ , alors le régresseur correspondant à cette variable, noté  $AO(m; A)_t$ , sera défini de la manière suivante :

$$AO(m; A)_t = \begin{cases} 1 & \text{si } t = (m; A) \\ 0 & \text{sin on} \end{cases}$$

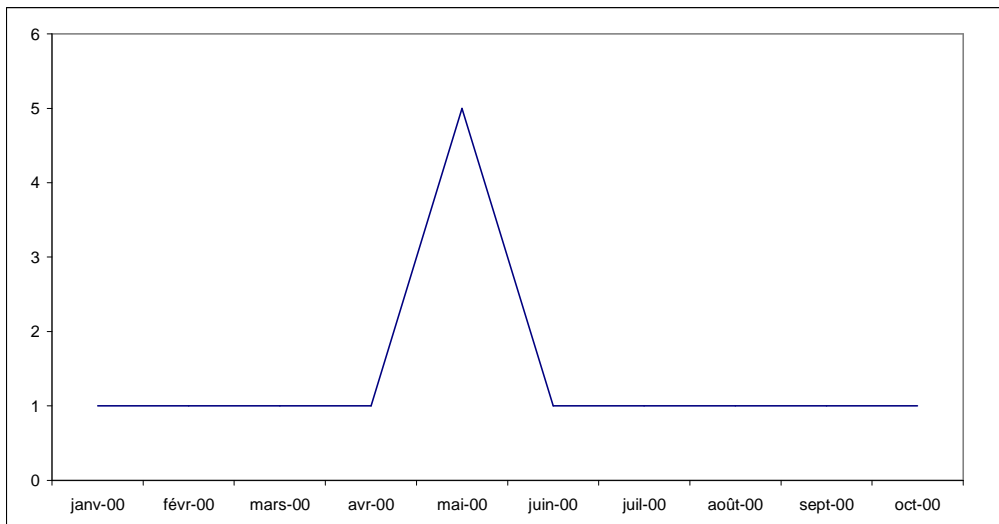
- le Transitory Change (TC) ou changement de niveau transitoire (cf. graphique 2.2) : le régresseur correspondant à un Transitory Change détecté en  $t_0$ , noté  $TC(t_0)_t$ , est défini de la manière suivante :

$$TC(t_0)_t = \begin{cases} 0 & \text{si } t < t_0 \\ 0,7^{t-t_0} & \text{si } t \geq t_0 \end{cases}$$

- le Level Shift (LS) ou changement de niveau (cf. graphique 2.3) : le régresseur correspondant à un Level Shift détecté en  $t_0$ , noté  $LS(t_0)_t$ , est défini de la manière suivante :

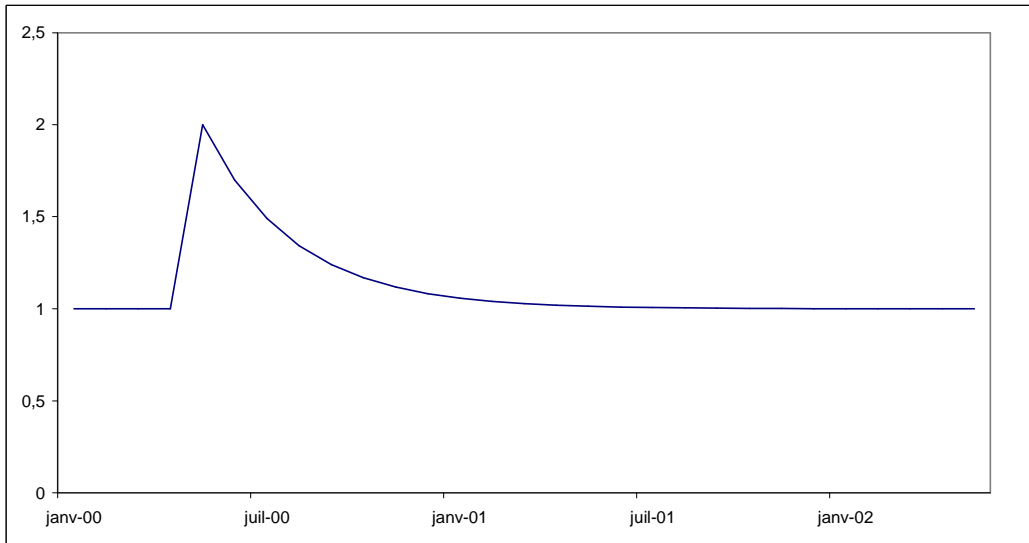
$$LS(t_0)_t = \begin{cases} -1 & \text{si } t < t_0 \\ 0 & \text{si } t \geq t_0 \end{cases}$$

**Graphique 1 : Exemple d'une série présentant un Additive Outlier en mai 2000**

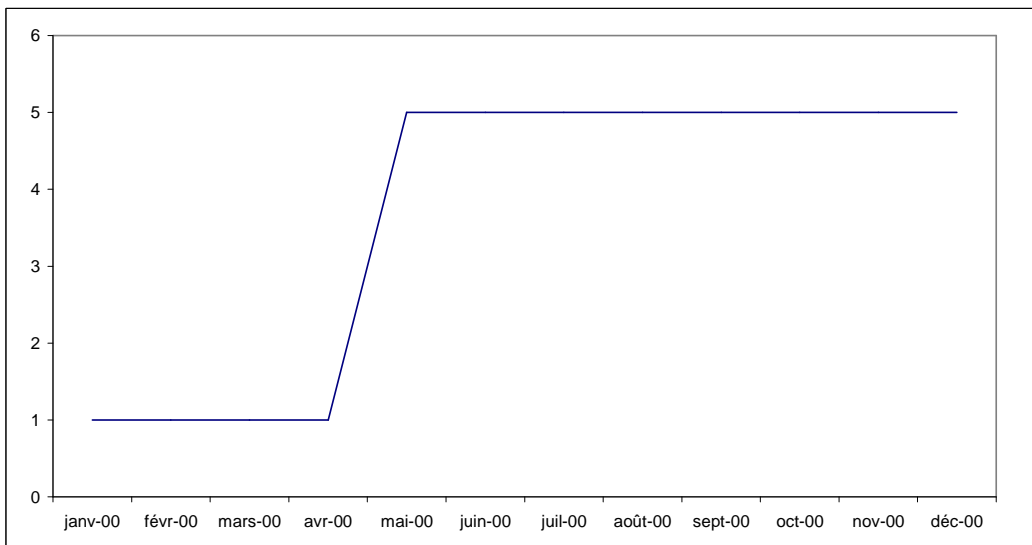




**Graphique 2 : exemple d'une série présentant un Transitory Change en mai 2000**



**Graphique 3 : exemple d'une série présentant un Level Shift en mai 2000**



### Annexe 3 : l'algorithme utilisé par X11

Dans X12-ARIMA, La désaisonnalisation est réalisée sur la série linéarisée obtenue à l'issue de la phase de pré-ajustement, i.e. sur la série brute corrigée des effets susceptibles de nuire à la qualité de la désaisonnalisation<sup>19</sup> (points atypiques, effets « jours ouvrables »,...). Cette série est ainsi définie de la manière suivante :

$$X\_linearisee_t = X_t - \sum_i \hat{\xi}_i Z_{it} \text{ dans le cas d'un schéma additif}$$

$$\ln(X\_linearisee_t) = \ln(X_t) - \sum_i \hat{\xi}_i Z_{it} \text{ dans le cas d'un schéma multiplicatif}$$

où les  $\hat{\xi}_i$  sont les estimateurs des  $\xi_i$

Avant de procéder à la désaisonnalisation, un choix de modèle de décomposition doit être fait selon le profil de la série considérée. Si on opte pour un schéma additif, alors on considérera le modèle suivant :

$$X\_linearisee_t = T_t + S_t + I_t$$

alors que si choisit un schéma multiplicatif, le modèle de décomposition se présentera sous la forme suivante :

$$X\_linearisee_t = T_t * S_t * I_t$$

où :

- $T_t$  est la tendance-cycle
- $S_t$  est la composante saisonnière
- $I_t$  est l'irrégulier

Dans X12-ARIMA, la désaisonnalisation est réalisée en utilisant la méthode X11. Cette méthode met en œuvre l'algorithme suivant qui repose sur l'utilisation de moyennes mobiles judicieusement choisies (nous nous plaçons ici dans le cas d'une série mensuelle pour laquelle un schéma additif a été retenue) :

#### 1. Estimation de la Tendence-Cycle par moyenne mobile 2x12 :

$$T_t^{(1)} = M_{2 \times 12}(X\_linearisee_t)$$

#### 2. Estimation de la composante Saisonnier-Irrégulier :

$$(S_t + I_t)^{(1)} = X\_linearisee_t - T_t^{(1)}$$

#### 3. Estimation de la composante saisonnière par moyenne mobile 3x3 sur chaque mois :

$$S_t^{(1)} = M_{3 \times 3}[(S_t + I_t)^{(1)}]$$

---

<sup>19</sup> Les moyennes mobiles étant des opérateurs linéaires, elles sont traditionnellement sensibles aux points atypiques notamment

Les coefficients saisonniers sont ensuite normalisés de telle sorte que leur somme sur toute période de 12 mois soit approximativement nulle :

$$S_{norm_t}^{(1)} = S_t^{(1)} - M_{2 \times 12}(S_t^{(1)})$$

4. Estimation de la série corrigée des variations saisonnières :

$$X\_linearisee\_CVS_t^{(1)} = X\_linearisee_t - S_{norm_t}^{(1)} = (T_t + I_t)^{(1)}$$

Cette première estimation de la série corrigée de la variation saisonnière doit, par construction, contenir moins de saisonnalité que la série linéarisée. La méthode X11 remet en œuvre ces 4 premières étapes en changeant les moyennes mobiles pour tenir compte de cette propriété.

5. Estimation de la Tendence-Cycle par moyenne mobile de Henderson sur 13 termes :

$$T_t^{(2)} = H_{13}(X\_linearisee\_CVS_t^{(1)})$$

6. Estimation de la composante Saisonnier-Irrégulier :

$$(S_t + I_t)^{(2)} = X_t - T_t^{(2)}$$

7. Estimation de la composante saisonnière par moyenne mobile 3x5 sur chaque mois :

$$S_t^{(2)} = M_{3 \times 5}[(S_t + I_t)^{(2)}]$$

Les coefficients saisonniers sont ensuite normalisés de telle sorte que leur somme sur toute période de 12 mois soit approximativement nulle :

$$S_{norm_t}^{(2)} = S_t^{(2)} - M_{2 \times 12}(S_t^{(2)})$$

8. Estimation de la série corrigée des variation saisonnières :

$$X\_linearisee\_CVS_t^{(2)} = X\_linearisee_t - S_{norm_t}^{(2)} = (T_t + I_t)^{(2)}$$

La série CVS finale est enfin obtenue en ajoutant les effets des différents points atypiques à la série linéarisée CVS en affectant les AO et les TC à l'irrégulier et les LS à la tendance. La série CVS finale s'écrit ainsi sous la forme suivante ;

$$X_t^{CVS} = T_t^{final} + I_t^{final}$$

avec :

$$T_t^{final} = T_t + \sum_i \hat{\xi}_i LS_{i,t}$$

$$I_t^{final} = I_t + \sum_j \hat{\xi}_j AO_{j,t} + \sum_k \hat{\xi}_k TC_{k,t}$$

## Annexe 4 : La méthode de construction des cinq jeux de régresseurs « CJO »

### Définition du calendrier français et comptage des jours ouvrables et des jours fériés

On commence tout d'abord par générer le calendrier Français pour une longue période, ici pour la période allant du 01/01/1990 au 01/12/4819 (4819 = 2020+2799). Pour cela, on construit une table, notée Frenchcalendar, dont le nombre de lignes correspond à l'ensemble des mois existant entre ces deux dates. Puis, pour chaque ligne de la table Frenchcalendar, i.e. pour chaque mois, on calcule le nombre de dimanches, le nombre de lundis,..., le nombre de samedis. On note dans la suite :

- **Day1(m,A)** : le nombre de dimanches correspondant au mois m de l'année A
- **Day2(m,A)** : le nombre de lundis correspondant au mois m de l'année A
- **Day3(m,A)** : le nombre de mardis correspondant au mois m de l'année A
- **Day4(m,A)** : le nombre de mercredis correspondant au mois m de l'année A
- **Day5(m,A)** : le nombre de jeudis correspondant au mois m de l'année A
- **Day6(m,A)** : le nombre de vendredis correspondant au mois m de l'année A
- **Day7(m,A)** : le nombre de samedis correspondant au mois m de l'année A

Puis, pour chaque ligne de la table Frenchcalendar, i.e. pour chaque mois m d'une année A, on définit la date correspondant au dimanche de Pâques pour l'année A, notée **EasterG**. On en déduit alors le jeudi de l'Ascension, notée **jascen**, et le lundi de Pentecôte, noté **lpent**, de l'année A. En effet, ces 2 jours sont définis de la manière suivante :

$$\text{Jascen} = \text{EasterG} + 39$$

Et

$$\text{Lpent} = \text{EasterG} + 50$$

Puis, on repère :

- le jour férié correspondant au lundi de Pâques en incrémentant de 1 le nombre de lundis fériés, noté **Off2**, pour le mois durant lequel le lundi de Pâques apparaît.
- Le jeudi de l'Ascension, en incrémentant de 1 le nombre de jeudis fériés, noté **Off5**, pour le mois durant lequel le jeudi de l'Ascension apparaît. Cependant, si le jeudi de l'Ascension tombe le 01/05 ou le 08/05, on ne fait pas cette incrémentation.
- Le lundi de Pentecôte, en incrémentant de 1 le nombre de lundis fériés (**Off2**) pour le mois durant lequel le lundi de Pentecôte apparaît. Cependant, pour l'année 2005, on ne fait qu'une incrémentation de 0,5<sup>20</sup>.
- Le 1<sup>er</sup> janvier ; Si on note  $i$  ( $i \in \{1, \dots, 7\}$ ) le jour de la semaine correspondant au 1<sup>er</sup> janvier de l'année A, alors on incrémente de 1 le nombre de jours fériés correspondant (**Offi**) pour le mois de janvier de l'année A.
- La fête du travail, i.e. le 1<sup>er</sup> mai ; Si on note  $i$  ( $i \in \{1, \dots, 7\}$ ) le jour de la semaine correspondant au 1<sup>er</sup> mai de l'année A, alors on incrémente de 1 le nombre de jours fériés correspondant (**Offi**) pour le mois de mai de l'année A.
- L'armistice de 1945, i.e. le 8 mai ; Si on note  $i$  ( $i \in \{1, \dots, 7\}$ ) le jour de la semaine correspondant au 8 mai de l'année A, alors on incrémente de 1 le nombre de jours fériés correspondant (**Offi**) pour le mois de mai de l'année A.
- Le 14 juillet ; Si on note  $i$  ( $i \in \{1, \dots, 7\}$ ) le jour de la semaine correspondant au 14 juillet de l'année A, alors on incrémente de 1 le nombre de jours fériés correspondant (**Offi**) pour le mois de juillet de l'année A.
- L'Assomption, i.e. le 15 août ; Si on note  $i$  ( $i \in \{1, \dots, 7\}$ ) le jour de la semaine correspondant au 15 août de l'année A, alors on incrémente de 1 le nombre de jours fériés correspondant (**Offi**) pour le mois d'août de l'année A.

<sup>20</sup> En raison de l'apparition de la journée de solidarité envers les personnes âgées initiée par la loi du 30 juin 2004

- La Toussaint, i.e. le 1<sup>er</sup> novembre ; Si on note  $i$  ( $i \in \{1, \dots, 7\}$ ) le jour de la semaine correspondant au 1<sup>er</sup> novembre de l'année A, alors on incrémente de 1 le nombre de jours fériés correspondant (**Offi**) pour le mois de novembre de l'année A.
- L'armistice de 1918, i.e. le 11 novembre ; Si on note  $i$  ( $i \in \{1, \dots, 7\}$ ) le jour de la semaine correspondant au 11 novembre de l'année A, alors on incrémente de 1 le nombre de jours fériés correspondant (**Offi**) pour le mois de novembre de l'année A.
- Noël, i.e. le 25 décembre ; Si on note  $i$  ( $i \in \{1, \dots, 7\}$ ) le jour de la semaine correspondant au 25 décembre de l'année A, alors on incrémente de 1 le nombre de jours fériés correspondant (**Offi**) pour le mois de décembre de l'année A.

## Construction des jeux de régresseurs « CJO »

On définit, tout d'abord, le régresseur « année bissextile », noté **leapyear**, de la manière suivante :

$$\text{leapyear}(m, A) = \begin{cases} 0 & \text{si le mois considéré n'est pas le mois de février} \\ 0,75 & \text{si le mois considéré est le mois de février et l'année considérée est bissextile} \\ -0,25 & \text{si le mois considéré est le mois de février et l'année considérée n'est pas bissextile} \end{cases}$$

Pour rappel, depuis l'ajustement du calendrier grégorien, on considère qu'une année est bissextile si une des 2 conditions suivantes est vérifiée :

- elle est divisible par 4 mais non divisibles par 100
- ou
- elle est divisible par 400

Afin d'avoir des régresseurs corrigés des variations saisonnières (« CVS »), on calcule pour chacun des 12 mois de l'année ( $m \in \{1, \dots, 12\}$ ) la moyenne sur l'ensemble des années comprises entre 1990 et 4819 des variables suivantes :

- $\text{Day1}(m, A), \dots, \text{Day7}(m, A)$
- Et
- $\text{Off1}(m, A), \dots, \text{Off7}(m, A)$  (Offi est le nombre de jours fériés correspondant au mois m de l'année A)

Ces moyennes, appelées **moyennes de long terme**, sont notées :

- $\text{mean}(\text{Day1}, m), \dots, \text{mean}(\text{Day7}, m)$
- et
- $\text{mean}(\text{Off1}, m), \dots, \text{mean}(\text{Off7}, m)$

A partir de ces moyennes, on redéfinit des variables « centrées » pour chacune des variables ci-dessus de la manière suivante :

$$\begin{aligned} \forall i = 1, \dots, 7, \\ \text{Dayi\_new}(m, A) &= \text{Dayi}(m, A) - \text{mean}(\text{Dayi}, m) \\ \text{Offi\_new}(m, A) &= \text{Offi}(m, A) - \text{mean}(\text{Offi}, m) \end{aligned}$$

On obtient alors pour chaque jour de la semaine, noté  $i$  ( $i \in \{1, \dots, 7\}$ ), le nombre de jours non fériés ( $y$  compris les ponts) « centré », noté  $\text{Ini\_new}$ , et défini de la manière suivante :

$$\forall i \in \{1, \dots, 7\}, \quad \text{Ini\_new}(m, A) = \text{Dayi\_new}(m, A) - \text{Offi\_new}(m, A)$$

Chaque jeu de régresseurs « CJO » repose sur des hypothèses portant sur les effets des différents types de jours. Elles sont récapitulées dans le tableau suivant :

**Tableau 1 : Définition des 3 jeux de régresseurs « CJO » regcjo\_1, regcjo\_2 et regcjo\_7**

Effet	Variable associée	Regcjo_1	Regcjo_2	Regcjo_7
Effet « dimanche non férié »	Group(In1_new)	0	0	0
Effet « lundi non férié »	Group(In2_new)	1	1	1
Effet « mardi non férié »	Group(In3_new)	2	1	2
Effet « mercredi non férié »	Group(In4_new)	3	1	3
Effet « jeudi non férié »	Group(In5_new)	4	1	4
Effet « vendredi non férié »	Group(In6_new)	5	1	5
Effet « samedi non férié »	Group(In7_new)	6	0	0
Effet « dimanche férié »	Group(Off1_new)	0	0	0
Effet « lundi férié »	Group(Off2_new)	0	0	0
Effet « mardi férié »	Group(Off3_new)	0	0	0
Effet « mercredi férié »	Group(Off4_new)	0	0	0
Effet « jeudi férié »	Group(Off5_new)	0	0	0
Effet « vendredi férié »	Group(Off6_new)	0	0	0
Effet « samedi férié »	Group(Off7_new)	0	0	0

On en déduit alors les régresseurs « CJO » corrigés des variations saisonnières correspondant à chacun des 3 jeux de régresseurs « CJO » regcjo\_1, regcjo\_2 et regcjo\_7 :

$$\forall k > 0,$$

$$\text{REGk\_contr}(m, A) = \text{REGk}(m, A) - \left( \frac{\sum_{i \in \{\text{In1\_new}, \dots, \text{In7\_new}, \text{Off1\_new}, \dots, \text{Off7\_new}\}} \text{Ind}_k(i)}{\sum_{i \in \{\text{In1\_new}, \dots, \text{In7\_new}, \text{Off1\_new}, \dots, \text{Off7\_new}\}} \text{Ind}_0(i)} \right) \times \text{REG0}(m, A)$$

où :

$$\forall k \geq 0, \quad \text{REGk}(m, A) = \sum_{i \in \{\text{In1\_new}, \dots, \text{In7\_new}, \text{Off1\_new}, \dots, \text{Off7\_new}\}} i(m, A) \times \text{Ind}_k(i)$$

avec :

$$\forall i \in \{\text{In1\_new}, \dots, \text{In7\_new}, \text{Off1\_new}, \dots, \text{Off7\_new}\},$$

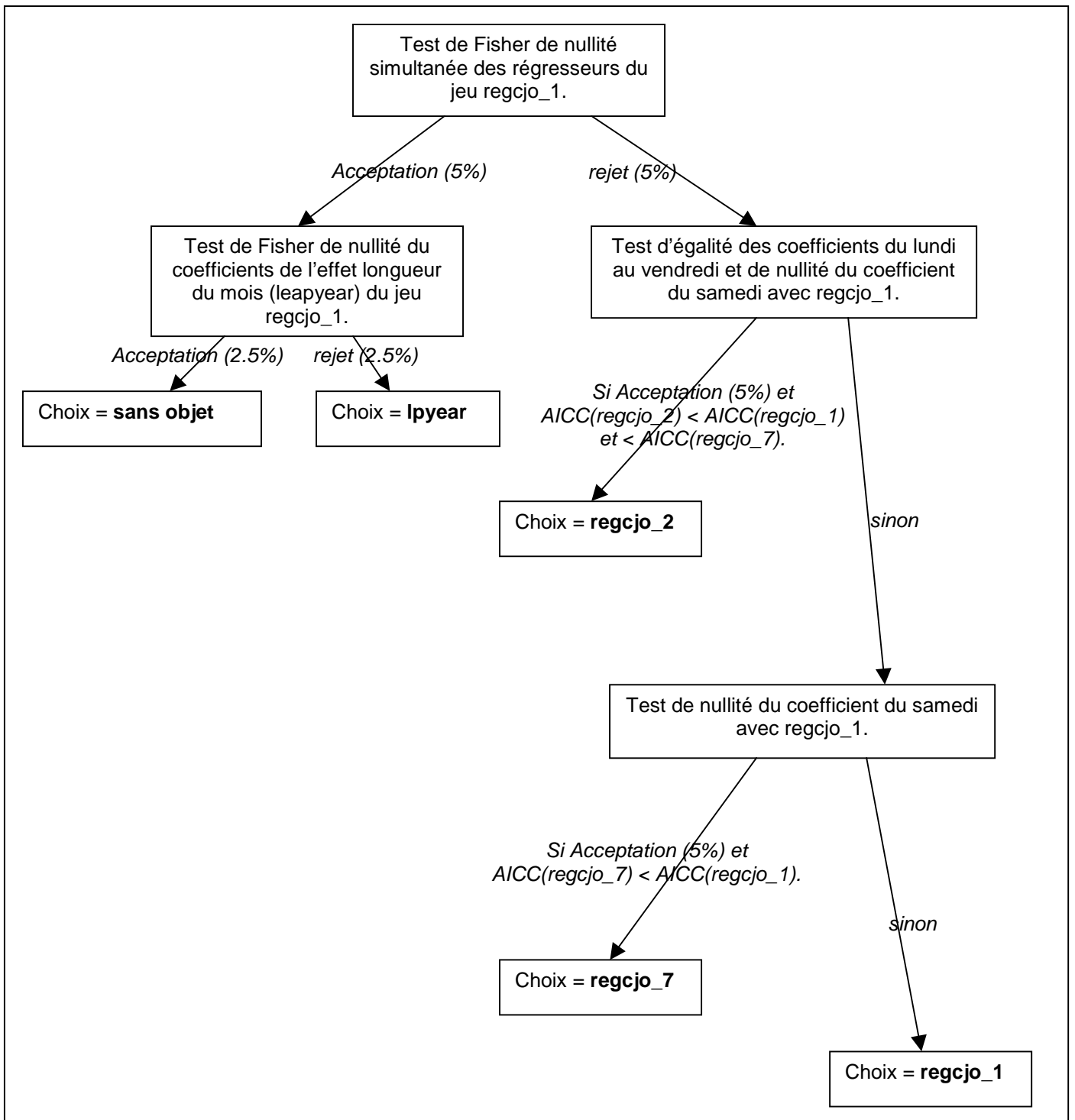
$$\text{Ind}_k(i) = \begin{cases} 1 & \text{si Group}(i) = k \\ 0 & \text{sin on} \end{cases}$$

Les 4 jeux de régresseurs « CJO » sont ainsi composés des variables suivantes :

**Tableau 2 : Les régresseurs composant les 4 jeux de régresseurs « CJO »**

<b>Leap_year</b>	<b>Regcjo_1</b>	<b>Regcjo_2</b>	<b>Regcjo_7</b>
Leapyear	Leapyear	Leapyear	Leapyear
	REG1_contr	REG1_contr	REG1_contr
	REG2_contr		REG2_contr
	REG3_contr		REG3_contr
	REG4_contr		REG4_contr
	REG5_contr		REG5_contr
	REG6_contr		

## Annexe 5: L'algorithme de sélection du meilleur jeu de régresseurs « CJO »





## Annexe 6 : Les notes du bilan qualité utilisées

Dans le cadre de cette étude, une désaisonnalisation va être jugée de bonne qualité si :

- la série CVS-CJO ne présente pas d'effets saisonniers résiduels
- la série CVS-CJO ne présente pas d'effets « jours ouvrables » résiduel
- la qualité de la décomposition peut être considérée comme bonne

A titre informatif, nous étudierons également la présence d'effets saisonniers résiduels et d'effets « jours ouvrables » résiduels sur l'irrégulier.

### La présence d'effets saisonniers résiduels

La présence d'effets saisonniers résiduels est testée sur la série CVS-CJO « entière » et sur la série CVS-CJO restreinte aux 36 derniers mois. Dans le cadre de cette étude, on considère que la série CVS-CJO « entière » est non saisonnière si et seulement si les 3 conditions suivantes sont vérifiées :

- le test combiné de « présence d'une saisonnalité identifiable » statue sur la présence d'une saisonnalité non identifiable (modalité « Not Present »)
- et
- l'hypothèse H0 de non présence de saisonnalité (égalité des différents mois) associée au test paramétrique de « saisonnalité stable » n'est pas rejetée au niveau 5% (i.e. p-value>5%)
- et
- l'hypothèse H0 de non présence de saisonnalité (égalité des différents mois) associée au test non paramétrique de Kruskal-Wallis de « saisonnalité stable » n'est pas rejetée au niveau 5%

Les résultats des différents tests sont obtenues en appliquant la PROC X12 de SAS à la série CVS-CJO<sup>21</sup>.

Quant à la série CVS-CJO restreinte aux 36 derniers mois, on considère qu'elle est non saisonnière si et seulement si l'hypothèse H0 de non présence de saisonnalité associée au test paramétrique de « saisonnalité stable » n'est pas rejetée au niveau 10%.

On définit alors la note suivante :

$$\text{note\_sais\_residuel\_sur\_cvscjo} = \begin{cases} 20 & \text{si saison\_entier} = \text{"non"} \text{ et saison\_fin\_periode} = \text{"non"} \\ 15 & \text{si saison\_entier} = \text{"oui"} \text{ et saison\_fin\_periode} = \text{"non"} \\ 5 & \text{si saison\_entier} = \text{"non"} \text{ et saison\_fin\_periode} = \text{"oui"} \\ 0 & \text{si saison\_entier} = \text{"oui"} \text{ et saison\_fin\_periode} = \text{"oui"} \end{cases}$$

### La présence d'effets « jours ouvrables » résiduel

Pour tester la présence d'effets « jours ouvrables » résiduel sur la série CVS-CJO, on modélise la série CVS-CJO par un modèle Reg-ARIMA en introduisant le jeu de régresseurs CJO « regcjo\_1 », puis on teste la nullité conjointe des coefficients associés à ce jeu de régresseurs. Cette modélisation est réalisée en utilisant la PROC X12 de SAS.

Si on note  $\hat{\beta}$  le vecteur des estimateurs des coefficients associés au jeu de régresseurs « regcjo\_1 », l'hypothèse nulle du test s'écrit :

<sup>21</sup> cf. annexe 3 pour une description détaillée de la mise en œuvre de ces tests

$$H_0 : \hat{\beta} = 0$$

On définit une note associée à ce test de la manière suivante :

$$\text{note\_jo\_residuel\_sur\_cvscjo} = \begin{cases} 20 & \text{si } p\_value\_jo\_residuel > 5\% \\ 0 & \text{sin on} \end{cases}$$

### La qualité de la décomposition

Dans le cadre de cette étude, on jugera la qualité de la décomposition réalisée par X11 à la lumière des valeurs prises par les statistiques Q et Q-M2 : si les valeurs prises par ces dernières sont toutes les deux strictement inférieures à 1, on considérera que la décomposition est de bonne qualité ; dans le cas contraire, la décomposition sera a priori de mauvaise qualité.

On définit ainsi la note suivante :

$$\text{note\_qualite\_decomposition} = \begin{cases} 0 & \text{si } Q \geq 1 \text{ ou } Q - M2 \geq 1 \\ 20 & \text{sin on} \end{cases}$$

## Bibliographie

[1] Fabre V., « Choix d'une méthode de désaisonnalisation : application aux créations d'entreprise », X<sup>èmes</sup> Journées de Méthodologie Statistique de l'Insee (2009)

[2] Attal-Toubert K., « Régresseurs pour effets de calendrier : comment les construire, comment les choisir ? » XI<sup>èmes</sup> Journées de Méthodologie Statistique de l'Insee (2012)

[3] Macaulay, F.R. (1931), « The Smoothing of Time Series », *National Bureau of Economic Research*

[4] Himpens S., « Comment juger de la qualité d'une désaisonnalisation ? », XI<sup>èmes</sup> Journées de Méthodologie Statistique de l'Insee (2012)