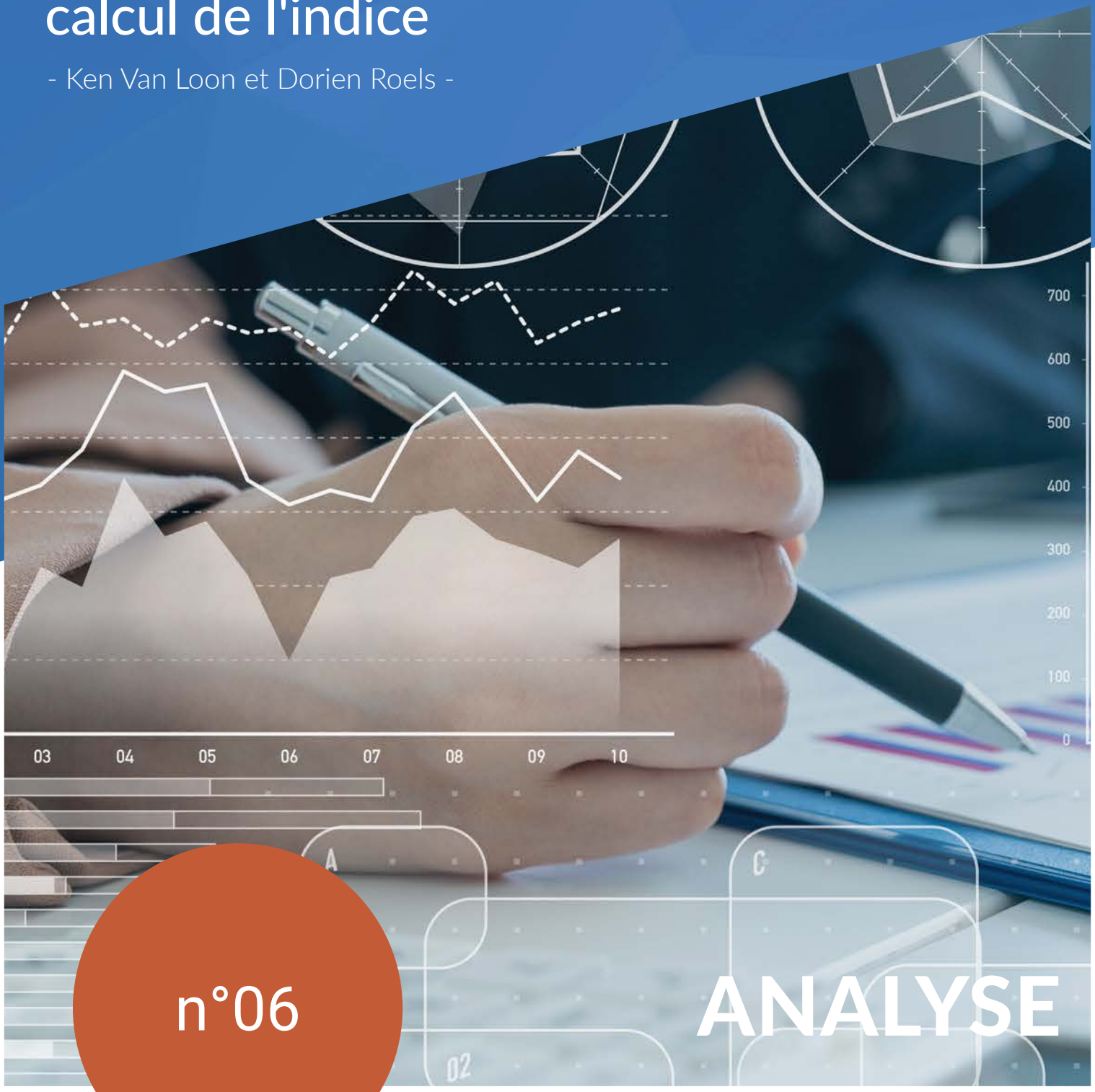
STATBEL
La Belgique en chiffres

Évaluation des méthodes multilatérales de calcul de l'indice

- Ken Van Loon et Dorien Roels -



n°06

ANALYSE

07.2019

economie
SPF Economie, P.M.E., Classes moyennes et Energie

.be

Évaluation des méthodes multilatérales de calcul de l'indice

Ken Van Loon, Dorien Roels[1]

[1] Statisticiens à Statbel (Direction générale Statistique - Statistics Belgium)

ABSTRACT

Depuis janvier 2015, Statbel, l'office belge de statistique, utilise des scanner data de supermarchés pour calculer l'indice des prix à la consommation. L'un des nombreux avantages de l'utilisation des scanner data réside dans le fait que les informations sur le chiffre d'affaires sont disponibles au niveau du produit. Lorsque les informations sur le chiffre d'affaires sont directement utilisées dans des indices en chaîne, caractérisés par une fréquence élevée (par exemple un indice en chaîne mensuel), il existe un risque de chain drift dans la série d'indices. Un chain drift signifie que l'indice ne revient pas à l'unité (ou à un niveau d'indice de 100) quand les prix et les informations sur le chiffre d'affaires de la période considérée reviennent au niveau de la période de référence. Dans ce cas, il n'y a pas eu de changement de prix à proprement parler et aucun changement d'indice ne peut donc être enregistré. Le chain drift survient parce que le prix et le chiffre d'affaires ne sont pas définis de manière unique dans un contexte intertemporel. Pendant les promotions, les produits sont achetés en grande quantité, ce qui entraîne la constitution de stocks. C'est pourquoi les quantités achetées au cours des mois qui suivent ne reviennent pas immédiatement au niveau initial, alors que le prix bien. L'effet haussier sur les prix a moins de poids que l'effet baissier des promotions, avec pour conséquence une dérive négative des indices en chaîne.

En raison de cette dérive, nous travaillons actuellement au niveau le plus bas avec des indices non pondérés pour les calculs où des scanner data sont utilisées. Les informations sur le chiffre d'affaires ne sont pas directement utilisées dans le calcul, bien qu'elles soient disponibles (elles sont toutefois utilisées pour l'échantillonnage, par exemple). Pour éviter ce problème de dérive, des méthodes multilatérales ont été développées depuis quelques années. Ces méthodes ne comparent pas deux périodes comme c'était traditionnellement le cas (par exemple la période actuelle et une période de base fixe ou la période actuelle et la période précédente), mais toutes les périodes. Cet article donne un aperçu théorique et mathématique des différentes méthodes multilatérales. Étant donné que les méthodes multilatérales sont par définition susceptibles de faire l'objet de révisions lorsque la série est étendue, il convient de travailler avec une période restreinte. Les périodes peuvent ensuite être couplées de différentes manières. Celles-ci sont expliquées visuellement et mathématiquement. Toutes ces méthodes seront ensuite testées sur des données réelles.

SOMMAIRE

Évaluation des méthodes multilatérales de calcul de l'indice *1*

Abstract *2*

Sommaire *3*

Introduction *4*

1. Méthode de calcul actuelle – méthode dynamique *5*

2. Chain drift *6*

3. Méthodes multilatérales *9*

3.1. GEKS 10

3.2. Geary-Khamis 11

3.3. Lehr (augmenté) 12

3.4. Time Product Dummy 13

3.5. Régression hédonique 14

4. Rolling year et splicing *14*

4.1. Movement splice 15

4.2. Window splice 17

4.3. Half splice 19

4.4. Mean splice 20

4.5. Fixed base monthly expanding window 22

4.6. Fixed base moving window 23

5. Recherche empirique *24*

5.1. Résultats pour une période entière 24

5.2. Résultats des méthodes de splicing et d'extension 26

5.3. Impact du dumping 27

5.4. Impact des relances de produits 28

6. Conclusion *28*

Bibliographie *30*

INTRODUCTION

Depuis janvier 2015, Statbel, l'office belge de statistique, utilise des scanner data de supermarchés pour calculer l'indice des prix à la consommation. L'un des nombreux avantages de l'utilisation des scanner data réside dans le fait que les informations sur le chiffre d'affaires sont disponibles au niveau du produit. Cependant, lorsque ces informations sont directement utilisées dans un indice en chaîne caractérisé par une fréquence de couplage élevée (par exemple tous les mois), il existe un risque de dérive (chain drift). En raison de cette dérive, l'indice ne revient pas à l'unité ou à 100 quand les prix et le chiffre d'affaires de la période actuelle sont identiques à ceux de la période de référence. Les indices obtenus ne seraient alors ni transitifs ni circulaires, ce qui signifie que l'indice en chaîne ne serait pas égal à l'indice direct.

Afin d'éviter le chain drift, nous travaillons actuellement avec des indices non pondérés au niveau le plus bas. Les informations sur le chiffre d'affaires sont bien entendu utilisées à d'autres fins, par exemple pour déterminer les produits repris dans l'échantillon. Entre-temps, des méthodes ont été proposées afin d'utiliser les informations sur le chiffre d'affaires au niveau du produit sans entraîner de chain drift. Ces méthodes sont également appelées "méthodes multilatérales", parce que plusieurs périodes sont comparées, contrairement aux calculs plus traditionnels de l'indice où deux périodes étaient comparées. L'une des conséquences du calcul de l'indice à l'aide de méthodes multilatérales est qu'à chaque fois que la série chronologique est étendue, les indices du passé sont révisés. Cela s'explique par le fait que les périodes plus anciennes sont également comparées à l'ajout le plus récent. Afin d'éliminer l'impact de la révision, nous travaillons avec une période limitée. Ces périodes limitées peuvent ensuite être couplées de différentes manières.

Tous ces points seront expliqués dans cet article, et les différentes méthodes seront testées sur des scanner data réelles. Nous commençons par une brève description de la méthode de calcul actuelle. Nous expliquerons ensuite ce qu'est exactement le chain drift et pourquoi il se produit. Nous examinerons alors de manière plus approfondie un certain nombre de méthodes multilatérales et la manière dont elles résolvent le problème du chain drift. Viendra ensuite une explication de l'utilisation des périodes limitées et des différentes manières de les coupler. Enfin, les différentes méthodes seront testées de manière empirique et les résultats seront présentés.

1. METHODE DE CALCUL ACTUELLE – METHODE DYNAMIQUE

Les calculs à l'aide de scanner data se font à un niveau inférieur aux groupes de produits publiés dans l'indice des prix à la consommation, où des groupes de produits détaillés spécifiques au détaillant ont été créés. Ce niveau est appelé le niveau élémentaire, et les calculs à ce niveau sont effectués pour un échantillon déterminé. Cet échantillon se compose de produits disponibles pendant deux périodes consécutives et dont la part dans les dépenses est supérieure au seuil dynamique déterminé. Ce seuil est défini comme:

$$\frac{s_m + s_{m-1}}{2} > \frac{1}{n * \lambda} \qquad \textit{Équ. 1}$$

Où :

- s_m est la part de marché de chaque produit correspondant au cours du mois m
- s_{m-1} est la part de marché de chaque produit correspondant au cours du mois $m-1$
- n est le nombre de produits
- λ = 1.25

Étant donné que le seuil est dynamique par l'inclusion du paramètre n, cette méthode de travail est également appelée méthode dynamique. La détermination de ces indices au niveau le plus bas (indices élémentaires) se fait par le biais d'une moyenne géométrique ou indice de Jevons P_J. La formule est la suivante:

$$P_J = \prod_{i \epsilon G_{0,1}} \left(\frac{p_i^1}{p_i^0}\right)^{1/N_{0,1}} \qquad \textit{Équ. 2}$$

où $G_{0,1}$ est l'échantillon. Il s'agit de l'ensemble de tous les produits qui correspondent durant deux périodes consécutives (période 0 et période 1) et qui se trouvent au-dessus du seuil, et où $N_{0,1}$ est le nombre total de produits dans l'échantillon. Dans un indice en chaîne de plusieurs périodes, ces indices mensuels sont couplés en multipliant les indices obtenus entre eux. Pour un indice de Jevons en chaîne $P_{J,\text{ketting}}$ comprenant trois périodes, la formule est la suivante:

$$P_{J,\text{ketting}}^{0,2} = \prod_{i \epsilon G_{0,1}} \left(\frac{p_i^1}{p_i^0}\right)^{1/N_{0,1}} \prod_{i \epsilon G_{1,2}} \left(\frac{p_i^2}{p_i^1}\right)^{1/N_{1,2}} \qquad \textit{Équ. 3}$$

La formule générale pour un $P_{J,\text{ketting}}^{0,T}$ où T est une série chronologique (=0, 1, ..., T) est la suivante:

$$P_{J,\text{ketting}}^{0,T} = \prod_{t=1}^{T} \left(\prod_{i \epsilon G_{t-1,t}} \left(\frac{p_i^t}{p_i^{t-1}}\right)^{1/N_{t-1,t}} \right) \qquad \textit{Équ. 4}$$

Ces indices pour un groupe de produits k, exprimés dans l'année en cours avec comme période de base décembre de l'année précédente (=100), peuvent s'écrire t pour la période $P_{J,k,dec=100}^{0,t}$. Cela tient au fait que les coefficients de pondération sont

toujours adaptés en décembre. Les indices obtenus sont ensuite agrégés avec les poids correspondants calculés à partir du chiffre d'affaires tiré des scanner data du groupe de produits concerné de l'année précédente.

Il convient de noter que lorsqu'un produit disparaît de l'échantillon, le prix du produit est ensuite imputé ($\hat{p}_i$) sur la base de l'évolution du prix du groupe de produits correspondant. Lors du retour dans l'échantillon, le prix de la période précédente (p_i^{t-1}) est dans ce cas un prix imputé ($\hat{p}_i^{t-1}$). L'objectif est de tenir compte de l'évolution des prix "manquante" lorsqu'un produit sort de l'échantillon pour ensuite y revenir.

2. CHAIN DRIFT

Un indice non pondéré est donc utilisé au niveau élémentaire. Cela peut sembler étrange de prime abord, étant donné que les scanner data contiennent des informations relatives au chiffre d'affaires jusqu'au niveau des produits individuels. Grâce à ces informations, il serait possible de passer à une formule d'indice superlatif. Dans une formule d'indice superlatif, les prix et les quantités, tant de la période de référence que de la période considérée, sont traités de manière symétrique.

Pourquoi une formule d'indice superlatif n'est-elle pas utilisée et pourquoi les informations relatives au chiffre d'affaires ne sont-elles pas directement utilisées dans le calcul de l'indice, mais seulement de manière indirecte pour composer l'échantillon mensuel sur la base du seuil dynamique? La raison en est que l'utilisation de ces informations relatives au chiffre d'affaires dans le calcul de l'indice entraîne un chain drift. Un chain drift signifie que l'indice en chaîne ne revient pas à l'unité (ou à un niveau d'indice de 100) quand les prix de la période considérée reviennent au niveau de la période de référence. Le concept du chain drift est expliqué à l'aide d'un indice de Törnqvist, P_T, dont la formule est égale à:

$$P_T^{0,t} = \prod_{i=1}^{n} \left(\frac{p_i^t}{p_i^0}\right)^{0.5\left(\frac{p_i^0 q_i^0}{\sum_{j=1}^{n} p_j^0 q_j^0} + \frac{p_i^t q_i^t}{\sum_{j=1}^{n} p_j^t q_j^t}\right)}$$

$$= \prod_{i=1}^{n} \left(\frac{p_i^t}{p_i^0}\right)^{0.5\,(s_i^0 + s_i^t)} \qquad \textit{Équ. 5}$$

L'équation 5 exprime l'indice de Törnqvist pour la période de 0 à t et est égale à la moyenne géométrique des ratios de prix (période actuelle par rapport à la base), pondérée à partir de la moyenne arithmétique des parts de dépenses au cours de la période de base et de la période en cours. Les informations sur le chiffre d'affaires sont donc utilisées directement. Pour un indice de Törnqvist en chaîne de deux périodes, la formule est la suivante:

$$P_{T,\text{ketting}}^{0,2} = \prod_{i=1}^{n} \left(\frac{p_i^1}{p_i^0}\right)^{0.5\,(s_i^0 + s_i^1)} \prod_{i=1}^{n} \left(\frac{p_i^2}{p_i^1}\right)^{0.5\,(s_i^1 + s_i^2)} \qquad \textit{Équ. 6}$$

Supposons maintenant que seul 1 produit *a* est suivi (*i* = *a*). L'équation peut alors être ramenée à la forme suivante:

$$P_{T,\text{ketting}}^{0,2} = \left(\frac{p_a^1}{p_a^0}\right)^{0.5\,(s_a^0 + s_a^1)} \left(\frac{p_a^2}{p_a^1}\right)^{0.5\,(s_a^1 + s_a^2)} \qquad \textit{Équ. 7}$$

Si ce produit *a* a fait l'objet d'une promotion temporaire lors de la période 1 ($p_a^0 > p_a^1$), mais revient à son niveau de prix de la période 0 lors de la période 2 ($p_a^0 = p_a^2$), alors l'équation 7 peut être réécrite comme suit:

$$\mathrm{P}_{\mathrm{T,ketting}}^{0,2} = \left(\frac{p_a^1}{p_a^0}\right)^{0.5\,(s_a^0 + s_a^1)} \left(\frac{p_a^0}{p_a^1}\right)^{0.5\,(s_a^1 + s_a^2)}$$

$$= \left(\frac{p_a^1}{p_a^0}\right)^{0.5\,(s_a^0 + s_a^1) - 0.5\,(s_a^1 + s_a^2)}$$

$$= \left(\frac{p_a^1}{p_a^0}\right)^{0.5\,(s_a^0 - s_a^2)} \qquad \textit{Équ. 8}$$

Étant donné qu'en théorie, l'indice doit revenir à l'unité parce que le niveau de prix de la période 2 est égal à celui de la période 0, cela ne peut se produire que lorsque $s_a^0 = s_a^2$, comme le montre l'équation 8. Cela signifie donc que les dépenses de la période 2 et de la période 0 doivent être égales, tout comme les niveaux de prix. La théorie microéconomique classique part du principe que les quantités sont déterminées de manière unique pour un ensemble de prix donné. Si c'était le cas, les poids de la période 2 et de la période 0 seraient identiques.

Si cette prévision théorique ne se concrétise pas, alors, $P_{T,ketting}^{02} < 1$ lorsque $s_a^0 > s_a^2$ et $P_{T,ketting}^{02} > 1$ si $s_a^0 < s_a^2$. Un exemple fictif simple dans lequel l'indice de Törnqvist en chaîne ne revient pas à l'unité est présenté dans le tableau ci-dessous pour les produits A et B sur une période de quatre mois. Le prix et le chiffre d'affaires du produit B restent stables durant toute la période. Par contre, ceux du produit A fluctuent.

Tableau 1: Exemple fictif de chain drift pour un indice de Törnqvist en chaîne

	Janvier		Février		Mars		Avril	
	Prix	Chiffre d'affaires	Prix	Chiffre d'affaires	Prix	Chiffre d'affaires	Prix	Chiffre d'affaires
Produit A	2,50	10.000	2,00	500.000	2,50	2.000	2,50	10.000
Produit B	3,00	10.000	3,00	10.000	3,00	10.000	3,00	10.000
Indice de Törnqvist en chaîne	100		84,78		96,35		96,35	

Dans cet exemple, le produit A est en promotion en février. Par conséquent, son chiffre d'affaires augmente fortement au cours de ce mois, entraînant ainsi un effet baissier sur l'indice (une diminution de 100 à 84,78). En mars, les prix ont retrouvé leur niveau de janvier. Cependant, le chiffre d'affaires du produit A n'égale pas celui de janvier (par exemple, parce que les consommateurs ont constitué des stocks pendant la promotion de février). L'indice ne retrouve dès lors pas le niveau escompté de 100. De plus, il subsiste un effet permanent comme le montre l'indice d'avril. Afin de démontrer cet effet permanent, en avril, les données de chiffre d'affaires et de prix sont identiques à celles de janvier, ce qui, sans aucun doute, devrait déboucher sur un indice de 100. Ce n'est toutefois pas le cas. La valeur de l'indice reste de 96,35 étant donné que le rapport de prix ne change pas. Ce phénomène est appelé chain drift. Si l'on travaillait avec un indice direct, où avril serait directement comparé à janvier, l'indice reviendrait également à 100. Cependant, en raison de la dynamique des scanner data (de nombreux produits qui apparaissent sur le marché ou en disparaissent), il est nécessaire de travailler avec un indice en chaîne.

Le graphique ci-dessous permet de visualiser le chain drift de l'indice de Törnqvist en chaîne. Les chiffres de l'IPC mentionnés sont les indices officiels du produit X:

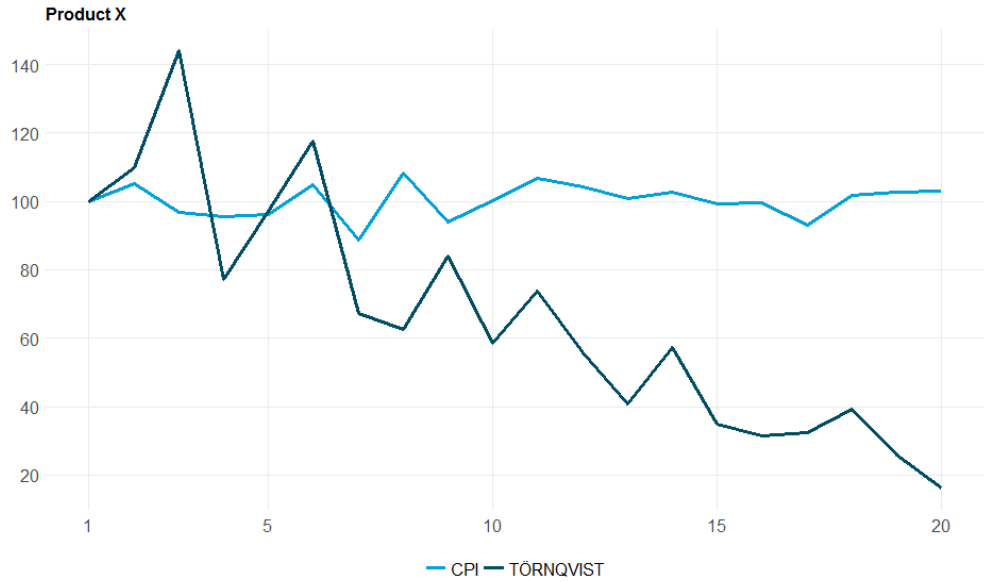


Figure 1 : Chain drift de l'indice de Törnqvist en chaîne

Il apparaît clairement que l'indice de Törnqvist présente une dérive négative. Le chain drift est également la conséquence de la non-transitivité d'un indice de Törnqvist en chaîne (ou de tout indice en chaîne superlatif). La non-transitivité signifie que l'évolution des prix mesurée entre deux périodes dépend de la période de base choisie, avec pour résultat le fait que l'indice en chaîne n'est pas égal à l'indice direct correspondant. Pour rappel, dans l'exemple, en cas de comparaison directe, l'indice est égal à 100. La raison de la non-transitivité peut être facilement démontrée. L'indice de Törnqvist direct pour un indice qui va de la période 1 à T est:

$$P_{T,\text{direct}}^{1,T} = \prod_{i=1}^{n} \left(\frac{p_i^T}{p_i^1}\right)^{0.5\,(s_i^1 + s_i^T)} \qquad \textit{Équ. 9}$$

Si l'indice de Törnqvist est transitif, alors $P_{T,\text{direct}}^{0,T} / P_{T,\text{direct}}^{0,1} = P_{T,\text{direct}}^{1,T}$, ce qui est peu probable parce que:

$$\frac{P_T^{0,T}}{P_T^{0,1}} = \frac{\prod_{i=1}^{n} \left(\frac{p_i^T}{p_i^0}\right)^{0.5\,(s_i^0 + s_i^T)}}{\prod_{i=1}^{n} \left(\frac{p_i^1}{p_i^0}\right)^{0.5\,(s_i^0 + s_i^1)}} = P_T^{1,T} \left(\prod_{i=1}^{n} (p_i^0)^{s_i^1 - s_i^T} \prod_{i=1}^{n} (p_i^1)^{s_i^T - s_i^0} \prod_{i=1}^{n} (p_i^T)^{s_i^0 - s_i^1} \right)^{0.5} \qquad \textit{Équ. 10}$$

La probabilité que le dernier facteur entre parenthèses soit égal à 1 est assez faible. La non-transitivité est donc la conclusion la plus logique et $P_{T,\text{direct}}^{0,T} / P_{T,\text{direct}}^{0,1} \neq P_{T,\text{direct}}^{1,T}$.

La question est désormais de savoir si ce chain drift sera plutôt positif ou négatif. S'agit-il d'un indice qui affiche une tendance à la baisse (vers 0) sur de longues périodes ou d'un indice orienté à la hausse (vers l'infini)?

Le plus logique est que l'indice s'approche de zéro, parce que le composant faisant l'objet d'une baisse de prix reçoit un poids plus élevé en raison de la forte hausse de son chiffre d'affaires, et ce contrairement à la hausse des prix de la période suivante qui entraîne un poids moins important étant donné que les consommateurs achètent moins que ce à quoi on s'attendrait normalement (par exemple en raison de la constitution d'un stock ou d'une substitution). La probabilité d'une dérive positive est en théorie possible, mais survient moins souvent en pratique. Une dérive positive peut se produire lorsqu'une promotion est d'application à la fin de la période de référence et se poursuit pendant une partie de la période considérée. La hausse du prix lors de la période considérée reçoit de cette manière un poids plus important, entraînant une dérive positive.

En raison du chain drift, il a donc été décidé de travailler avec un indice de Jevons non pondéré au niveau le plus bas. L'indice de Jevons est en réalité une variante particulière de l'indice de Törnqvist mentionné ci-dessus, dans laquelle les poids de tous

les produits correspondent dans un agrégat élémentaire. Par conséquent, le poids d'un produit j pendant la période 0 $\left(s_j^0\right)$ est égal au poids durant la période t $\left(s_j^t\right)$. De même, les poids de tous les produits 1 à n sont égaux dans l'agrégat élémentaire. Cela signifie donc également que $s_j^0 = s_j^t = 1/N$. Dans ce cas, la formule de Törnqvist est donc égale à:

$$P_T^{0,t} = \prod_{i=1}^{n} \left(\frac{p_i^t}{p_i^0}\right)^{0.5\left(s_j^0 + s_j^t\right)} = \prod_{i=1}^{n} \left(\frac{p_i^t}{p_i^0}\right)^{\left(s_j^0\right)} = \prod_{i=1}^{n} \left(\frac{p_i^t}{p_i^0}\right)^{\frac{1}{n}} = P_J^{0,t} \qquad \text{Équ. 11}$$

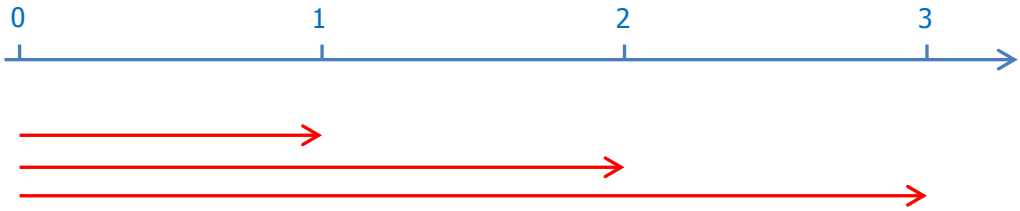
ce qui donne l'indice de Jevons. Il apparaît aussi clairement qu'un indice de Jevons est transitif. Lorsque dans l'équation 10, la valeur identique s_i est complétée, le facteur entre parenthèses est égal à 1. Par conséquent, $P_{J,\text{direct}}^{0,T} / P_{J,\text{direct}}^{0,1} = P_{J,\text{direct}}^{1,T}$.

3. MÉTHODES MULTILATERALES

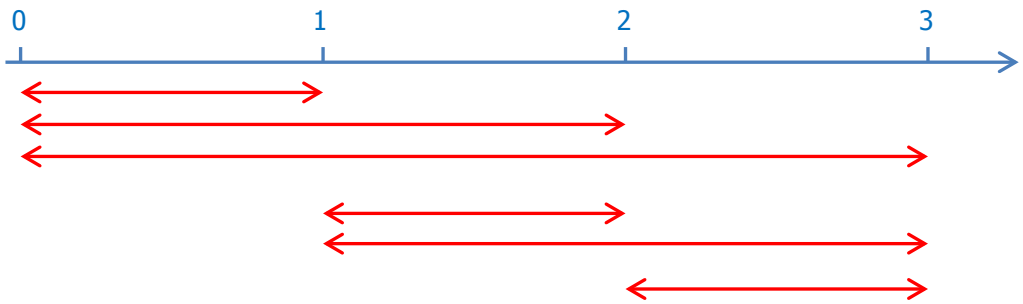
Pour tenter de remédier au problème de chain drift mentionné ci-dessus, différentes méthodes multilatérales sont proposées. Avant de se pencher sur ces méthodes, il convient de s'arrêter sur la différence entre la méthode utilisée actuellement pour les scanner data (une méthode bilatérale) et les méthodes multilatérales.

Dans une méthode bilatérale telle que l'indice de Jevons en chaîne, les prix sont comparés entre deux périodes consécutives pour exprimer l'évolution des prix à long terme. S'il y a quatre périodes ($t = 0, 1, 2, 3$), l'évolution des prix entre les périodes 3 et 0 est calculée en faisant une chaîne de toutes les comparaisons de prix intermédiaires. Cette chaîne est schématisée ci-dessous.

Un indice bilatéral en chaîne est différent d'un indice bilatéral direct. Avec un indice bilatéral direct, la période actuelle est comparée à une période de base fixe. Dans le cas décrit ci-dessus de 4 périodes, chaque période est directement comparée à la période de base fixe, comme montré ci-dessous.



Les méthodes multilatérales diffèrent des méthodes bilatérales en ce sens que plus de 2 périodes (au moins 3) sont comparées directement entre elles. Dans notre exemple, chaque période est comparée aux trois autres périodes, comme indiqué dans la figure suivante.



Différentes méthodes multilatérales seront abordées dans ce chapitre. La première est la méthode GEKS.

3.1. GEKS

Les indices GEKS (nommés d'après Gini, Eltetö, Köves & Szulc) ont été développés pour les comparaisons spatiales de prix (ils sont notamment utilisés pour les parités de pouvoir d'achat). L'objectif de la procédure GEKS traditionnelle est que lorsque le pouvoir d'achat du pays A est directement comparé à celui du pays C, la comparaison devrait produire le même résultat que si cette comparaison était effectuée indirectement en comparant d'abord le pays A avec le pays B, puis le pays B avec le pays C. La procédure GEKS est donc transitive en ce sens qu'elle est indépendante du pays que l'on choisit comme base. Cette méthode n'est donc logiquement pas utilisée pour les comparaisons spatiales des prix pour les indices des prix à la consommation, mais pour les comparaisons de prix au fil du temps (temporelles) (Ivancic, Diewert & Fox, 2011). La transitivité intertemporelle par l'intégration de plusieurs périodes assure l'absence de chain drift dû à la variation du chiffre d'affaires de mois en mois.

La méthode utilise tous les produits qui correspondent les uns aux autres – matched model - en appliquant ensuite une moyenne géométrique non pondérée de tous les indices de prix bilatéraux possibles, chaque période servant de période de référence. Les indices de prix bilatéraux utilisés comme input sont des indices superlatifs. Les indices bilatéraux de Fisher ont été initialement proposés (= moyenne géométrique des indices de Laspeyres et de Paasche qui utilisent pour la pondération respectivement la période de référence et la période actuelle). Logiquement, d'autres indices superlatifs, comme un indice de Törnqvist, sont également possibles. Étant donné qu'ils ont déjà été utilisés précédemment, nous utiliserons ici et dans les prochaines parties du texte les indices de Törnqvist dans l'indice GEKS. Un indice GEKS pour la période de 0 à t peut être exprimé de la manière suivante:

$$P_{GEKS}^{0,t} = \prod_{l=0}^{T} \left({P^{0l}}/{P^{tl}} \right)^{\left(1/{T+1}\right)} = \prod_{l=0}^{T} \left(P^{0l} P^{lt}\right)^{\left(1/{T+1}\right)} \qquad \textit{Équ. 12}$$

où P^{0l} et P^{lt} sont des indices de Törnqvist bilatéraux entre respectivement la période 0 et la période l (avec 0 comme période de base), et entre la période l et la période t (avec l comme période de référence), et étant donné qu'un indice de Törnqvist satisfait au test de réversibilité dans le temps, $P^{lt} = 1/P^{tl}$. Les paramètres 0, l et t sont des périodes de la série chronologique T (=0, 1, …, T). Dans les indices des prix bilatéraux, seuls les éléments qui correspondent (matched items) sont utilisés. Dans la formule ci-dessus, l'indice des prix de la période t dépend aussi des prix des périodes futures qui ne sont pas encore disponibles (t+1, t+2 , …). Dans la pratique, on ne recalcule que la période la plus récente T. La formule peut dès lors être réécrite comme suit:

$$P_{GEKS}^{0,T} = \prod_{t=0}^{T} \left(P^{0t} P^{tT}\right)^{\left(1/{T+1}\right)} \qquad \textit{Équ. 13}$$

Maintenant, lorsque l'indice GEKS est calculé dans le tableau ci-dessous pour les mêmes données que dans l'exemple de chain drift d'un indice en chaîne, on peut constater que l'indice d'avril revient à 100 et qu'il n'y a pas de chain drift.

Tableau 2: Exemple fictif de chain drift pour un indice de Törnqvist en chaîne par rapport à un indice GEKS

	Janvier		Février		Mars		Avril	
	Prix	Chiffre d'affaires	Prix	Chiffre d'affaires	Prix	Chiffre d'affaires	Prix	Chiffre d'affaires
Produit A	2,50	10.000	2,00	500.000	2,50	2.000	2,50	10.000
Produit B	3,00	10.000	3,00	10.000	3,00	10.000	3,00	10.000
Indice de Törnqvist en chaîne	100		84,78		96,35		96,35	
GEKS (Törnqvist)	100		85,57		99,07		100	

3.2. Geary-Khamis

Une autre méthode de comparaison spatiale des prix qui a été transformée en méthode de comparaison intertemporelle de prix est la méthode Geary-Khamis (Chessa, 2016). Elle repose sur le principe d'indice de valeur unitaire, avec pour un produit i un prix unitaire p_i^t et une quantité vendue q_i^t. Lorsqu'un ensemble de produits G_t n'est pas homogène, les quantités ne sont pas facilement agrégées. À l'aide des facteurs d'ajustement de la qualité v_i, la combinaison entre les différents produits peut toutefois être réalisée. Ces facteurs transforment les quantités vendues q_i^t en "unités communes" $v_i q_i^t$. Et les prix p_i^t des différents produits i peuvent ainsi être transformés en "prix ajustés en fonction de la qualité" $\frac{p_i^t}{v_i}$. On obtient ainsi une "valeur unitaire ajustée en fonction de la qualité" $\tilde{p}^t$ pour un ensemble (non homogène) de produits au mois t:

$$\tilde{p}^t = \frac{\sum_{i \in G_t} p_i^t q_i^t}{\sum_{i \in G_t} v_i \, q_i^t} \qquad \textit{Équ. 14}$$

L'indice qui en résulte (parfois aussi appelé "indice de valeur unitaire ajusté en fonction de la qualité") est alors:

$$P^{0,t} = \frac{\tilde{p}^t}{\tilde{p}^0}$$

$$= \frac{\sum_{i \in G_t} p_i^t q_i^t / \sum_{i \in G_0} p_i^0 q_i^0}{\sum_{i \in G_t} v_i q_i^t / \sum_{i \in G_0} v_i q_i^0} \qquad \textit{Équ. 15}$$

L'équation 15 peut également être perçue comme le rapport entre un indice de chiffre d'affaires (numérateur) et un indice de quantité pondéré (dénominateur). La méthode Geary-Khamis définit les facteurs d'ajustement de la qualité v_i comme:

$$v_i = \frac{\sum_{z=0}^{T} q_i^z / P^{0,z}}{\sum_{z=0}^{T} q_i^z} \qquad \textit{Équ. 16}$$

La valeur unitaire ajustée en fonction de la qualité est donc, selon la méthode Geary-Khamis, basée sur toutes les données disponibles du mois 0 au mois T. Le calcul de l'indice de prix de valeur unitaire ajusté en fonction de la qualité $P^{0,t}$ dépend des facteurs d'ajustement de la qualité v_i, qui dépendent eux-mêmes de l'indice de prix $P^{0,t}$. Le calcul de ces valeurs doit donc être effectué au même moment, éventuellement à l'aide d'une procédure itérative, où l'on part d'un indice de prix fictif et où l'on travaille ensuite de manière itérative jusqu'à atteindre la convergence.

De nouveau, l'exemple fictif peut démontrer qu'aucun chain drift ne se produit avec la méthode Geary-Khamis (l'indice revient à 100 en avril). Naturellement, les indices des mois de février et mars diffèrent légèrement de ceux de la méthode GEKS. Cela sera également le cas pour les autres méthodes multilatérales abordées ci-dessous. La différence vient du fait que les facteurs d'ajustement de la qualité sont calculés différemment ou que le chiffre d'affaires (ou les unités vendues) est utilisé d'une autre manière.

Tableau 3: Exemple fictif d'un indice Geary-Khamis sans dérive

	Janvier		Février		Mars		Avril	
	Prix	Chiffre d'affaires	Prix	Chiffre d'affaires	Prix	Chiffre d'affaires	Prix	Chiffre d'affaires
Produit A	2,50	10.000	2,00	500.000	2,50	2.000	2,50	10.000
Produit B	3,00	10.000	3,00	10.000	3,00	10.000	3,00	10.000
Geary-Khamis	100		84,75		96,53		100	

3.3. Lehr (augmenté)

La méthode de Lehr est similaire à la méthode de Geary-Khamis (Lamboray, 2017; Von Auer, 2017), mais n'a pas recours à une méthode itérative complexe. Les facteurs d'ajustement de la qualité v_i de l'indice de Lehr sont définis comme suit:

$$v_i = \frac{p_i^0 q_i^0 + p_i^T q_i^T}{q_i^0 + q_i^T} \qquad \textit{Équ. 17}$$

Lorsque ceux-ci sont complétés dans l'équation 15, les valeurs unitaires ajustées en fonction de la qualité $\tilde{p}^0$ et $\tilde{p}^T$ peuvent être calculées, et l'indice de Lehr peut alors être déduit.

Cet indice n'utilise que les données du mois 0 et du mois T. Par conséquent, il ne s'agit également que d'une méthode bilatérale. La méthode de Lehr peut toutefois être étendue avec toutes les informations possibles en calculant les facteurs d'ajustement de la qualité comme suit:

$$v_i = \frac{\sum_{t=0}^{T} p_i^t q_i^t}{\sum_{t=0}^{T} q_i^t} \qquad \textit{Équ. 18}$$

Le fait de compléter ces facteurs dans l'équation 15 nous donne d'autres valeurs unitaires ajustées en fonction de la qualité $\tilde{p}^0$ et $\tilde{p}^T$, grâce auxquelles "l'indice de Lehr augmenté" peut être calculé. Cet indice est toujours un indice bilatéral, mais utilise de manière implicite toutes les informations disponibles, devenant ainsi multilatéral.

Aucun chain drift ne se produit dans la méthode de Lehr.

Tableau 4: Exemple fictif d'un indice de Lehr augmenté sans dérive

	Janvier		Février		Mars		Avril	
	Prix	Chiffre d'affaires	Prix	Chiffre d'affaires	Prix	Chiffre d'affaires	Prix	Chiffre d'affaires
Produit A	2,50	10.000	2,00	500.000	2,50	2.000	2,50	10.000
Produit B	3,00	10.000	3,00	10.000	3,00	10.000	3,00	10.000
Lehr augmenté	100		89,59		93,35		100	

3.4. Time Product Dummy

La méthode Time Product Dummy (TPD) est dérivée de la méthode Country Product Dummy (CPD), également utilisée pour les comparaisons spatiales de prix, et a recours aux estimations de régression. De Haan et Krsinich (2014) ont développé cette méthode pour réaliser une comparaison intertemporelle. Supposons N différents produits pendant une période $0, \dots, T$ dont bon nombre d'entre eux ne sont peut-être pas toujours disponibles. Un modèle de régression sur la base de toutes les données disponibles pour toutes les périodes $0, \dots, T$ peut alors être écrit comme suit:

$$\ln p_i^t = \alpha + \sum_{t=1}^{T} \delta^t D_i^t + \sum_{i=1}^{N-1} \gamma_i D_i + \varepsilon_i^t \qquad \textit{Équ. 19}$$

Où D_i^t est une variable time dummy qui prend la valeur de 1 si le produit i est disponible pendant la période t et de 0 dans les autres cas, et D_i est une variable dummy avec une valeur de 1 si l'observation est liée au produit i ou de 0 dans les autres cas. Les effets fixes spécifiques au produit sont présentés par les paramètres γ_i. δ^t représente les paramètres time dummy. À l'aide de la méthode des moindres carrés pondérés (avec les parts du chiffre d'affaires comme poids), les paramètres α, δ^t et γ_i peuvent être estimés.

L'indice TPD peut donc être obtenu en prenant l'exponentiel du paramètre $\hat{\delta}^t$ estimé:

$$P_{TPD}^{0,t} = \exp(\hat{\delta}^t) \qquad \textit{Équ. 20}$$

Pour un ensemble de produits G_t, les prix ajustés en fonction de la qualité au mois t peuvent être écrits comme suit:

$$\tilde{p}^t = \prod_{i \in G_t} \left(\frac{p_i^t}{v_i}\right)^{s_i^t} \qquad \textit{Équ. 21}$$

Où le prix est corrigé au moyen des effets fixes spécifiques au produit, les facteurs d'ajustement de la qualité sont $v_i = \exp(\gamma_i)$. L'indice TPD peut donc être écrit comme suit:

$$P_{TPD}^{0,t} = \frac{\tilde{p}^t}{\tilde{p}^0} = \frac{\prod_{i \in G_t} \left(\frac{p_i^t}{\exp(\hat{\gamma}_i)}\right)^{s_i^t}}{\prod_{i \in G_0} \left(\frac{p_i^0}{\exp(\hat{\gamma}_i)}\right)^{s_i^0}} \qquad \textit{Équ. 22}$$

Comme le montre le tableau ci-dessous, aucune dérive ne survient.

Tableau 5: Exemple fictif d'un indice Time Product Dummy sans dérive

	Janvier		Février		Mars		Avril	
	Prix	Chiffre d'affaires	Prix	Chiffre d'affaires	Prix	Chiffre d'affaires	Prix	Chiffre d'affaires
Produit A	2,50	10.000	2,00	500.000	2,50	2.000	2,50	10.000
Produit B	3,00	10.000	3,00	10.000	3,00	10.000	3,00	10.000
Time Product Dummy	100		80,60		99,78		100	

3.5. Régression hédonique

Lorsque les données contiennent également des caractéristiques sur le produit, elles peuvent être utilisées dans des méthodes dites hédoniques pour enregistrer les changements en termes de qualité. Les techniques de régression des indices des prix à la consommation sont appelées régressions hédoniques. Le prix est toujours exprimé en fonction des caractéristiques.

Lors du remplacement d'un ancien produit par un nouveau, les méthodes hédoniques n'ont pas pour objectif d'estimer la différence de qualité sur la base du modèle de régression (après quoi le prix peut être adapté pour tenir compte de cette différence). Cependant, elles estiment quel serait le prix relatif (ou le prix) d'un nouveau produit au cours de la période de référence ou d'un produit disparu au cours de la période actuelle.

En outre, les ajustements de qualité sont pris en compte dans l'ensemble de la population, et non seulement dans l'échantillon des produits pour lesquels des remplacements ont eu lieu. Ainsi, les remplacements 1 pour 1 ne seraient pas possibles, par exemple parce que le nombre de produits dans un segment diminuerait au fil du temps, et qu'il n'y aurait pas de nouveau remplaçant pour chaque produit disparu.

La forme fonctionnelle de régression hédonique la plus couramment utilisée est un modèle semi-logarithmique linéaire dans lequel le prix subit une transformation logarithmique[3]. Cette transformation logarithmique fait notamment en sorte que les résidus du modèle sont moins caractérisés par l'hétéroscédasticité. La forme fonctionnelle peut être exprimée comme suit pour une série chronologique avec deux périodes adjacentes (t=0,1):

$$\ln(p_i^t) = \alpha^t + \sum_{k=1}^{K} \beta_k^t x_{ik} + \varepsilon_i^t \qquad \textit{Équ. 23}$$

Où p_i^t est le prix du produit i pendant la période t, dont on prend ensuite le logarithme naturel. x_{ik} représente la caractéristique k (k=1, …, K) du produit i, avec le paramètre correspondant β_k^t. α^t est l'intercept, les termes de perturbation ε_i^t sont supposés indépendamment avec une valeur probable de 0 et une variance constante.

La variante hédonique time dummy est notée comme suit:

$$ln(p_i^t) = \alpha + \sum_{t=1}^{T} \delta^t D_i^t + \sum_{k=1}^{K} \beta_k x_{ik} + \varepsilon_i^t \qquad \textit{Équ. 24}$$

Il s'agit d'une extension de l'équation 23 avec une time dummy D_i^t qui prend la valeur de 1 si le produit i est disponible pendant la période t et de 0 autrement, similaire à la méthode Time Product Dummy décrite ci-dessus.

Cette méthode ne s'applique pas aux scanner data des supermarchés, étant donné que pour cette gamme de produits, les caractéristiques déterminant le produit sont trop limitées. Elle peut toutefois être utilisée dans des calculs de scanner data de l'électronique grand public.

4. ROLLING YEAR ET SPLICING

Dans les méthodes multilatérales décrites ci-dessus, les indices calculés précédemment sont revus lors de l'ajout de nouvelles données mensuelles. Par exemple, dans le cas d'un indice GEKS où la période est étendue (T+1), par exemple au mois suivant pour l'IPC, l'équation prendra la forme suivante:

$$P_{GEKS}^{0,T+1} = \prod_{t=0}^{T+1} (P^{0t} P^{tT+1})^{(1/T+2)} \qquad \textit{Équ. 25}$$

[3] Pour plus d'informations sur la régression hédonique : Triplet, J. (2006), Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes. Paris: OCDE. de Haan, J. (2010) 'Hedonic Price Indexes: A Comparison of Imputation, Time Dummy and 'Re-Pricing' Methods', Journal of Economics and Statistics Vol. 230, No. 6, Themenheft: Index Number Theory and Price Statistics pp. 772-791.

Comme tous les indices bilatéraux de l'indice GEKS sont calculés sur la base de toutes les correspondances dans l'ensemble de données, les indices pour les périodes 1,2,..., T - calculés sur la base de l'équation ci-dessus - sont donc différents des indices calculés précédemment. Cela signifierait également que les indices précédents doivent être revus à chaque fois, ce qui serait naturellement problématique pour les indices déjà publiés. Le même problème des révisions s'applique également aux autres méthodes, étant donné que les facteurs d'ajustement de la qualité v_i sont modifiés par l'ajout de la nouvelle période, modifiant ainsi également les indices du passé.

De telles révisions doivent être évitées pour les calculs officiels de l'IPC. Pour résoudre ce problème de révisions, plusieurs solutions sont possibles, en utilisant à chaque fois une période mobile de 13 mois (rolling year) pour calculer les indices. On utilise souvent une période de 13 mois parce que cela correspond à un "an" pour la mesure de l'inflation (par exemple, de décembre à décembre) et représente également la période la plus courte possible pour les produits soumis à des variations saisonnières. Une période de 13 mois correspond également à la période utilisée pour les imputations des produits "hors échantillons" dans la méthode dynamique. Si l'on travaille avec une rolling year dans le cadre de la méthode GEKS, la méthode s'appelle alors la méthode rolling year GEKS (RYGEKS)[4]. L'utilisation d'une rolling year permet de ne pas réviser les indices publiés des périodes précédentes. Reste une question: comment coupler différentes périodes (splicing) ? Pour cela, il existe un certain nombre de méthodes qui diffèrent dans la manière dont le splicing est effectué entre l'indice qui est estimé avec la période précédente avec $t = 0, \dots, T$ et l'indice de la période suivante $t = 1, \dots, T + 1$ (où T est égal à la durée de la période). Il s'agit du movement splice, window splice, half splice et mean splice. Chessa (2016) propose également la méthode Fixed Base Monthly Expanding Window (FBEW) avec une période limitée et chaque année un mois de base fixe (décembre). Dans cette méthode, la période est étendue chaque mois au mois suivant. Lamboray (2017) propose un mélange entre la méthode FBEW et le movement splice. Cette méthode utilise une période mobile (rolling window) où le dernier mois est comparé au mois de décembre précédent. Ce mois de décembre est donc le mois de base fixe (fixed base) et la méthode est intitulée Fixed Base Moving Window (FBMW).

Toutes les méthodes de couplage seront abordées en détail ci-dessous. En raison de l'utilisation de différentes périodes, la notation de l'indice doit être adaptée. L'indice de l'indice mentionnera la période utilisée, et l'exposant de l'indice indiquera la période pour laquelle l'indice est calculé. Dès lors, $P_{2,14}^{3,4}$ est égal à un indice calculé pour le mois 4 par rapport au mois 3 sur la base d'une période égale à 13 mois qui couvre les mois $t = 2, \dots, 14$. Dans une formule plus générale, un indice GEKS peut par exemple s'écrire comme suit:

$$GEKS_{0,T}^{a,b} = \prod_{l=0}^{T} (P^{al} P^{lb})^{(1/T+1)} \qquad \textit{Équ. 26}$$

4.1. Movement splice

La première méthode de splicing présentée est la méthode movement splice (MS) (de Haan & van der Grient, 2011). Dans cette méthode, l'évolution mensuelle la plus récente (movement) de l'indice, calculée à l'aide de la nouvelle période de 13 mois ($t = 1, \dots, 13$), est couplée à la série chronologique existante sur la base de l'indice des 13 premiers mois ($t = 0, \dots, 12$). La formule générale peut s'écrire comme suit:

$$P_{MS}^{0,t} = P_{MS}^{0,t-1} P_{t-T+1,t}^{t-1,t} \qquad \textit{Équ. 27}$$

Voici un exemple plus détaillé utilisant la méthode (RY)GEKS[5], dans lequel l'évolution de mois en mois est écrite de manière explicite en tant que rapport entre deux indices. Avec cette option de splicing, la méthode RYGEKS commencera donc par $RYGEKS_{MS}^{0,12}$, qui est égal à un indice GEKS "standard":

---

[4] Bien entendu, il est possible de travailler avec une période d'une autre durée que 13 mois. Dans ce cas, il n'est plus vraiment question d'une année, et le nom rolling window est plus logique que rolling year, bien que naturellement, on peut se poser la question de savoir dans quelle mesure 13 mois représentent une année. Dans cet article, on utilise simplement la dénomination standard RYGEKS, même si la durée de la période (T) reste ouverte dans les formules générales. Dans la pratique, on suppose toutefois normalement une période de 13 mois.

[5] Le renvoi est similaire pour les autres méthodes.

$$RYGEKS_{MS}^{0,12} = GEKS_{0,12}^{0,12} = \prod_{t=0}^{12} (P^{0t} P^{t12})^{(1/13)} \qquad \textit{Équ. 28}$$

Le $RYGEKS_{MS}^{0,12}$ obtenu ci-dessus peut désormais être multiplié par les indices GEKS mensuels les plus récents $(GEKS_{1,13}^{1,13}/GEKS_{1,13}^{1,12})$ issus de la nouvelle période afin d'obtenir le $RYGEKS_{MS}^{0,13}$ pour le mois actuel. La formule est la suivante:

$$RYGEKS_{MS}^{0,13} = GEKS_{0,12}^{0,12} \left(\frac{GEKS_{1,13}^{1,13}}{GEKS_{1,13}^{1,12}} \right) \qquad \textit{Équ. 29}$$

L'équation ci-dessus du $RYGEKS_{MS}^{0,13}$ peut être réécrite de manière plus simple en tant que produit de l'ancien $RYGEKS_{MS}^{0,12}$ et du $GEKS_{1,13}^{12,13}$ qui mesure l'évolution des prix entre le dernier mois de l'ancienne période $(t = 12)$ et le mois supplémentaire $(t = 13)$ dans la nouvelle période sur la base de l'entièreté de la nouvelle période. Cela s'explique par le fait que

$$\begin{aligned}
RYGEKS_{MS}^{0,13} &= GEKS_{0,12}^{0,12} \left(\frac{GEKS_{1,13}^{1,13}}{GEKS_{1,13}^{1,12}} \right) \\
&= \prod_{t=0}^{12} (P^{0t} P^{t12})^{(1/13)} \frac{\prod_{t=1}^{13} (P^{1t} P^{t13})^{(1/13)}}{\prod_{t=1}^{13} (P^{1t} P^{t12})^{(1/13)}} \\
&= \prod_{t=0}^{12} (P^{0t} P^{t12})^{(1/13)} \frac{\prod_{t=1}^{13} (P^{t13})^{(1/13)}}{\prod_{t=1}^{13} (P^{t12})^{(1/13)}} \\
&= \prod_{t=0}^{12} (P^{0t} P^{t12})^{(1/13)} \prod_{t=1}^{13} (P^{12t} P^{t13})^{(1/13)} \\
&= GEKS_{0,12}^{0,12} GEKS_{1,13}^{12,13} \qquad \textit{Équ. 30}
\end{aligned}$$

Le mois suivant $(t = 0, 1, \ldots, 14)$, l'indice qui vient d'être obtenu peut à nouveau être multiplié par le nouvel indice mensuel.

$$RYGEKS_{MS}^{0,14} = RYGEKS_{MS}^{0,13} \; GEKS_{2,14}^{13,14} \qquad \textit{Équ. 31}$$

Visuellement, le movement splice $RYGEKS_{MS}^{0,14}$ peut également être représenté comme suit:

		Mois														
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Période	0															
	1															
	2															

où chaque ligne représente la période et chaque colonne représente les mois. Les cellules vertes indiquent quels indices sont utilisés à partir de quelle période pour obtenir l'indice d'un mois donné. Les cellules rouges représentent par période les informations utilisées, mais pas publiées.

Dans la forme plus générale, le $RYGEKS_{MS}^{0,t}$ est donc égal à:

$$RYGEKS_{MS}^{0,t} = RYGEKS_{MS}^{0,t-1} GEKS_{t-T+1,t}^{t-1,t} \qquad \textit{Équ. 32}$$

Où $RYGEKS_{MS}^{0,t}$ est toujours étendu en raison du couplage de l'indice précédent ($RYGEKS_{MS}^{0,t-1}$) avec un indice multilatéral qui mesure l'évolution des prix entre le dernier mois de la période $(t-1)$ précédente et le dernier mois de la nouvelle période (t) calculée pour l'ensemble de la période $(t = t-T+1, t)$.

4.2. Window splice

Le movement splice décrit ci-dessus tient donc seulement compte de la dernière évolution mensuelle de la nouvelle période. La méthode du window splice proposée par Krsinich (2016) prend l'autre extrémité et couple l'estimation de l'ensemble des 13 nouveaux mois - encore une fois seulement dans le cas où T représente à chaque fois 13 mois - à l'indice des 12 derniers mois (issu de l'estimation des 13 mois précédents).

La formule générale de la méthode du window splice est la suivante:

$$P_{WS}^{0,t} = P_{WS}^{0,t-1} \frac{P_{t-T+1,t}^{t-T+1,t}}{P_{t-T,t-1}^{t-T+1,t-1}} \qquad \textit{Équ. 33}$$

Voici à nouveau un exemple plus détaillé dérivé de la formule générale à l'aide de l'indice GEKS. L'indice avec le window splice $P_{RYGEKS_{WS}}^{0,12}$ pour les 13 premiers mois est calculé comme un indice GEKS standard et est donc identique à celui issu de la méthode du movement splice:

$$RYGEKS_{WS}^{0,12} = \prod_{t=0}^{12} (P^{0t} P^{t12})^{(1/13)} = RYGEKS_{MS}^{0,12} \qquad \textit{Équ. 34}$$

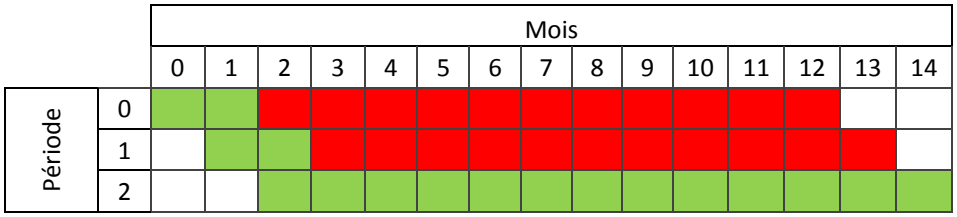
À partir du 14e mois, l'indice du window splice ($RYGEKS_{WS}^{0,13}$) diffère de celui du movement splice ($RYGEKS_{MS}^{0,13}$). $RYGEKS_{WS}^{0,13}$ est calculé en couplant l'évolution des prix de l'ensemble de la nouvelle période de 13 mois $(t = 1, \dots, 13)$, à savoir $GEKS_{1,13}^{1,13}$, à l'indice des prix des 12 derniers mois ($GEKS_{0,12}^{0,1}$) obtenu à l'aide de la période précédente $(t = 0, \dots, 12)$. L'indice $RYGEKS_{WS}^{0,13}$ pour toute la période d'estimation $(t = 0, \dots, 13)$ peut donc s'écrire en tant que produit de deux indices, tout comme le movement splice:

$$\begin{aligned} RYGEKS_{WS}^{0,13} &= \prod_{t=0}^{12} (P^{0t} P^{t1})^{(1/13)} \prod_{t=1}^{13} (P^{1t} P^{t13})^{(1/13)} \\ &= GEKS_{0,12}^{0,1} GEKS_{1,13}^{1,13} \end{aligned} \qquad \textit{Équ. 35}$$

Si la période d'estimation est étendue à l'aide d'une période supplémentaire $(t = 0, \dots, 14)$, l'indice est alors égal à:

$$\begin{aligned} RYGEKS_{WS}^{0,14} &= \prod_{t=0}^{12} (P^{0t} P^{t1})^{(1/13)} \prod_{t=1}^{13} (P^{1t} P^{t2})^{(1/13)} \prod_{t=2}^{14} (P^{2t} P^{t14})^{(1/13)} \\ &= GEKS_{0,12}^{0,1} GEKS_{1,13}^{1,2} GEKS_{2,14}^{2,14} \end{aligned} \qquad \textit{Équ. 36}$$

Visuellement, le window splice pour un $RYGEKS_{WS}^{0,14}$ peut être représenté comme suit:



L'indice issu de la nouvelle période ($t = 2, \dots, 14$) est donc couplé à l'indice du même mois ($t = 2$) qui correspond au départ de la nouvelle période. Ce dernier est l'indice d'une seule période d'estimation précédente ($GEKS_{1,13}^{1,2}$). L'indice obtenu est alors couplé à l'indice de deux périodes d'estimation précédentes $GEKS_{0,12}^{0,1}$. Dans la forme plus générale, le $RYGEKS_{WS}^{0,t}$ est égal à:

$$RYGEKS_{WS}^{0,t} = GEKS_{0,T}^{0,1} GEKS_{1,T+1}^{1,2} \dots GEKS_{t-T,t}^{t-T+1,t} \qquad \textit{Équ. 37}$$

Sur la base de la formule ci-dessus, l'indice window splice d'une seule période supplémentaire ($RYGEKS_{WS}^{0,t+1}$) ne peut pas facilement être exprimé en fonction de l'indice window splice précédent ($RYGEKS_{WS}^{0,t}$). Cela est toutefois possible si le calcul est quelque peu réécrit. Ainsi, $RYGEKS_{WS}^{0,13} = GEKS_{0,12}^{0,1} GEKS_{1,13}^{1,13}$ peut être réécrit en tant que produit de l'indice précédent ($GEKS_{0,12}^{0,12}$) et le ratio des indices EKS des deux périodes qui commencent chacune au mois de splicing ($GEKS_{1,13}^{1,13}/GEKS_{0,12}^{1,12}$). Il en découle:

$$\begin{aligned} RYGEKS_{WS}^{0,13} &= GEKS_{0,12}^{0,1} GEKS_{1,13}^{1,13} \left(GEKS_{0,12}^{0,12}/GEKS_{0,12}^{0,12}\right) \\ &= GEKS_{0,12}^{0,12} \left(GEKS_{1,13}^{1,13}/GEKS_{0,12}^{1,12}\right) \end{aligned} \qquad \textit{Équ. 38}$$

Si la formule est prolongée d'une période supplémentaire ($RYGEKS_{WS}^{0,14}$) sur la base de la méthode décrite ci-dessus - dans laquelle le ratio des indices EKS est tiré des périodes actuelle et précédente, à chaque fois à partir du mois de splicing - il apparaît alors clairement que tout comme dans la méthode du movement splice, un window splice peut être exprimé en tant que produit couplé à l'indice précédent:

$$\begin{aligned} RYGEKS_{WS}^{0,14} &= RYGEKS_{WS}^{0,13} \left(GEKS_{2,14}^{2,14}/GEKS_{1,13}^{2,13}\right) \\ &= GEKS_{0,12}^{0,1} GEKS_{1,13}^{1,13} \left(GEKS_{2,14}^{2,14}/GEKS_{1,13}^{2,13}\right) \\ &= GEKS_{0,12}^{0,1} \left(GEKS_{1,13}^{1,13}/GEKS_{1,13}^{2,13}\right) GEKS_{2,14}^{2,14} \\ &= GEKS_{0,12}^{0,1} GEKS_{1,13}^{1,2} GEKS_{2,14}^{2,14} \end{aligned} \qquad \textit{Équ. 39}$$

Où l'équation est égale à celle obtenue dans l'équation 36. La méthode de calcul dans laquelle l'indice window splice actuel ($RYGEKS_{WS}^{0,t}$) est exprimé par rapport à l'indice window splice précédent ($RYGEKS_{WS}^{0,t-1}$) peut naturellement aussi s'écrire de manière plus générale:

$$RYGEKS_{WS}^{0,t} = RYGEKS_{WS}^{0,t-1} \frac{GEKS_{t-T+1,t}^{t-T+1,t}}{GEKS_{t-T,t-1}^{t-T+1,t-1}} \qquad \textit{Équ. 40}$$

Où $RYGEKS_{WS}^{0,t}$ est à chaque fois prolongé en couplant l'indice précédent ($RYGEKS_{WS}^{0,t-1}$) à un facteur qui est le rapport de

deux indices multilatéraux qui mesurent l'évolution des prix sur la base de la nouvelle période, d'une part, et de la période précédente, d'autre part. La période de départ de l'évolution des prix mesurée dans ce facteur est à chaque fois le premier mois des deux périodes ($t - T + 1$). La période de fin est à chaque fois le dernier mois de la période respective.

Quelle est la différence entre le $RYGEKS_{WS}^{0,t}$ et le $RYGEKS_{MS}^{0,t}$? Cela peut être facilement vérifié en examinant l'indice après le premier splicing $RYGEKS^{0,13}$ et en exprimant ensuite le $RYGEKS_{WS}^{0,13}$ en fonction du $RYGEKS_{MS}^{0,13}$.

Il ressort de l'équation 30 que $RYGEKS_{MS}^{0,13} = GEKS_{0,12}^{0,12} GEKS_{1,13}^{12,13}$, qui peut être réécrit comme suit: $RYGEKS_{MS}^{0,13} = GEKS_{0,12}^{0,1} GEKS_{0,12}^{1,12} GEKS_{1,13}^{12,13}$. Il ressort de l'équation 35 que $RYGEKS_{WS}^{0,13} = GEKS_{0,12}^{0,1} GEKS_{1,13}^{1,13}$, qui peut être réécrit comme suit: $RYGEKS_{WS}^{0,13} = GEKS_{0,12}^{0,1} GEKS_{1,13}^{1,12} GEKS_{1,13}^{12,13}$. La seule différence entre les deux formules est logiquement le calcul de l'indice pour la période 1 à 12, calculé dans la méthode du movement splice sur la base de la première période (0,...,12) $GEKS_{0,12}^{1,12}$ et dans la méthode du window splice sur la base de la deuxième période (1,...,13) $GEKS_{1,13}^{1,12}$. Par conséquent, le window splice peut être calculé à partir du movement splice (et inversement) en le multipliant par le rapport de ces deux indices différents ($GEKS_{1,13}^{1,12} / GEKS_{0,12}^{1,12}$).

$$\frac{RYGEKS_{WS}^{0,13}}{RYGEKS_{MS}^{0,13}} = \frac{GEKS_{0,12}^{0,1} GEKS_{1,13}^{1,12} GEKS_{1,13}^{12,13}}{GEKS_{0,12}^{0,1} GEKS_{0,12}^{1,12} GEKS_{1,13}^{12,13}}$$

$$RYGEKS_{WS}^{0,13} = \frac{GEKS_{1,13}^{1,12}}{GEKS_{0,12}^{1,12}} RYGEKS_{MS}^{0,13}$$

Équ. 41

4.3. Half splice

Outre le movement splice et le window splice qui effectuent le splicing sur les extrémités des mois communs aux périodes (à savoir la fin et le début, respectivement), celui-ci peut également être effectué en milieu de période. C'est ce qu'on appelle le half splice (de Haan, 2015).

Le half splice (couplage en milieu de période) se produirait alors au moment $t = \frac{T+1}{2}$ si T est impair et au moment $t = \frac{T}{2}$ si T est pair. L'indice $P_{HS}^{0,12}$ des 13 premiers mois est logiquement identique aux deux méthodes de splicing précédentes. Lorsque la période compte 13 mois, qui est un nombre impair, le splicing se produira au 7e mois ($\frac{13+1}{2}$) de la période. À partir de la nouvelle période ($t = 1, \dots, 13$), l'indice half splice $P_{HS}^{0,13}$ peut être écrit comme suit:

$$P_{HS}^{0,13} = P_{0,12}^{0,7} P_{1,13}^{7,13}$$

Équ. 42

La formule générale de l'indice half splice est la suivante:

$$P_{HS}^{0,t} = P_{HS}^{0,t-1} \frac{P_{t-T+1,t}^{t-\frac{T+1}{2}+1,t}}{P_{t-T,t-1}^{t-\frac{T+1}{2}+1,t-1}}$$

Équ. 43

La formule dérivée est illustrée ci-dessous à l'aide de l'indice GEKS.

Dans le cas d'un indice GEKS, l'équation 42 devient:

$$RYGEKS_{HS}^{0,13} = GEKS_{0,12}^{0,7} GEKS_{1,13}^{7,13}$$

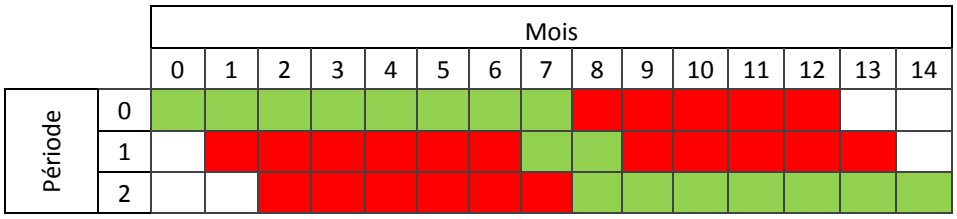
Équ. 44

Lorsque la série est prolongée d'un mois supplémentaire, le splicing se produira toujours au 7e mois de cette nouvelle période. Il s'agit toujours d'une période de plus que le mois de splicing précédent. L'indice half splice $RYGEKS_{HS}^{0,14}$ est alors égal à:

$$RYGEKS_{HS}^{0,14} = GEKS_{0,12}^{0,7} GEKS_{1,13}^{7,8} GEKS_{2,14}^{8,14}$$

Équ. 45

Visuellement, le half splice pour un $RYGEKS_{HS}^{0,14}$ peut être représenté comme suit:



$RYGEKS_{HS}^{0,14}$ peut désormais s'exprimer à nouveau en fonction de $RYGEKS_{HS}^{0,13}$

$$\frac{RYGEKS_{HS}^{0,14}}{RYGEKS_{HS}^{0,13}} = \frac{GEKS_{0,12}^{0,7} GEKS_{1,13}^{7,8} GEKS_{2,14}^{8,14}}{GEKS_{0,12}^{0,7} GEKS_{1,13}^{7,13}}$$

$$RYGEKS_{HS}^{0,14} = RYGEKS_{HS}^{0,13} \frac{GEKS_{2,14}^{8,14}}{GEKS_{1,13}^{8,13}}$$

Équ. 46

Dans la forme plus générale, le $RYGEKS_{HS}^{0,t}$ est égal à:

$$RYGEKS_{HS}^{0,t} = RYGEKS_{HS}^{0,t-1} \frac{GEKS_{t-T+1,t}^{t-\frac{T+1}{2}+1,t}}{GEKS_{t-T,t-1}^{t-\frac{T+1}{2}+1,t-1}}$$

Équ. 47

Où $RYGEKS_{HS}^{0,t}$ est à chaque fois prolongé en couplant l'indice précédent ($RYGEKS_{HS}^{0,t-1}$) à un facteur qui est le rapport de deux indices multilatéraux qui mesurent l'évolution des prix sur la base de la nouvelle période, d'une part, et de la période précédente, d'autre part. La période de départ de l'évolution des prix mesurée dans ce facteur est à chaque fois le mois du milieu de la période actuelle ($t - \frac{T+1}{2} + 1$) si T est impair. La période de fin est à chaque fois le dernier mois de la période respective.

4.4. Mean splice

Tant pour le movement splice et le window splice que le half splice, un choix est effectué quant au mois commun aux deux périodes qui sera utilisé comme mois de couplage. On peut toutefois considérer que le choix d'un mois de couplage est toujours arbitraire et qu'il n'est pas nécessaire de choisir un mois commun spécifique. La méthode qui tente de procéder de la sorte est appelée la méthode mean splice (Diewert & Fox, 2017). Cette méthode utilise la moyenne géométrique de tous les choix de splicing possibles, c'est-à-dire tous les mois communs à la période actuelle et à la période précédente. L'indice de départ est à nouveau identique à celui de toutes les méthodes précédentes. À partir de la nouvelle période ($t = 1, \dots, 13$), l'extension mensuelle sur la base de la méthode mean splice peut être écrite de manière générale comme suit:

$$P_{GMS}^{t-1,t} = \prod_{l=t-T+1}^{t-1} \left(\frac{P_{t-T+1,t}^{l,t}}{P_{t-T,t-1}^{l,t-1}} \right)^{\frac{1}{T-1}}$$

Équ. 48

Pour les méthodes générales de splicing, la formule peut donc s'écrire comme suit:

$$P_{GM}^{0,t} = P_{GMS}^{0,t-1} \prod_{l=t-T+1}^{t-1} \left(\frac{P_{t-T+1,t}^{l,t}}{P_{t-T,t-1}^{l,t-1}} \right)^{\frac{1}{T-1}} \qquad \textit{Équ. 49}$$

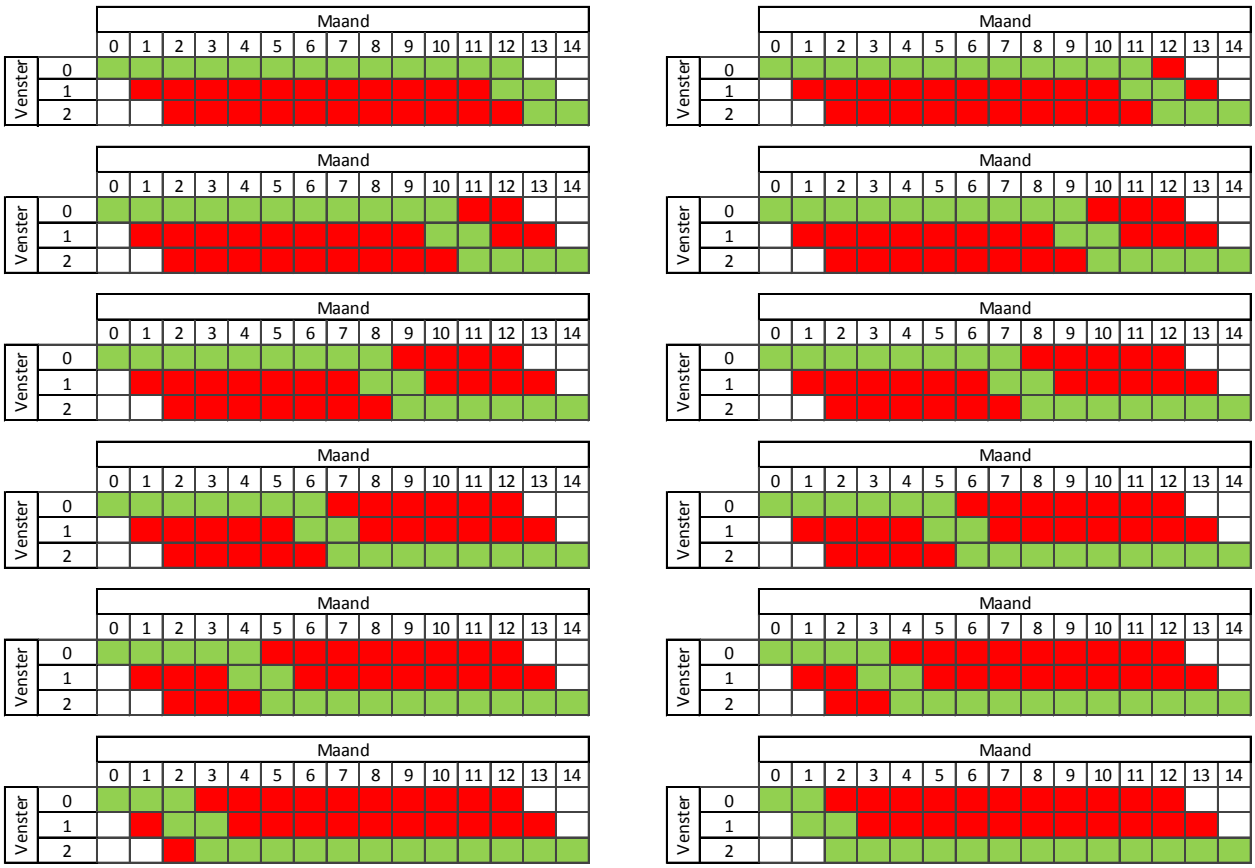
Dans le cas de la méthode GEKS et du premier glissement de la période, la formule devient:

$$RYGEKS_{GMS}^{12,13} = \prod_{l=1}^{12} \left(\frac{GEKS_{1,13}^{l,13}}{GEKS_{0,12}^{l,12}} \right)^{\frac{1}{12}} \qquad \textit{Équ. 50}$$

Celle-ci peut être couplée à l'indice calculé lors de la période précédente $RYGEKS_{GMS}^{0,12}$:

$$RYGEKS_{GMS}^{0,13} = RYGEKS_{GMS}^{0,12} \prod_{l=1}^{12} \left(\frac{GEKS_{1,13}^{l,13}}{GEKS_{0,12}^{l,12}} \right)^{\frac{1}{12}} \qquad \textit{Équ. 51}$$

Où l s'étend sur les mois communs à l'ancienne et à la nouvelle période. Visuellement, l'indice mean splice pour la période suivante ($RYGEKS_{GMS}^{0,14}$) peut être représenté comme la moyenne géométrique des combinaisons suivantes:



Dans la forme plus générale, le $RYGEKS_{GMS}^{0,t}$ est égal à:

$$RYGEKS_{GMS}^{0,t} = RYGEKS_{GMS}^{0,t-1} \prod_{l=t-T+1}^{t-1} \left(\frac{GEKS_{t-T+1,t}^{l,t}}{GEKS_{t-T,t-1}^{l,t-1}} \right)^{\frac{1}{T-1}} \qquad \textit{Équ. 52}$$

Où $RYGEKS_{GMS}^{0,t}$ est toujours prolongé par le couplage de l'ancien indice ($RYGEKS_{GMS}^{0,t-1}$) à un facteur qui représente la moyenne géométrique de tous les rapports des indices multilatéraux qui mesurent les évolutions de prix pour deux périodes consécutives. On utilise alors cette moyenne géométrique de tous les choix de splicing possibles au cours des mois communs aux deux périodes.

4.5. Fixed base monthly expanding window

Dans les quatre méthodes de splicing décrites ci-dessus, aucun mois ne joue le rôle de mois de base traditionnel. Cela n'est pas tout à fait surprenant, étant donné que le but des méthodes multilatérales est précisément de ne donner à aucun mois ce sens spécifique. Naturellement, dans les méthodes du movement splice, du window splice et du half splice, un mois donné de la période joue toutefois un rôle spécifique en tant que mois de couplage. Il s'agit ici toujours d'une signification numérique, à savoir respectivement le dernier mois, le premier mois ou le mois du milieu d'une période.

L'absence d'un "mois de base traditionnel" est contraire à la méthode actuelle d'indice en chaîne utilisée tant pour l'IPC national que pour l'IPCH[6] harmonisé, où décembre est toujours le mois de couplage annuel (ou mois de base). C'est pourquoi Chessa (2016) a également proposé une méthode d'extension ou de prolongation dans laquelle décembre continuerait à jouer ce rôle: la méthode fixed base monthly expanding window (FBEW).

Dans le cadre de la méthode FBEW, on commence toujours avec un mois de base fixe (fixed base month) (décembre de l'année précédente dans notre cas) et la série chronologique est prolongée d'un mois tous les mois. En janvier, la série chronologique est donc composée de 2 mois (indice bilatéral) et la durée totale de 13 mois est atteinte en décembre. Dans cette méthode, l'indice compare à chaque fois le mois actuel et décembre de l'année précédente et utilise pour chaque extension toutes les données disponibles. Un indice est donc à chaque fois exprimé par rapport à décembre de l'année précédente, comme dans les calculs traditionnels de l'indice. En multipliant les indices de décembre entre eux, on obtient une série à long terme. La formule de l'indice, avec le mois de base b, peut s'écrire comme suit:

$$P_{FBEW}^{0,t} = P_{b-T,b}^{b-T,b} P_{b,t}^{b,t} \qquad \textit{Équ. 53}$$

Cette méthode avec extension de la série chronologique permet de réviser légèrement l'indice de janvier à novembre lorsque la série est étendue. Cette révision n'est toutefois pas publiée. L'un des avantages de cette méthode est que cette méthode d'extension ne nécessite aucune rolling year pour pouvoir commencer. En principe, un seul couplage est effectué, à savoir le chaînage annuel avec le mois de décembre. Cela signifie que le chain drift, qui en théorie pourrait encore éventuellement se produire dans les autres méthodes étant donné qu'un couplage mensuel est encore nécessaire, est minimisé dans la méthode FBEW.

Dans le cas de la méthode GEKS, la FBEW est notée comme suit pour le premier glissement après 13 mois:

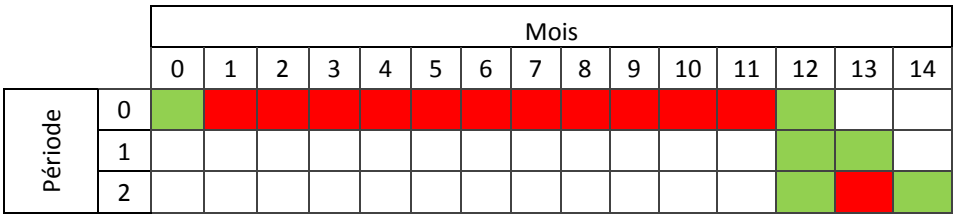
$$RYGEKS_{FBEW}^{0,13} = GEKS_{0,12}^{0,12} GEKS_{12,13}^{12,13} \qquad \textit{Équ. 54}$$

En cas d'extension à la période 14, l'équation devient:

$$RYGEKS_{FBEW}^{0,14} = GEKS_{0,12}^{0,12} GEKS_{12,14}^{12,14} \qquad \textit{Équ. 55}$$

où $GEKS_{12,14}^{12,13} \neq GEKS_{12,13}^{12,13}$ et il y a donc une révision implicite de l'indice du mois précédent. Visuellement, $RYGEKS_{FBEW}^{0,14}$ peut être représenté comme suit:

[6] L'IPCH est également calculé par Statbel. Les principes et la méthodologie de l'IPCH sont harmonisés autant que possible dans les différents pays européens.



L'indice du mois 12 est estimé sur la base des informations de la période 0 à 12[7]. Dans ce cas, cet indice est égal à celui des 13 premiers mois selon les méthodes du movement, half, window et mean splice. L'extension à la période 13 a recours aux informations des mois 12 et 13 et exprime l'indice par rapport à la période 12. Pour la période 14, on utilise les informations des périodes 12 à 14. L'indice est également exprimé par rapport à la période 12. Le couplage avec l'indice obtenu pour la période 12 crée une série à long terme.

4.6. Fixed base moving window

La méthode fixed base moving window (FBMW) est une méthode hybride proposée par Lamboray (2017). Elle utilise une période de 13 mois qui se déplace à chaque fois d'un mois et est donc similaire aux méthodes de splicing qui utilisent une année mobile. Contrairement à ces méthodes, dans le cadre de la méthode FBMW, l'indice est à chaque fois comparé au mois de décembre de l'année précédente, tout comme dans la méthode FBEW. L'avantage de cette méthode par rapport à la méthode FBEW est qu'elle utilise plus d'informations au cours des premiers mois de l'année. L'inconvénient par rapport à la méthode FBEW est que 13 mois sont nécessaires pour pouvoir commencer.

La formule de la méthode FBMW, avec à nouveau b comme mois de référence, peut s'écrire comme suit:

$$P_{FBMW}^{0,t} = P_{b-T,b}^{b-T,b} P_{t-T,t}^{b,t} \qquad \textit{Équ. 56}$$

Les indices des 13 premiers mois de la méthode FBMW sont identiques à ceux de la méthode FBEW. Ensuite, chaque comparaison des mois de base est identique dans les deux méthodes. Toutefois, les indices intermédiaires sont différents, étant donné que l'on travaille avec une période complète dans la méthode FBMW au lieu d'une série chronologique incrémentielle étendue dans la méthode FBEW.

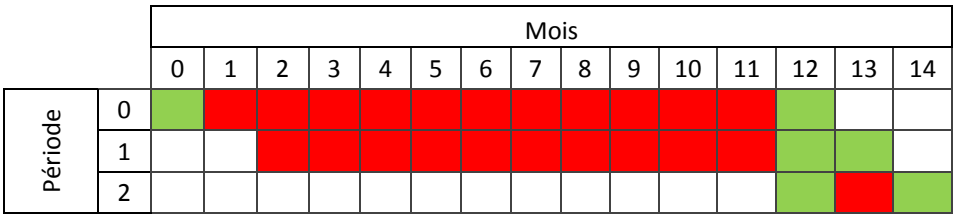
La différence entre les deux méthodes est plus claire lorsque l'on regarde l'indice GEKS. Pour la période 13, cet indice devient:

$$RYGEKS_{FBMW}^{0,13} = GEKS_{0,12}^{0,12} GEKS_{1,13}^{12,13} \qquad \textit{Équ. 57}$$

L'extension $GEKS_{1,13}^{12,13}$ dans l'équation 57 diffère de $GEKS_{12,13}^{12,13}$, comme dans l'équation 54 dans la méthode FBEW. Pour la période 13, il apparaît clairement que l'indice FBMW est égal à l'indice movement splice (équ. 30). L'indice pour la période 14 est:

$$RYGEKS_{FBMW}^{0,14} = GEKS_{0,12}^{0,12} GEKS_{2,14}^{12,14} \qquad \textit{Équ. 58}$$

et diffère du movement splice étant donné que l'extension $GEKS_{2,14}^{12,14}$ est utilisée au lieu de $GEKS_{2,14}^{13,14}$. Visuellement, cet indice peut être représenté comme suit:



[7] Pour cette méthode, il n'est pas nécessaire d'avoir une période de 13 mois pour pouvoir démarrer, la méthode peut directement être utilisée.

Il apparaît clairement que cet indice diffère de toutes les méthodes précédentes, mais qu'après 13 mois, cette méthode est identique à la méthode FBEW, étant donné que la période de 13 mois a glissé et que l'indice de décembre est lié à l'indice de décembre de l'année précédente.

5. RECHERCHE EMPIRIQUE

Les différentes méthodes multilatérales et options de splicing décrites ci-dessus ont été appliquées aux scanner data d'un détaillant. Les données comprennent 37 mois et englobent des produits des groupes COICOP "01.1 Produits alimentaires ", "01.2 Boissons non alcoolisées ", "02.1 Boissons alcoolisées " en "12.1.3.2 Articles pour l'hygiène corporelle et le bien-être, produits ésotériques et produits de beauté ". Il s'agit des groupes de produits ayant le poids le plus élevé pour lesquels des scanner data sont actuellement utilisées dans l'IPC. Les produits saisonniers (fruits et légumes frais) du groupe COICOP 01.1 ont été exclus, étant donné qu'ils sont calculés à l'aide d'une autre méthode dans l'indice officiel (Roels & Van Loon, 2017). Par ailleurs, un filtre de dumping et un filtre des valeurs aberrantes ont été utilisés dans les méthodes multilatérales comme ceux qui sont également utilisés dans la méthode dynamique.

Les méthodes multilatérales ont été appliquées aux groupes de produits qui se trouvent sous le niveau ECOICOP5. Il s'agit du même niveau auquel la méthode dynamique est actuellement appliquée. Pour les groupes COICOP visés ci-dessus, cela revient à environ 480 groupes de produits au total. Le même niveau a été choisi parce que le but de l'étude est d'examiner uniquement l'impact d'un changement de méthodologie à ce niveau, les procédures d'agrégation aux niveaux supérieurs sont donc restées inchangées. Les agrégats à un niveau supérieur sont obtenus à l'aide d'indices en chaîne annuels standards de type Laspeyres (pour plus d'informations, voir Roels & Van Loon, 2017).

Les relances (relaunches) de produits sont prises en compte en couplant les codes de produits (combinaison de travail manuel et de text mining). En cas de modification des quantités, par exemple lorsque le contenu de l'ancien produit s'élève à 250 grammes et celui du nouveau produit à 300 grammes, le prix est corrigé en fonction de la quantité. Cette méthode de travail est comparable à la manière dont les relances de produits sont actuellement effectuées dans le cadre de la méthode dynamique.

5.1. Résultats pour une période entière

En premier lieu, une comparaison entre la méthode dynamique et les méthodes multilatérales est effectuée lors du calcul sur l'ensemble de la période de 37 mois (full window). Ces indices multilatéraux sont par définition exempts de toute forme de chain drift, étant donné qu'aucune forme de chaînage n'est appliquée. Cet indice full window peut ensuite également être utilisé comme "référence" pour la comparaison avec les différentes options de splicing. L'utilisation de ces indices comme référence est naturellement discutable, étant donné qu'il faut faire un compromis entre la transitivité et la caractéristicité. Plus la période est longue, plus il est probable que les récentes évolutions des prix soient influencées par des prix du passé lointain, ce qui est naturellement défavorable (Fox, 2017). Étant donné que la période compte 37 mois, on suppose que l'impact pour la plupart des méthodes n'est pas particulièrement grand.

La figure 1 montre la différence entre la méthode dynamique et les quatre méthodes multilatérales testées (GEKS, Geary-Khamis (GK), Lehr, Time Product Dummy (TPD)) pour les groupes COICOP 01.1, 01.2, 02.1 et 12.1.3.2. Dans toutes les comparaisons effectuées, la période 1 a été assimilée à 100.

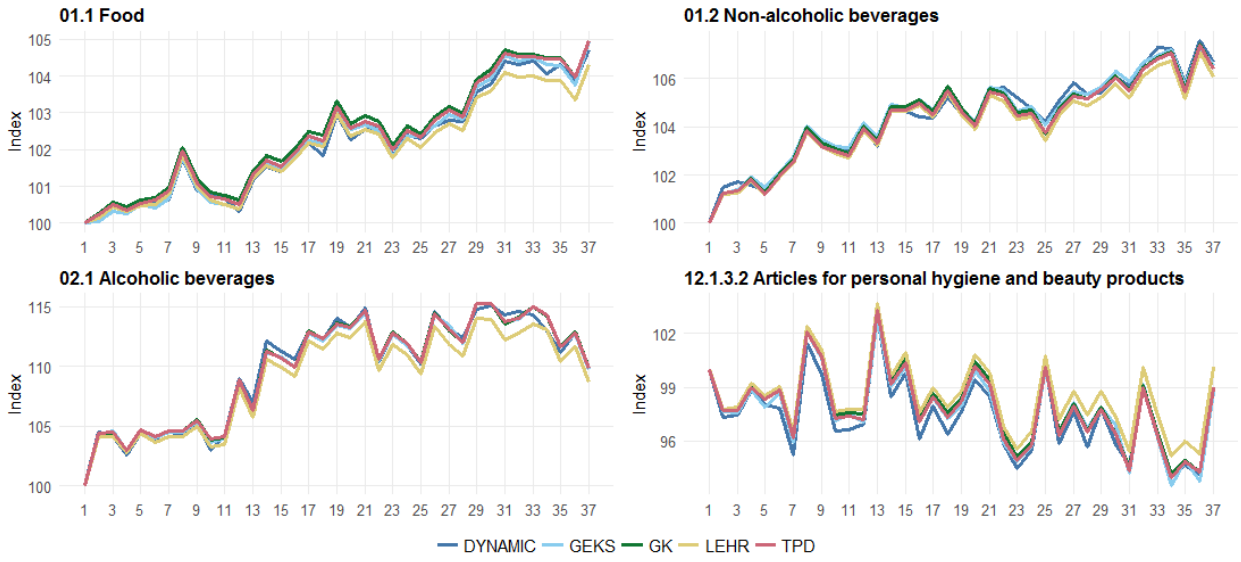


Figure 2 : Différence entre la méthode dynamique et les méthodes multilatérales (période entière)

Le tableau ci-dessous présente les différences moyennes, l'écart-type et la différence finale entre les méthodes multilatérales et la méthode dynamique. La différence est calculée par rapport à la méthode dynamique (c-à-d que la différence moyenne pour l'indice GEKS est égale à l'indice GEKS moyen moins l'indice dynamique moyen):

Tableau 6: Différences entre les méthodes multilatérales (période complète) et la méthode dynamique

COICOP	METHODE	GEKS	GK	LEHR	TPD
01.1	moyenne	0.05	0.25	-0.10	0.16
	(écart-type)	(0.14)	(0.13)	(0.17)	(0.11)
	Indice final	0.22	0.24	-0.41	0.24
01.2	moyenne	0.05	-0.05	-0.29	-0.14
	(écart-type)	(0.25)	(0.27)	(0.31)	(0.25)
	Indice final	-0.14	-0.27	-0.62	-0.28
02.1	moyenne	-0.02	0.05	-0.79	0.03
	(écart-type)	(0.45)	(0.43)	(0.57)	(0.44)
	Indice final	-0.10	-0.05	-1.18	-0.07
12.1.3.2	moyenne	0.33	0.58	1.05	0.42
	(écart-type)	(0.42)	(0.39)	(0.36)	(0.34)
	Indice final	0.01	0.27	1.38	0.29

Toutes les méthodes multilatérales, à l'exception de la méthode de Lehr, donnent des résultats similaires. La méthode GEKS donne le plus petit écart moyen pour tous les groupes COICOP. Les différences avec la méthode dynamique actuelle sont également très faibles. La méthode actuelle semble donc mesurer correctement l'inflation aux niveaux agrégés, malgré le fait qu'elle n'utilise pas de pondération explicite aux niveaux détaillés. De même, le couplage mensuel semble ne pas entraîner de dérive.

La méthode de Lehr diffère fortement de la méthode dynamique et des autres méthodes multilatérales. Elle présente également le plus grand écart final pour tous les groupes COICOP. Pour tous les groupes COICOP, la méthode de Lehr commence à s'écarter au fil du temps, alors que les trois autres méthodes multilatérales évoluent de la même manière et sont également comparables à la méthode dynamique. La méthode de Lehr sous-estime fortement l'inflation en cas de tendance à la hausse des prix et la surestime lorsque les prix diminuent. Les résultats obtenus diffèrent en partie de ceux de Lamboray (2017), qui montrait formellement que dans un contexte de hausse (baisse) des prix, l'indice de Lehr sous-estimait (surestimait) l'indice de Geary-Khamis, mais d'un point de vue empirique, les résultats étaient très similaires. Nous croyons même que de modestes variations de prix sur un grand nombre de produits peuvent entraîner une différence significative. Au vu de ces résultats, la méthode de Lehr semble susceptible de sous-estimer l'inflation lorsque les prix augmentent et de la surestimer lorsque les prix baissent, et n'est pas examinée de manière plus approfondie (le recours à des méthodes de splicing ou d'extension n'a pas résolu les problèmes décrits ci-dessus).

5.2. Résultats des méthodes de splicing et d'extension

Les six méthodes de splicing et d'extension décrites ci-dessus s'appliquent aux méthodes multilatérales, à l'exception de la méthode de Lehr pour les raisons visées ci-dessus. Le tableau ci-dessous donne les écarts moyens entre les indices qui ont recours à une année mobile combinée à une méthode de splicing ou d'extension et ceux qui utilisent la période complète (par exemple GEKS MOVE = GEKS MOVE moins GEKS période complète). Les différences sont à nouveau calculées pour les quatre groupes COICOP. L'écart-type et les écarts d'indice finaux sont également présentés.

Tableau 7: Différences entre les méthodes de splicing et d'extension pour les quatre groupes COICOP

COICOP 01.1

MÉTHODE/TYPE		MOVE	WINDOW	HALF	MEAN	FBEW	FBMW
GEKS	moyenne	0.02	0.02	0.02	0.02	0.01	0.02
	(écart-type)	(0.04)	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)
	Indice final	-0.03	-0.01	0.03	0.00	0.01	0.01
GK	moyenne	-0.16	-0.03	-0.09	-0.09	-0.16	-0.11
	(écart-type)	(0.10)	(0.10)	(0.06)	(0.06)	(0.09)	(0.05)
	Indice final	-0.29	0.15	0.02	0.00	-0.05	-0.05
TPD	moyenne	0.01	-0.10	-0.03	-0.04	-0.08	-0.02
	(écart-type)	(0.04)	(0.07)	(0.04)	(0.03)	(0.05)	(0.05)
	Indice final	0.04	-0.24	0.02	-0.02	-0.06	-0.06

COICOP 01.2

MÉTHODE/TYPE		MOVE	WINDOW	HALF	MEAN	FBEW	FBMW
GEKS	moyenne	-0.11	-0.11	-0.11	-0.11	-0.11	-0.09
	(écart-type)	(0.05)	(0.05)	(0.06)	(0.05)	(0.10)	(0.09)
	Indice final	-0.05	-0.05	0.00	-0.02	-0.02	-0.02
GK	moyenne	-0.21	-0.06	-0.09	-0.11	-0.15	-0.12
	(écart-type)	(0.11)	(0.11)	(0.08)	(0.08)	(0.20)	(0.15)
	Indice final	-0.21	0.11	-0.08	-0.11	-0.17	-0.17
TPD	moyenne	-0.10	-0.09	-0.04	-0.07	-0.09	-0.06
	(écart-type)	(0.07)	(0.07)	(0.07)	(0.06)	(0.18)	(0.14)
	Indice final	-0.06	-0.08	-0.08	-0.12	-0.21	-0.21

COICOP 02.1

MÉTHODE/TYPE		MOVE	WINDOW	HALF	MEAN	FBEW	FBMW
GEKS	moyenne	-0.01	-0.01	-0.04	-0.04	0.05	0.02
	(écart-type)	(0.05)	(0.05)	(0.06)	(0.06)	(0.10)	(0.07)
	Indice final	0.03	0.06	-0.03	-0.03	0.16	0.16
GK	moyenne	-0.14	-0.15	-0.17	-0.14	-0.04	-0.04
	(écart-type)	(0.10)	(0.11)	(0.10)	(0.09)	(0.19)	(0.13)
	Indice final	-0.10	-0.02	-0.16	-0.03	0.18	0.18
TPD	moyenne	0.00	-0.16	-0.10	-0.08	0.04	0.02
	(écart-type)	(0.08)	(0.12)	(0.08)	(0.07)	(0.16)	(0.11)
	Indice final	0.14	-0.21	-0.14	-0.04	0.14	0.14

COICOP 12.1.3.2

MÉTHODE/TYPE		MOVE	WINDOW	HALF	MEAN	FBEW	FBMW
GEKS	moyenne	-0.22	-0.22	-0.24	-0.23	-0.21	-0.18
	(écart-type)	(0.17)	(0.17)	(0.19)	(0.18)	(0.21)	(0.18)
	Indice final	0.00	-0.03	0.12	0.05	0.04	0.04
GK	moyenne	-1.04	-0.40	-0.67	-0.68	-0.81	-0.84
	(écart-type)	(0.53)	(0.30)	(0.26)	(0.27)	(0.27)	(0.27)
	Indice final	-1.56	0.24	-0.59	-0.63	-0.95	-0.95
TPD	moyenne	-0.70	-0.47	-0.50	-0.53	-0.61	-0.61
	(écart-type)	(0.34)	(0.21)	(0.20)	(0.22)	(0.24)	(0.23)
	Indice final	-0.94	-0.35	-0.51	-0.59	-0.86	-0.86

Pour tous les groupes COICOP, l'indice GEKS est celui qui présente le plus petit écart entre les options de couplage; en fait, l'inflation annuelle moyenne ne changerait guère si elle était arrondie à un dixième de point de pourcentage. Les petits écarts entre les différentes options de couplage pour la méthode GEKS sont à peine visibles dans la figure ci-dessous. Seuls les résultats pour COICOP 01.1 sont présentés (les résultats sont similaires pour les autres groupes).

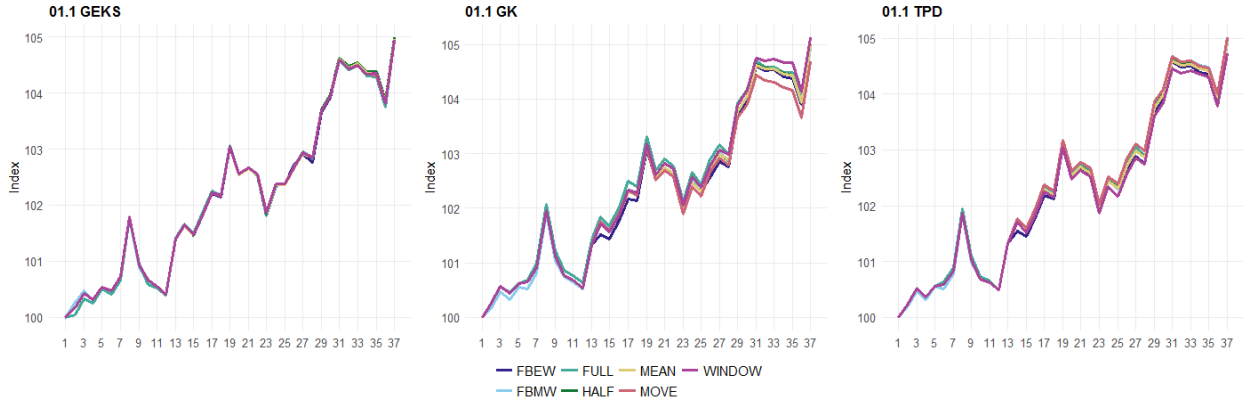


Figure 3 : Écarts entre les différentes méthodes de splicing et d'extension

L'indice Geary-Khamis semble présenter le plus grand écart entre les différentes options, la plus grande différence se situant entre le window splice et le movement splice. Ce n'est pas tout à fait surprenant, étant donné qu'il s'agit des options de splicing les plus extrêmes (début ou fin d'une période). En outre, la différence semble être la plus significative pour le groupe COICOP 12.1.3.2. Pour ce groupe de produits, l'indice fluctue également d'une période à l'autre et est donc couplé à des valeurs "extrêmes". L'indice TPD se trouve en grande partie entre les indices GEKS et Geary-Khamis pour les différentes options.

5.3. Impact du dumping

La méthode dynamique utilise un filtre de dumping pour éviter un biais négatif, sans quoi ce biais surviendrait étant donné que les produits en fin de série quittent la gamme de produits à des prix fortement réduits et n'y reviennent plus. Les niveaux de l'indice seraient dès lors inférieurs à la réalité.

Afin de tester l'impact du dumping sur les méthodes multilatérales, deux bases de données ont été créées: 1 avec des filtres de dumping et 1 sans filtre de dumping. Contrairement à la méthode dynamique, dans le cas des méthodes multilatérales, on n'observe aucun impact à long terme au niveau agrégé. La figure ne montre que le résultat de la méthode GEKS pour le groupe COICOP 01.1. Les résultats sont identiques pour les autres méthodes et les autres groupes COICOP.

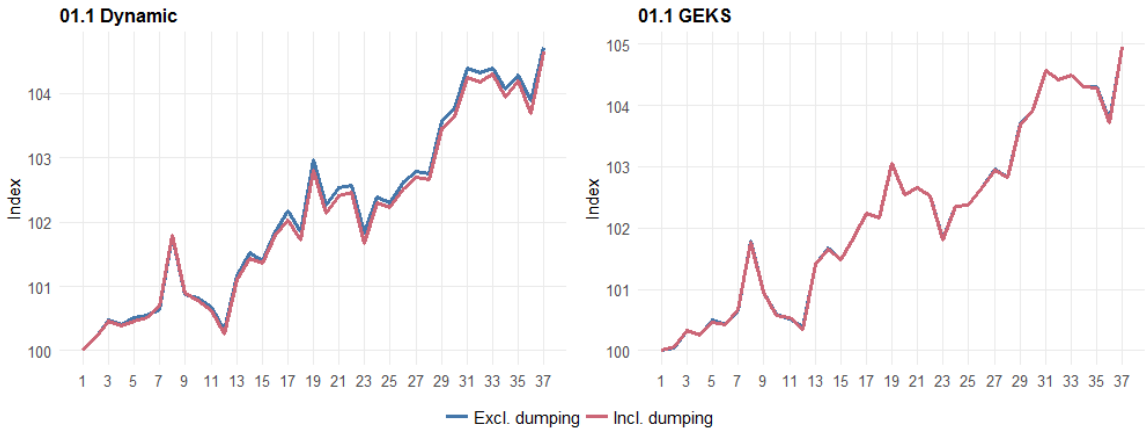


Figure 4 : Effet du dumping sur la méthode dynamique et la méthode GEKS pour COICOP 01.1.

Cependant, aux niveaux inférieurs, l'impact des filtres de dumping a été constaté lorsque la méthode GEKS est utilisée. Les recherches d'autres auteurs ont obtenu des résultats similaires (par exemple ABS (2016) & Chessa, Verburg et Willenborg (2017)).

Le graphique ci-dessous explique les différences pour les méthodes Geary-Khamis et TPD par rapport à la méthode GEKS pour les lames de rasoir (la différence est égale à GK/TPD moins GEKS). À gauche, les filtres de dumping sont activés (*les prix de*

dumping sont donc exclus) et l'écart varie entre +1 et -1. Sans les filtres de dumping (*les prix de dumping sont bien repris dans l'ensemble de données*), nous pouvons remarquer de plus grands écarts pour les périodes 29 à 31. Cela est dû au fait que l'indice GEKS est plus sensible au dumping et par conséquent, diminue plus que les autres indices. Toutefois, lorsque les méthodes multilatérales seront introduites, les filtres de dumping seront utilisés, conformément à la méthode de travail générale d'exclure les prix de dumping dans l'IPC.

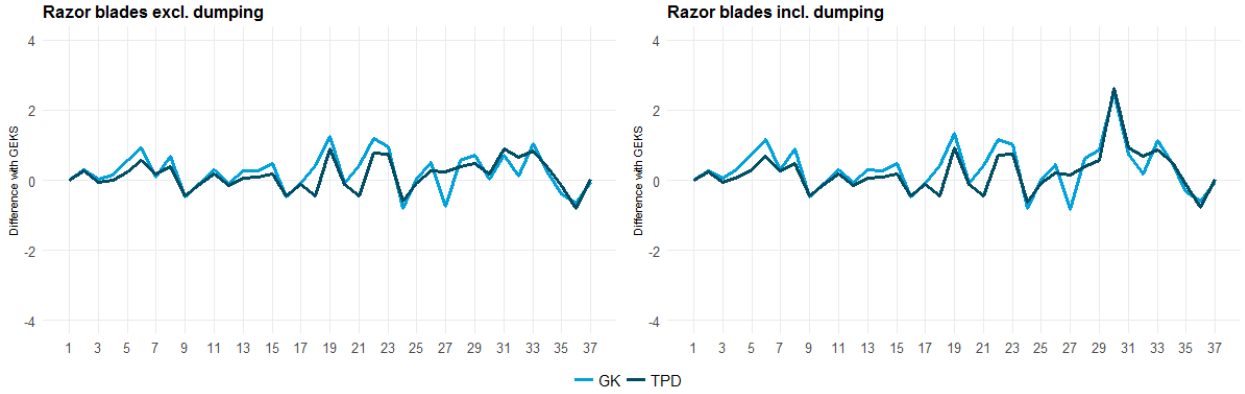


Figure 5: Effet du dumping entre Geary-Khamis/TPD et GEKS pour les lames de rasoir

5.4. Impact des relances de produits

Nous avons déjà expliqué ci-dessus que tant la méthode dynamique actuelle que les méthodes multilatérales tiennent compte des relances de produits en reliant les codes de produits. L'effet de ce lien semble important pour toutes les méthodes (dynamique et multilatérales) et entraîne un effet haussier. La figure ci-dessous présente un exemple pour le groupe COICOP 01.1.

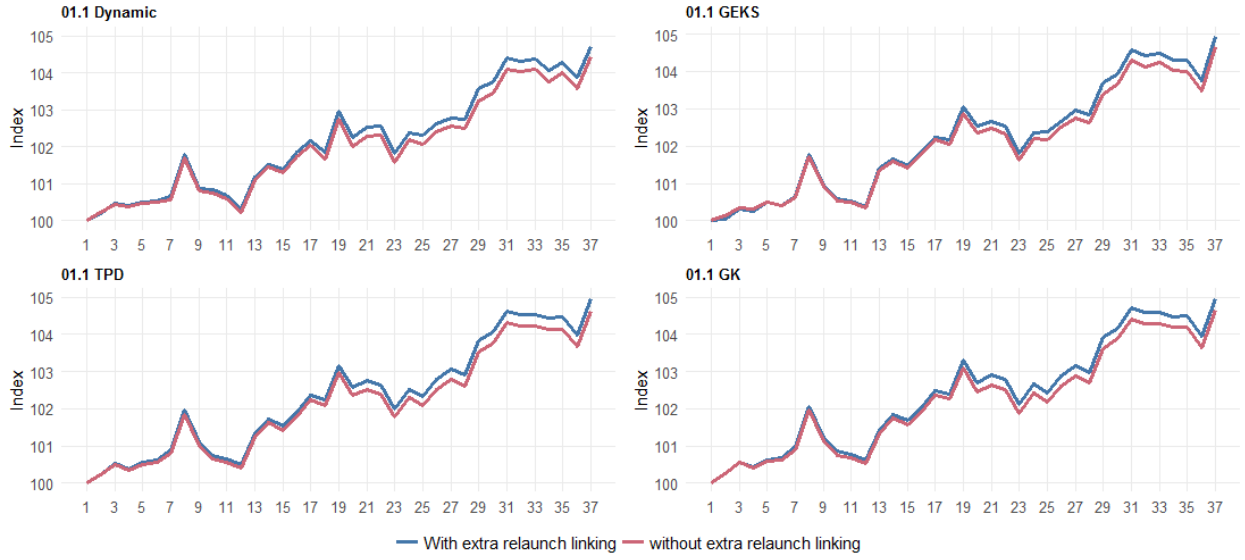


Figure 6: Effet des relances pour le groupe 01.1. avec les différentes méthodes

6. CONCLUSION

Les bons résultats de la méthode dynamique dans la recherche empirique soulèvent la question de savoir pourquoi la méthode de calcul devrait changer en faveur d'une méthode multilatérale. Le recours à une méthode multilatérale permettrait d'utiliser explicitement toutes les informations disponibles sur le chiffre d'affaires. Par ailleurs, de plus grandes différences ont été observées à court terme aux niveaux inférieurs (non agrégés). En outre, l'utilisation d'une méthode multilatérale est plus efficace parce que plus aucun échantillon n'est nécessaire et que moins de contrôles sont nécessaires. Dans la méthode dynamique, un nouvel échantillon est établi chaque mois, et le couplage des relances et des contrôles supplémentaires doivent toujours être effectués pour les produits qui disparaissent de l'échantillon ou les nouveaux produits

qui apparaissent dans l'échantillon. Avec une méthode multilatérale, on ne doit se pencher que sur les produits qui quittent ou qui entrent effectivement sur le marché.

Naturellement, il faudra décider quelle méthode multilatérale sera utilisée. Toutes les méthodes (à l'exception de la méthode de Lehr) et options de splicing testées donnent des résultats relativement similaires (le window splice et movement splice entraînent pour certaines méthodes une faible dérive), le choix sera donc plutôt théorique qu'empirique. Au niveau européen également, des discussions ont été lancées sur la (les) méthode(s) la (les) plus conforme(s) à la réglementation sur l'IPCH. La préférence interne actuelle va à la méthode GEKS combinée à une fixed based moving window de 13 mois (décembre comme mois fixe). La méthode GEKS est conforme aux calculs "classiques" de l'indice et est également plus facile à comprendre. La fixed based moving window est conforme à la pratique actuelle de donner un sens spécifique au mois de décembre (contrairement aux options de splicing). Par rapport à la fixed base enlarging window, elle présente l'avantage que les indices utilisent davantage d'informations au cours des premiers mois et sont donc plus stables. Des recherches complémentaires avec notamment des données provenant d'un second détaillant sont prévues en 2019. La mise en œuvre finale d'une méthode multilatérale devrait se faire en janvier 2020.

BIBLIOGRAPHIE

(ABS) Australian Bureau of Statistics (2016) Making Greater Use of Transactions Data to Compile the Consumer Price Index. - *Information Paper 6401.0.60.003, 29 November 2016, Canberra*.

Chessa A. (2016) A new methodology for processing scanner data in the Dutch CPI. - *Eurostat review of National Accounts and Macroeconomic Indicators, 1/2016, 49-69*.

Chessa A.G., Verburg, J., Willenborg, L. (2017) A Comparison of Price Index Methods for Scanner Data. - *Paper presented at the 15th Ottawa Group meeting, 10-12 May 2017, Elville am Rhein, Germany*.

de Haan J. (2015) A Framework for Large Scale Use of Scanner Data in the Dutch CPI. - *Paper presented at the 14th Ottawa Group meeting, Tokyo, Japan*.

de Haan J. (2010) Hedonic Price Indexes: A Comparison of Imputation, Time Dummy and 'Re-Pricing' Methods. -*Journal of Economics and Statistics 230, 772-791*.

de Haan J., Krsinich, F. (2014) Scanner Data and the Treatment of Quality Change in Non-Revisable Price Indexes. - *Journal of Business and Economic Statistics 32(3), 341-358*.

de Haan J., van der Grient, H. (2011) Eliminating chain drift in price indexes based on scanner data. - *Journal of Econometrics 161, 36-46*.

Diewert E.W., Fox, K.J. (2017) Substitution Bias in Multilateral Methods for CPI Construction using Scanner Data. - *Discussion Paper No. 17-02, Vancouver School of Economics, University of British Columbia*.

Fox K. (2017) Comment on "Making greater use of transactions data to compile the consumer price index, Australia".

Ivancic L., Diewert, E. W., Fox, K.J. (2011) Scanner data, time aggregation and the construction of price indexes. - *Journal of Econometrics 161, 24-35*.

Krsinich F. (2016) The FEWS Index: Fixed Effects with a Window Splice. - *Journal of Official Statistics 32 (2), 375-404*.

Lamboray C. (2017) The Geary Khamis index and the Lehr index: how much do they differ? - *Paper presented at the 15th Ottawa Group meeting, 10-12 May 2017, Elville am Rhein, Germany*.

Roels D., Van Loon K. (2017), Het gebruik van scannerdata van supermarkten in de consumptieprijsindex.

Von Auer L. (2017) Processing scanner data by an augmented GUV index. - *Eurostat review of National Accounts and Macroeconomic Indicators, 1/2017, 73-91*.

À PROPOS DE STATBEL

Statbel, l'office belge de statistique, collecte, produit et diffuse des chiffres fiables et pertinents sur l'économie, la société et le territoire belges.

La collecte s'effectue à l'aide de sources de données administratives et d'enquêtes. La production est réalisée de manière qualitative et scientifique. Les statistiques sont diffusées en temps opportun et de manière conviviale.

Statbel garantit que, d'une part, la vie privée et les données confidentielles sont protégées et que, d'autre part, les données sont utilisées à des fins exclusivement statistiques.

Visitez notre site internet
www.statbel.fgov.be

ou contactez-nous
e-mail: statbel@economie.fgov.be

Statbel (Direction générale Statistique - Statistics Belgium)
North Gate - Boulevard du Roi Albert II, 16, 1000 Bruxelles
E-mail: statbel@economie.fgov.be

Numéro d'entreprise
0314.595.348

Editeur responsable
Nicolas Waeyaert

North Gate
Boulevard du Roi Albert II, 16
1000 Bruxelles