Comparaison des indices de prix des vêtements et des chaussures à partir de données de caisse et de données moissonnées sur le Web

Comparing Price Indices of Clothing and Footwear for Scanner Data and Web Scraped Data

Antonio G. Chessa* et Robert Griffioen**

Résumé – Pour collecter les prix des biens de consommation, les instituts de statistique envisagent le recours à des données moissonnées sur le Web comme une alternative possible aux données de caisse. Les données de transaction étant rares, il est naturel de questionner la pertinence des données du Web pour le calcul d'indices de prix. On propose ici de comparer les indices de prix obtenus à partir de données du Web ou de données de caisse pour des vêtements et chaussures vendus par un même magasin en ligne. Les prix constatés en caisse et moissonnés sur le Web sont souvent égaux, bien que ceux du Web soient légèrement supérieurs en moyenne. Le nombre de produits dont les prix sont moissonnés sur le Web est très corrélé au nombre de produits vendus. Compte tenu du taux de renouvellement élevé des articles dans le secteur de l'habillement, une méthode multilatérale (celle de Geary-Khamis) a été utilisée pour calculer les indices de prix. Pour 16 catégories de produits, les indices montrent de légers écarts globaux entre les deux sources de données : les indices en glissement annuel ne diffèrent que de 0.3 point de pourcentage au niveau de la nomenclature COICOP (vêtements pour hommes et pour femmes). Reste à savoir si ces résultats prometteurs pour les données moissonnées sur le Web se confirmeront pour d'autres points de vente.

Abstract *– Statistical institutes are considering web scraping of online prices of consumer goods as a feasible alternative to scanner data. The lack of transaction data generates the question whether web scraped data are suited for price index calculation. This article investigates this question by comparing price indices based on web scraped and scanner data for clothing and footwear in the same webshop. Scanner data and web scraped prices are often equal, with the latter being slightly higher on average. Numbers of web scraped product prices and products sold show remarkably high correlations. Given the high churn rates of clothing products, a multilateral method (Geary-Khamis) was used to calculate price indices. For 16 product categories, the indices show small overall differences between the two data sources, with year on year indices differing only by 0.3 percentage point at COICOP level (men's and women's clothing). It remains to be investigated whether such promising results for web scraped data will also be found for other retailers.*

Codes JEL / JEL Classification : C43, E31
Mots-clés : IPC, données de caisse, moissonnage du Web, méthodes multilatérales, méthode Geary-Khamis
Keywords: CPI, scanner data, web scraping, multilateral methods, Geary Khamis method

** Centraal Bureau voor de Statistiek (CBS), équipe IPC (auteur correspondant ag.chessa@cbs.nl)*
*** Centraal Bureau voor de Statistiek, équipe IPC au moment où ces recherches ont été menées*

Les auteurs tiennent à remercier Eurostat pour la subvention qui leur a été accordée pour mener cette recherche.

Reçu le 31 juillet 2017, accepté après révisions le 1er avril 2019
Traduit de la version originale en anglais

Pour citer cet article : Chessa, A. G. & Griffioen, R. (2019). Comparing Price Indices of Clothing and Footwear for Scanner Data and Web Scraped Data. *Economie et Statistique / Economics and Statistics*, 509, 49–68. https://doi.org/10.24187/ecostat.2019.509.1984

Les données de caisse sont de plus en plus utilisées pour mesurer l'indice des prix à la consommation (IPC). En effet, elles sont quasiment idéales pour remplacer les données des enquêtes classiques, car elles contiennent des données de transaction. Les prix et les dépenses sont connus pour chaque article vendu, grâce au code-barres (le *Global Trade Item Number*, ou GTIN, émis et géré par la société internationale GS1). Les dépenses correspondant à chaque article, obtenues grâce aux données de caisse, peuvent servir à établir des indices de prix pondérés, ce qui leur confère un avantage conséquent sur les données tirées des enquêtes.

En Europe, jusqu'en 2014, quatre instituts nationaux de statistique (INS) utilisaient les données de caisse dans leur IPC. En janvier 2018, ils étaient dix à le faire (voir également Leclair *et al.*, ce numéro). Bien que les INS soient habilités à développer leur propre méthode de traitement des données de caisse et de calcul des indices de prix pour les agrégats élémentaires, il est néanmoins souhaitable que les méthodes utilisées dans différents pays soient comparables pour ces agrégats. Dans cette optique, afin d'encourager les INS à commencer à traiter les données de caisse, Eurostat a émis des directives et établi une liste descriptive des pratiques actuelles (Eurostat, 2017).

La collecte des données de caisse est un processus potentiellement long. Différents facteurs entrent en jeu. Par exemple, il faut identifier les personnes à contacter chez le détaillant, chercher à savoir si le détaillant est disposé à coopérer et le temps qu'il peut consacrer à préparer un jeu de données dans un format que l'institut de statistique pourra utiliser. Dans plusieurs pays, comme par exemple aux Pays-Bas, une loi sur la statistique peut être invoquée pour demander des données de caisse. En revanche, dans les pays dépourvus d'une telle loi, il peut être difficile d'obtenir ce type de données et les INS se concentrent alors sur les données collectées en ligne (par exemple, voir Breton *et al.*, 2016). Le moissonnage du Web, dans le but de collecter les prix en ligne et des informations sur les caractéristiques des articles, est de plus en plus utilisé depuis quelques années (Breton *et al.*, 2016 ; Cavallo, 2016 ; Griffioen & ten Bosch, 2016) et offre de nouvelles perspectives pour la statistique publique. Comme avec les données de caisse, la taille des échantillons peut être considérablement augmentée, et la collecte et le traitement des données peuvent être largement automatisés. La collecte automatisée de données en ligne permet également de réduire la charge administrative liée à la collecte de prix, non seulement pour les INS mais aussi pour les détaillants. Pour cette raison, le remplacement des enquêtes par sondage par une collecte automatisée de données de prix en ligne représente une opportunité pour les instituts de statistique – mais aussi un défi de taille.

Compte tenu de la popularité grandissante du moissonnage du Web, il est important d'envisager les fonctionnalités et les limitations de l'utilisation des prix en ligne pour le calcul des indices de prix. Le moissonnage du Web ne permet de collecter que les prix en ligne, car les dépenses relatives aux articles proposés sur un site Web ne sont évidemment pas disponibles en ligne. Certes, c'est également le cas pour la collecte de prix classique. Toutefois, maintenant que les données de caisse sont disponibles, on sait quantifier l'effet d'informations présentes ou manquantes sur un indice de prix. Par exemple, les indices de prix peuvent largement différer selon que l'on utilise des poids basés sur les dépenses ou des poids égaux pour les produits dans une formule d'indice (Chessa *et al*., 2017)[1].

De telles différences débouchent sur une question importante : le nombre de prix de produits moissonnés sur le Web est-il bien corrélé au nombre de ventes dans les données de caisse ? Si la réponse est oui, les indices de prix exclusivement basés sur les quantités et les prix moissonnés sur le Web sont susceptibles de fournir une bonne approximation des indices de prix basés sur les données de caisse. Bien sûr, le résultat dépend de plusieurs facteurs, comme la politique pratiquée par les magasins en ligne, la conception de leur site Web (par exemple, quels produits sont mis en avant et apparaissent le plus souvent sur un site) et la stratégie de moissonnage (le site entier est-il moissonné, et si oui à quelle fréquence et à quel moment ?). Certes, une comparaison entre des indices de prix basés sur des données de caisse ou basés sur des données moissonnées n'est raisonnable que si les mêmes métadonnées relatives aux articles peuvent être utilisées dans le calcul des indices concernés.

1. Nous utilisons le mot « produit » en tant que concept générique et le mot « article » en référence au GTIN. Un produit est équivalent à un article lorsque les GTIN affichent un taux de renouvellement peu élevé, c'est-à-dire lorsque les assortiments restent stables au fil du temps. Si les assortiments ne sont pas stables, par exemple lorsque les GTIN sont de courte durée en raison de relances, les GTIN doivent être reliés et classés en différents groupes. Les GTIN de chaque groupe ont les mêmes caractéristiques d'article. Nous appelons ces groupes des « produits ». La façon dont les caractéristiques sont sélectionnées, ainsi que le fait pour les GTIN d'être considérés comme des produits ou non, sont des questions complexes qui mériteraient une étude séparée.

Le Centraal Bureau voor de Statistiek (CBS) reçoit des données de caisse de la part d'un grand magasin en ligne néerlandais depuis plusieurs années. En octobre 2012, CBS a commencé à moissonner des prix en ligne et des métadonnées du même magasin. Disposer de données de caisse et de données du Web fournit donc une excellente opportunité de comparer les prix des produits, les quantités et les indices de prix entre ces deux sources de données. Les indices de prix calculés avec des données de caisse peuvent servir de référence pour évaluer l'exactitude des indices de prix calculés avec des données moissonnées. Cet article vise à comparer les indices de prix basés sur ces deux sources de données.

La suite s'articule comme suit. La prochaine section décrit brièvement les informations contenues dans les données de caisse et les données moissonnées sur le Web du magasin en ligne néerlandais. Dans la section suivante, nous décrivons la méthode appliquée aux données de caisse et aux données du Web, que nous appelons la « méthode Q-U » de l'anglais *Quality-adjusted Unit value*, c'est-à-dire la valeur unitaire ajustée en fonction de la qualité. Les indices de prix calculés sur la base des deux sources de données sont ensuite comparés au niveau de la catégorie et de la nomenclature COICOP[2]. Nous présentons enfin les principales conclusions de cette étude, ainsi que quelques suggestions de recherches complémentaires.

Données de caisse et données moissonnées pour un magasin en ligne néerlandais

Durant les premières années du programme de développement du moissonnage de données du Web, qui a débuté il y a plus de cinq ans, CBS s'est concentré sur les vêtements et les chaussures, dans le but de moins utiliser les enquêtes classiques pour ces catégories de produits au sein de son IPC. En conséquence, la comparaison des prix, des quantités et des indices de prix entre les données moissonnées et les données de caisse portera principalement sur les vêtements et les chaussures. Les résultats d'une analyse statistique des quantités et des prix des produits sont également présentés.

Données de caisse

CBS reçoit des données de caisse d'un grand magasin en ligne néerlandais depuis janvier 2011. Le détaillant spécifie et envoie les données une fois par semaine, et cet accord a également été conclu avec d'autres détaillants. Les données de caisse couvrent les transactions effectuées sur la totalité de l'assortiment du magasin. L'assortiment est extrêmement varié : hormis des vêtements et des chaussures, le magasin vend des appareils électroniques, des articles de maison et de jardin, des produits destinés aux loisirs, etc.

Pour chaque article (GTIN), les jeux de données de caisse contiennent les informations suivantes, qui sont communiquées en tant que champs distincts :

- année et semaine des ventes (informations groupées en un seul champ) ;
- GTIN ;
- numéro de l'article (code à 6 chiffres spécifique au détaillant pour chaque article);
- chaîne de texte contenant une (courte) description de l'article ;
- groupe dans lequel l'article est classé par le détaillant ;
- numéro du groupe ;
- nombre d'articles vendus ;
- chiffre d'affaires (dépenses) ;
- nombre d'articles retournés ;
- chiffre d'affaires des articles retournés ;
- TVA.

Depuis la fin 2013, le nombre d'articles retournés et le chiffre d'affaires correspondant sont également inclus dans les données par le détaillant, et sont communiqués chaque semaine depuis mars 2014. La valeur des articles retournés est déduite des champs « nombre d'articles vendus » et « chiffre d'affaires », de sorte que cette somme est une valeur nette. Pour cette raison, les champs « nombre d'articles vendus » et « chiffre d'affaires » peuvent présenter des valeurs négatives si le nombre d'articles retournés et le chiffre d'affaires correspondant sont supérieurs au nombre d'articles initialement vendus et au chiffre d'affaires correspondant.

Données moissonnées sur le Web

Certains types de produits, comme l'habillement, peuvent présenter un taux de renouvellement

2. Cela correspond aux catégories « Vêtements pour hommes » et « Vêtements pour femmes » de la nomenclature COICOP.

élevé. Les nouveaux articles doivent être reliés aux articles sortants de qualité identique ou similaire, afin que les variations de prix « cachées » soient prises en compte lors du calcul des indices de prix. Ce remplacement d'article prend également le nom de « relance ». Les articles peuvent être reliés en fonction de caractéristiques communes. Pour cette raison, il est important que les jeux de données de caisse contiennent des informations sur ces caractéristiques.

Toutefois, les instituts de statistique dépendent de ce que les détaillants peuvent leur communiquer, de sorte que les métadonnées incluses dans les données de caisse peuvent ne pas suffire à relier les articles. Malheureusement, c'est le cas pour les données de caisse du magasin en ligne traité dans cet article (voir plus bas dans cette section). Les instituts de statistique peuvent contacter les détaillants pour leur demander de plus amples informations. Le moissonnage du Web est une alternative intéressante pour compléter les informations relatives aux articles présentes dans les données de caisse.

L'outil de moissonnage du Web élaboré pour le magasin en ligne néerlandais collecte des données chaque jour depuis son lancement le 6 octobre 2012. Les informations suivantes sont collectées pour chaque article :

- année, mois et jour durant laquelle/lequel les données ont été moissonnées (un seul champ) ;

- numéro d'article spécifique au détaillant ;

- description de l'article ;

- nom de la marque ;

- trois niveaux de classification de l'article ;

- prix de l'article ;

- prix habituel de l'article.

La description de l'article collectée sur le Web contient plus d'informations que celle fournie dans les données de caisse (dans celles-ci, la description de l'article est souvent, par exemple, « Pantalons pour hommes »). Par ailleurs, les chaînes de texte moissonnées contiennent le nom de la marque et le contenu du lot (par exemple le nombre d'articles uniques dans un lot comprenant plusieurs articles), et la taille, le tissu et le style sont précisés pour certains vêtements. Le nom de la marque est également indiqué dans un champ distinct.

Sur le site Web, on peut aller au niveau de l'article à partir du menu principal, en suivant deux menus secondaires, de sorte que les articles sont classés en fonction de trois niveaux de groupe. Comme mentionné plus haut, l'assortiment du magasin en ligne est assez varié. Le moissonnage vise en priorité à collecter des informations sur les vêtements et les chaussures. Les trois niveaux de classification de l'article s'appliquant aux vêtements et aux chaussures se résument comme suit :

- le niveau supérieur (menu principal) divise les vêtements et les chaussures en cinq groupes, à savoir « Vêtements pour hommes », « Vêtements pour femmes », « Vêtements pour enfants », « Haut de gamme » et « En promotion ». Dans cet article, nous appelons ce niveau supérieur le « groupe principal » ;

- le niveau intermédiaire est appelé « catégorie ». L'outil de moissonnage a collecté des informations relatives à 145 catégories durant la période couverte par cette étude (c'est-à-dire de mars 2014 à décembre 2016) ;

- le niveau le plus détaillé est appelé « type », et contient 1 131 groupes.

Les groupes principaux « Haut de gamme » et « En promotion » peuvent contenir des articles dont le prix est réduit. En conséquence, un article peut être atteint à partir du groupe principal « En promotion » ou à partir de l'un des trois groupes principaux « Vêtements pour hommes », « Vêtements pour femmes » ou « Vêtements pour enfants ». L'outil de moissonnage « navigue » parmi ces cinq groupes principaux, de sorte que chaque article peut être moissonné plus d'une fois par jour. Les articles moissonnés plus d'une fois sont comptabilisés à chaque fois.

Il va de soi que le groupe « En promotion » ne contient pas seulement des vêtements et des chaussures mais aussi d'autres articles en promotion. Pour cette raison, l'outil de moissonnage collecte également les informations susmentionnées pour les appareils électroniques, les articles de maison et de jardin, les produits de beauté et de soin, etc. Les données moissonnées contiennent deux prix pour les articles en promotion : le prix réduit de l'article et son prix habituel. Le prix habituel d'un article en promotion est collecté en même temps que le prix réduit ; il correspond au prix en vigueur juste avant la période de promotion. Pour nos calculs d'indices, nous utilisons bien sûr les prix réduits – et non pas les prix habituels – pour les articles en promotion.

Analyse statistique des données de caisse et du Web

Dans cette sous-section, nous analysons plusieurs aspects des données de caisse et des données moissonnées ayant un impact direct sur le calcul des indices de prix. Notre priorité est de comparer les prix calculés à partir de ces deux sources. Les quantités vendues servent à calculer la valeur unitaire des produits et, associées aux prix, permettent d'établir le poids des produits. Dans ce contexte, il convient donc également de se demander comment les quantités vendues peuvent être comparées aux nombres de prix de produits moissonnés sur le Web.

Propriétés des deux jeux de données

Avant d'utiliser de grands jeux de données numériques dans l'IPC ou à des fins de recherche, une première étape clé consiste à effectuer plusieurs vérifications. Les articles sur la qualité des données publiés par Daas & van Nederpelt (2010) et Daas & Ossen (2010) proposent plusieurs « dimensions de qualité » pouvant servir à vérifier les données. Nous résumons ci-dessous nos conclusions sur certaines des dimensions que nous avons analysées dans le cadre des données de caisse et des données du Web.

- Exhaustivité : les variables (c'est-à-dire les colonnes ou les champs) des deux jeux de données présentent un degré d'exhaustivité élevé. Tous les enregistrements des données de caisse sont consignés, à l'exception du code GTIN qui présente un pourcentage élevé de valeurs manquantes (46.4 %). La raison de ce grand nombre de valeurs manquantes est inconnue, mais pourrait être liée au fait que le détaillant a ses propres codes produit, qui sont disponibles pour chaque enregistrement. Une description du produit est également disponible pour chaque enregistrement. Les données moissonnées sur le Web présentent, elles aussi, un degré d'exhaustivité élevé. Le prix et la description du produit manquent dans 21 enregistrements, ce qui est négligeable sur un total de plusieurs millions d'enregistrements.

- Stabilité : la stabilité est un autre facteur essentiel devant être vérifié avant d'utiliser un jeu de données à des fins de production statistique régulière. La production de l'IPC est entravée si, au cours d'un mois donné, le nombre total d'enregistrements semble largement inférieur à la normale. Ni les données de caisse ni les données moissonnées ne reflètent les augmentations ou diminutions rapides du nombre total d'enregistrements mensuels. Le nombre d'enregistrements augmente au fil du temps, ce qui s'explique par l'accroissement de l'assortiment.

- Niveau de détail : le volume de métadonnées incluses dans les données de caisse du magasin en ligne est limité ; ainsi, à titre indicatif 25 % des descriptions d'articles ne contiennent qu'un seul mot et 62 % ne comprennent pas plus de deux mots.

L'outil de moissonnage a collecté des informations relatives à 385 833 articles entre mars 2014 et décembre 2016. Ce chiffre est proche de celui de 407 253 articles vendus indiqué dans les données de caisse, bien que ces dernières couvrent la totalité de l'assortiment (contrairement aux données moissonnées sur le Web). Le grand nombre d'articles moissonnés découle, d'une part, du fait que l'outil de moissonnage collecte également des informations sur des articles ne faisant pas partie du secteur de l'habillement dans les groupes « Haut de gamme » et « En promotion » et, d'autre part, du fait que le site Web peut également comporter des articles non vendus.

En combinant le nom de la marque avec les trois niveaux de classification de l'article afin de regrouper ou relier les articles, les 385 833 articles moissonnés sur le Web se retrouvent divisés en 59 588 groupes d'articles. Le rapport nombre d'articles/groupes d'articles est donc assez faible. Il est beaucoup plus faible qu'avec les données de caisse (1 635 groupes pour 407 253 articles), ce qui indique le plus grand niveau de détail des métadonnées collectées par l'outil de moissonnage. Cela favorise l'homogénéité des produits lorsque les caractéristiques du produit sont utilisées pour définir le produit.

- Respect des délais : CBS reçoit des données de caisse une fois par semaine, habituellement dans les délais impartis, de tous les détaillants. L'outil de moissonnage du Web collecte des données une fois par jour, au cours de la nuit afin de pas occasionner de gêne durant les heures de pointe sur le site du magasin. Les données sont disponibles dès qu'elles ont été collectées. Toutefois, dans certaines circonstances, les délais peuvent ne pas être respectés, par exemple, si un site Web est indisponible ou a changé. À notre connaissance, les sites Web sont rarement indisponibles. En revanche, les changements de site sont plus fréquents, et c'est pourquoi nous avons créé une équipe « DevOps » dédiée au développement et aux opérations, afin

d'adapter l'outil de moissonnage et d'assurer son fonctionnement continu (pour plus de détails sur sa mise en œuvre à CBS, voir Griffioen *et al.,* 2016).

Comparaisons de prix

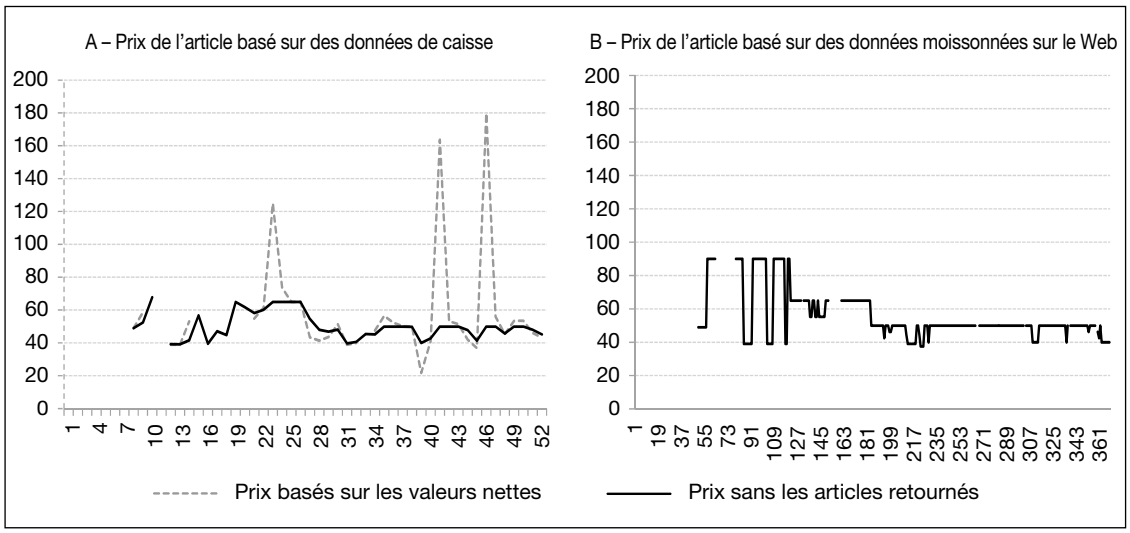
Il convient de noter que les données de caisse, d'une part, permettent de calculer les prix des transactions (c'est-à-dire les prix réellement payés par les consommateurs) et, d'autre part, peuvent avoir des composantes différentes, comme par exemple des réductions spéciales pour les titulaires de cartes ou de bons de réduction. Cela n'est pas le cas pour les prix moissonnés sur le Web, qui ne sont pas les prix des transactions mais les prix offerts par le détaillant sur son site Web.

Le prix d'un ensemble de transactions différentes sur un même article, ou sur des articles de même qualité, peut être calculé en tant que valeur unitaire : il s'agit du ratio des dépenses totales divisées par la somme des quantités vendues (ILO *et al.*, 2004, p. xxii). Ce calcul est habituellement très simple, mais des complications peuvent néanmoins survenir si les consommateurs retournent fréquemment certains articles. Le magasin en ligne propose une politique de retour favorable à ses clients, leur permettant de renvoyer leurs articles gratuitement dans un délai de 14 jours à compter de la livraison.

Les quantités retournées et les dépenses correspondantes sont déduites des quantités vendues et des dépenses durant la semaine pendant laquelle les articles sont retournés et traités par un détaillant. Les quantités vendues et les dépenses représentent donc une valeur nette dans les données de caisse. La semaine durant laquelle le retour est traité peut être différente de celle durant laquelle l'article a été acheté. Cela a deux conséquences importantes : premièrement, les dépenses et les quantités nettes peuvent être négatives ; deuxièmement, la valeur unitaire tirée des deux valeurs nettes diffère du prix initialement payé si le prix d'achat de l'article est différent du prix en vigueur pendant la semaine durant laquelle l'article est retourné. En outre, les consommateurs tendent à acheter plus d'un article lorsqu'il est en promotion. Pour cette raison, les premières semaines suivant une promotion doivent faire l'objet d'une attention particulière lors de la comparaison des prix basés sur les données de caisse avec ceux basés sur les données moissonnées sur le Web.

Lors de toute demande de données de caisse, CBS demande des informations distinctes sur les quantités retournées et sur les dépenses correspondantes. Les données de caisse du grand magasin en ligne néerlandais contiennent ces informations depuis la 12e semaine de l'année 2014. Nous pouvons donc quantifier l'impact des retours d'articles sur les dépenses nettes, les quantités vendues et les valeurs unitaires.

Figure I
Prix hebdomadaires basés sur des données de caisse et prix quotidiens basés sur des données du Web pour un même article (« jeans pour hommes ») en 2015



Note : deux calculs de prix sont indiqués pour les données de caisse : avec les articles retournés (c'est-à-dire sur la base des valeurs nettes) et sans ces articles retournés. Les prix sont exprimés en euros. L'axe horizontal indique le numéro de la semaine (données de caisse) et le jour (données moissonnées sur le Web).
Source : données de caisse pour les prix de vêtements (gauche) et prix moissonnés sur le Web (droite).

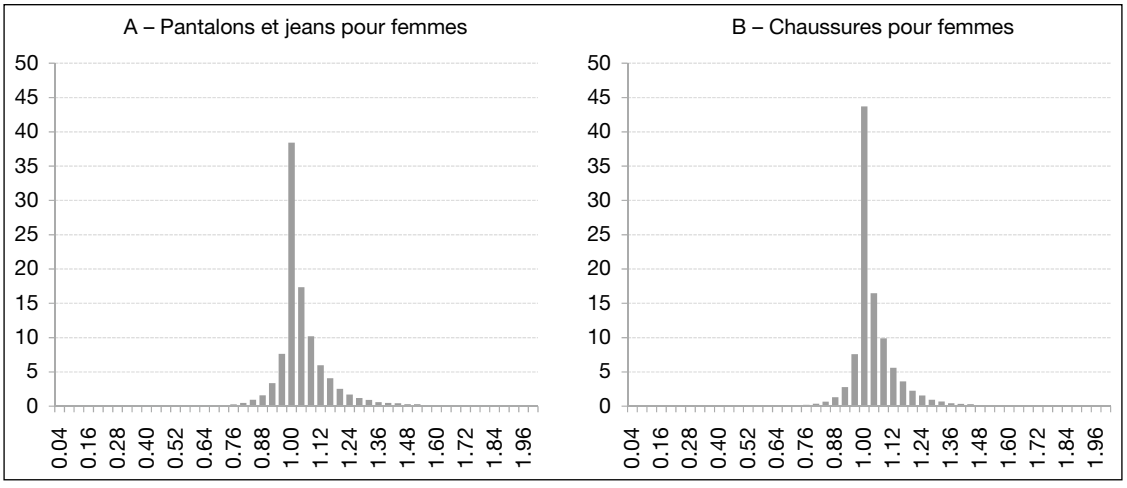
La figure I montre les prix basés sur les données de caisse et ceux basés sur les données du Web pour un même article durant une année entière. Les prix tirés des données de caisse (figure I-A) incluent les retours d'articles, c'est-à-dire que les quantités et les dépenses liées aux articles retournés sont déduites du chiffre d'affaires des semaines durant lesquelles les articles ont été retournés afin de parvenir à la valeur nette. Les prix ont été calculés uniquement si les quantités et les dépenses nettes sont supérieures à zéro. Trois pics importants apparaissent. Chacun de ces pics fait suite à une semaine durant laquelle les prix ont baissé. Les valeurs unitaires calculées à partir des quantités et des dépenses nettes produisent des prix supérieurs aux prix en vigueur pendant la semaine durant laquelle les articles ont été retournés. Les pics de prix surviennent lorsque les quantités d'articles retournés sont proches des quantités vendues pendant la semaine durant laquelle les articles sont retournés.

En soustrayant ces valeurs, on peut calculer le « vrai » prix initial de la transaction (ligne continue sur le graphique de la figure I-A). Cela montre combien il est important de demander des informations séparées sur les dépenses et sur les quantités d'articles retournés. Les prix corrigés sont beaucoup plus proches des prix moissonnés sur le Web indiqués figure I-B. Les prix moissonnés sont plus élevés, en moyenne, durant les premières semaines (c'est-à-dire jusqu'à la 19e semaine ou la 109e journée sur la figure I-B). L'article s'est vendu pour la première fois durant la 8e semaine de 2015. Il semblerait qu'il ait été inclus dans l'assortiment à un prix élevé, mais la ligne continue du graphique I-A suggère que les consommateurs l'ont acheté principalement lorsqu'il était en promotion. Après la période initiale, les écarts entre les prix dans les deux jeux de données se réduisent.

Compte tenu de l'impact que les articles retournés peuvent avoir sur les quantités et les dépenses nettes, nous avons décidé d'exclure les articles retournés des quantités et des dépenses pour la comparaison avec les prix et les quantités moissonnés sur le Web. Nous avons calculé deux statistiques de base : d'une part, le rapport entre les prix moissonnés sur le Web et les prix basés sur les données de caisse et, d'autre part, la corrélation entre le nombre de produits vendus et le nombre de prix de produits moissonnés sur le Web au fil du temps. Nous calculons la corrélation car il est difficile de faire une comparaison bijective entre les nombres de produits vendus et les nombres de prix des produits moissonnés sur le Web.

Les histogrammes représentant ces rapports sont indiqués à la figure II pour les catégories combinées de vêtements « Pantalons et jeans » et « Chaussures » pour femmes. Nous avons combiné dans le même groupe les articles ayant la même marque et le type de classification de l'article le plus détaillé. Nous avons fait le même choix pour le calcul de l'indice de prix (*infra*). Les articles des groupes principaux « Haut de gamme » et « En promotion » ont également été inclus afin de tenir

Figure II
Distribution des fréquences des rapports entre les prix de produits moissonnés sur le Web et les valeurs unitaires pour les données de caisse, pour deux catégories de produits



Note : pour chaque graphique, la somme des fréquences est égale à 100 %. Les rapports de prix des axes horizontaux sont centrés sur les valeurs de classe, avec une amplitude de classe de 0.04.
Source : données de caisse et données moissonnées sur le Web pour des vêtements et des chaussures.

compte des prix réduits. Un groupe [marque×type] peut être, par exemple « Shorts en jean » de la marque X. Toute combinaison [marque×type] est appelée « produit » dans cet article.

Les graphiques de la figure II montrent les rapports de prix combinés de tous les produits au cours de chaque mois. Ils présentent des pics élevés aux alentours de 1 (prix égaux) et sont tous les deux orientés vers des rapports supérieurs à 1. Les prix moissonnés sur le Web tendent à être plus élevés, en moyenne, que les prix des transactions. Nous avons déjà constaté la même chose pour les prix d'un même article (cf. figure I). Le plus faible niveau des prix tirés des données de caisse peut s'expliquer par la réorientation des ventes vers des articles moins chers, par exemple lorsque ces articles sont en promotion (« effet quantité »).

Comparaisons de quantités

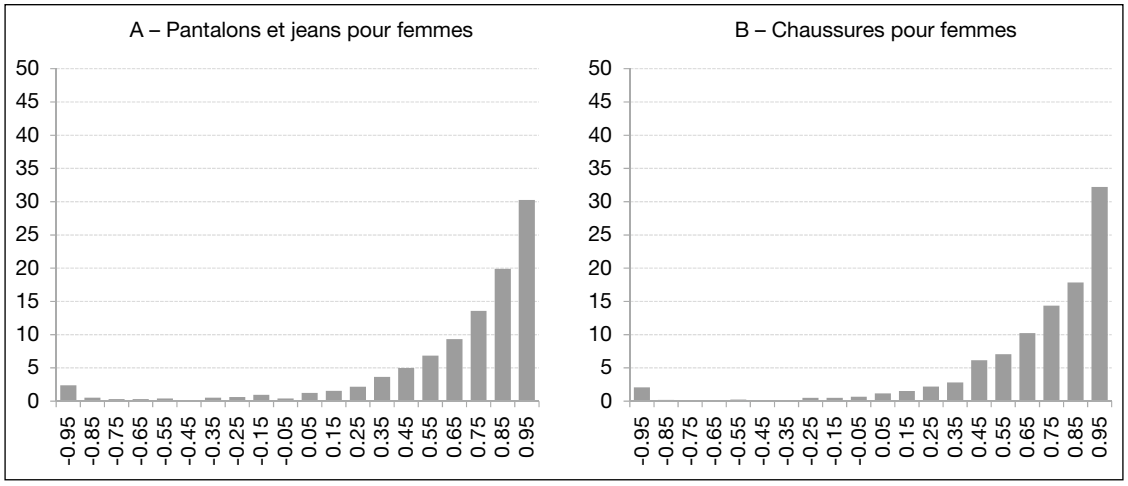
Nous avons calculé la corrélation entre le nombre de produits vendus et le nombre de prix de produits moissonnés sur le Web. Pour chaque produit, nous avons calculé la corrélation entre les couples de nombres vendus et de nombres de prix moissonnés pour tous les mois de la série. Les deux graphiques montrent des corrélations très élevées, et les fréquences les plus importantes s'observent pour les classes où la corrélation est la plus élevée (figure III). Ces tendances n'auraient pas été observées si les nombres moissonnés sur le Web étaient indépendants des nombres de produits vendus, ce qui engendrerait des distributions centrées sur une corrélation nulle. Les légères hausses constatées dans la classe où la corrélation est la plus faible s'expliquent dans une large mesure par des produits dont les prix ne sont observés que pendant deux mois. Si l'on exclut ces produits des calculs, ces hausses disparaissent.

Les fréquences auxquelles les articles peuvent être trouvés dans différents menus d'un site Web au fil du temps semblent correspondre assez bien aux quantités vendues. Cela découle de la politique du détaillant, qui consiste à promouvoir les articles les plus vendus du site. Les autres catégories de produits montrent des résultats semblables, tant en termes de prix que de quantités, ce qui crée des conditions favorables pour comparer les indices de prix basés sur les deux jeux de données. Il est donc important d'être en contact avec le détaillant afin d'obtenir des informations sur sa stratégie d'organisation de son site Web.

Dynamique de l'assortiment

Les vêtements et les chaussures se caractérisent habituellement par un taux de renouvellement élevé. Nous avons analysé la dynamique des assortiments de différentes catégories de produits pour les données de caisse et les données moissonnées sur le Web. Cette dynamique a été quantifiée en

Figure III
Distribution des fréquences des corrélations entre le nombre de prix de produits moissonnés sur le Web et le nombre de produits vendus



Note : pour chaque graphique, la somme des fréquences est égale à 100 %. Les corrélations des axes horizontaux sont centrées sur les valeurs de classe, avec une amplitude de classe de 0.1.
Source : données de caisse et données moissonnées sur le Web pour des vêtements et des chaussures.

introduisant trois mesures : (i) la part des produits vendus ou disponibles durant de longues périodes, c'est-à-dire « le flux », (ii) la part des produits ajoutés à un assortiment durant une année donnée, c'est-à-dire « les flux entrants », et (iii) la part des produits retirés d'un assortiment, c'est-à-dire « les flux sortants ». Nous avons calculé ces trois statistiques de flux en bilatéral, c'est-à-dire pour des groupes de deux mois. Le premier mois, choisi comme mois de référence, est resté fixe. Les produits vendus ou disponibles durant le mois de référence et durant le deuxième mois (mois en cours) sont comptabilisés dans le flux, les produits non vendus/disponibles le mois de base mais vendus/disponibles seulement durant le mois en cours sont comptabilisés dans les flux entrants et les produits disponibles le mois de référence mais non disponibles durant le mois en cours sont comptabilisés dans les flux sortants[3].

Les trois statistiques sont calculées pour chaque mois de la période allant de mars 2014 à décembre 2016, en utilisant respectivement les mois de mars 2014, décembre 2014 et décembre 2015 comme mois de référence. Les statistiques résultent de comptages au niveau du produit, c'est-à-dire pour les groupes [marque×type]. La figure IV montre les trois statistiques de flux pour les « pantalons et jeans » pour hommes.

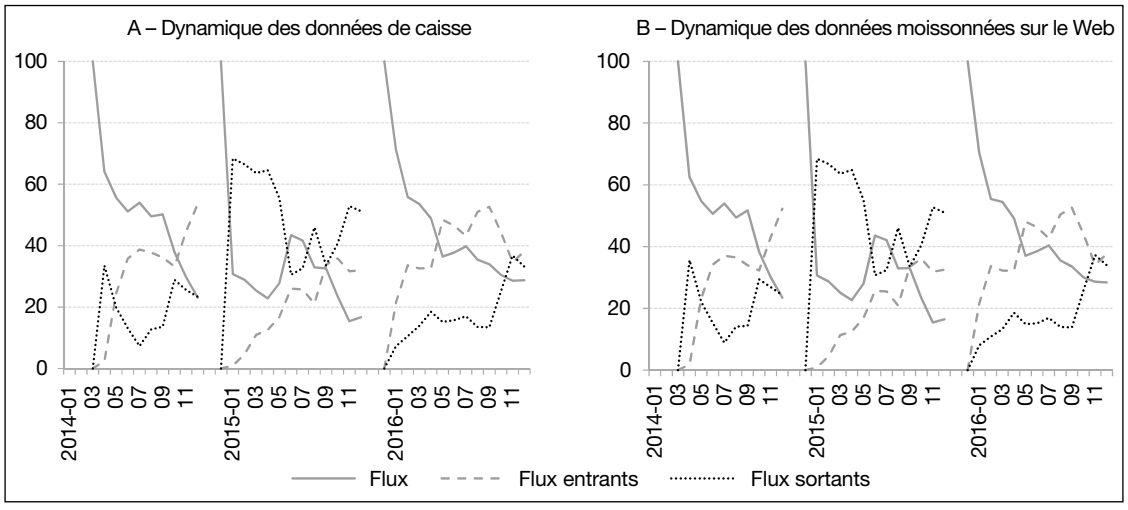
Le taux du flux est, par définition, égal à 100 % pour les mois de référence. La baisse rapide du taux du flux et la hausse, ainsi que le niveau élevé, des flux entrants indiquent un assortiment très dynamique. Les deux graphiques montrent clairement qu'il y a très peu de différence entre les statistiques de flux basées sur les données de caisse et sur les données du Web. Cela signifie que les articles qui ne se vendent plus sont rapidement retirés du site Web. On note également que la forte dynamique s'observe au niveau du produit, c'est-à-dire à un niveau moins détaillé que celui de l'article/du GTIN, ce qui joue un rôle important dans le choix de la méthode d'indice.

La méthode Q-U

Le secteur de l'habillement est notoirement complexe en termes de calcul des indices de prix, car les catégories de produits peuvent présenter des taux de renouvellement élevés. Les méthodes d'indice bilatérales peuvent être problématiques : les méthodes bilatérales directes n'incluent pas les nouveaux produits dans le calcul de l'indice au cours d'une année donnée, mais seulement lors du mois de base suivant, tandis que les méthodes des indices chaînés mensuellement peuvent souffrir du biais de dérive en chaîne. L'étude comparative de Chessa *et al.* (2017) montre que les indices bilatéraux pondérés peuvent varier de manière significative par rapport aux indices transitifs, ce

3. Des mesures bilatérales ont été choisies afin de faciliter les calculs. Il est bien sûr possible de rallonger la période à un plus grand nombre de mois, mais dans ce cas il est plus difficile de caractériser la dynamique. Pour des informations plus détaillées, voir Willenborg (2017).

Figure IV
Dynamique du flux avec les données de caisse et les données moissonnées sur le Web pour les pantalons et jeans pour hommes, par année



Note : les trois mesures du flux sont exprimées en pourcentage et atteignent 100 % chaque mois.
Source : données de caisse et données moissonnées sur le Web pour des vêtements.

qui est contraire à la condition que les méthodes des indices de prix doivent respecter pour éviter la dérive en chaîne.

Contrairement aux méthodes bilatérales, qui utilisent des informations de deux périodes afin de calculer l'indice, les méthodes multilatérales utilisent des informations de plusieurs périodes. L'un des avantages des méthodes multilatérales sur les méthodes bilatérales est que des indices transitifs ne contenant aucun biais de chaînage peuvent être calculés selon différents poids pour différents produits, et peuvent même varier d'un mois à l'autre. Toutefois, certaines méthodes (dont celle dite GEKS, pour Gini-Eltetö-Köves-Szulc) sont sensibles aux biais baissiers pour des assortiments dynamiques dont les produits sortent des prix de liquidation (Chessa *et al.,* 2017). Ces situations ne sont pas rares dans le secteur de l'habillement (Chessa, 2016a). Pour cette raison, nous avons choisi une méthode qui ne souffre pas des problèmes susmentionnés, à savoir la « méthode Q-U » (de l'anglais *Quality-adjusted Unit value*, c'est-à-dire la valeur unitaire ajustée en fonction de la qualité), pour les données de caisse et les données moissonnées du magasin en ligne. Cette méthode a été introduite dans l'IPC néerlandais en janvier 2016 (Chessa, 2016a). Quand on l'utilise pour des comparaisons de prix entre différents pays, elle est appelée « méthode Geary-Khamis » (GK) et représente alors un cas particulier au sein de la classe globale des méthodes Q-U. Pour cette raison, nous préférons le second terme, ou aussi « Q-U–GK ».

Formule d'indice

Chessa *et al.* (2017) comparent les méthodes d'indice bilatérales et multilatérales pondérées et non pondérées sur les jeux de données de caisse de quatre catégories de produits d'un grand magasin néerlandais autre que celui traité dans cet article. Les poids utilisés dans les formules d'indice peuvent engendrer des résultats largement différents de ceux des méthodes basées sur l'équipondération. Cela étant, les poids utilisés dans les méthodes bilatérales peuvent eux aussi être problématiques, notamment lorsqu'ils sont utilisés pour calculer les indices chaînés mensuellement. Ces indices peuvent engendrer un biais de chaînage important, qui découle directement du caractère intransitif des indices bilatéraux chaînés mensuellement.

Les indices bilatéraux directs ne tiennent pas compte des nouveaux produits en temps voulu, ces derniers n'étant inclus que lors du mois de base suivant, sauf si les prix sont imputés pour les mois précédant le mois durant lequel le produit est ajouté à l'assortiment. Une comparaison faite dans le secteur de l'habillement montre que la contribution des nouveaux produits à un indice peut être considérable (Chessa *et al.*, 2017). Les méthodes multilatérales ne souffrent d'aucun biais de chaînage, ce qui permet d'inclure les nouveaux produits en temps voulu et évite d'avoir à imputer les prix.

La dynamique de l'assortiment justifie également de choisir une méthode multilatérale pour les données de caisse et les données moissonnées du magasin en ligne néerlandais. Les écarts constatés dans Chessa *et al.* (2017) entre les indices de prix basés sur différentes méthodes multilatérales ne sont pas grands, mais peuvent avoir une importance significative. La méthode GEKS, ainsi que la méthode CCDI récemment proposée par Diewert & Fox (2017), sont sensibles aux prix de liquidation des articles sortants, ce qui engendre des biais baissiers (Chessa *et al.*, 2017). D'autres méthodes, comme la méthode Q-U et l'indicatrice temps/produit, n'ont pas ce désavantage.

La méthode Q-U a été introduite au sein de l'IPC néerlandais en janvier 2016, et sa première application à l'indice concernait alors les téléphones portables. Depuis janvier 2017, elle est également appliquée aux données de caisse du grand magasin néerlandais susmentionné. La méthode Q-U peut être considérée comme une famille de méthodes, incluant également certaines méthodes bilatérales bien connues comme les indices de Laspeyres, Paasche et Fisher (voir également Auer, 2014). Mais son but premier consiste à établir des indices transitifs multilatéraux. De fait, la méthode élargit le concept de valeur unitaire à des ensembles de biens hétérogènes. Il faut ainsi tenir compte des différences de qualité qui existent entre les produits, raison pour laquelle nous parlons de « valeur unitaire ajustée en fonction de la qualité ». D'autres auteurs, par exemple Auer (2014), parlent de « valeur unitaire généralisée ».

Pour bien expliquer le concept qui étaye la méthode Q-U, introduisons quelques notations. Soit G_0 et G_t des ensembles de produits appartenant à une catégorie de produit G, pour un mois de base 0 et un mois en cours t. Les ensembles de produits de 0 et de t peuvent être différents. Soit $p_{i,t}$ et $q_{i,t}$ les prix et les quantités vendues du produit $i \in G_t$, respectivement, au cours du mois t. Nous voulons identifier des facteurs d'échelle, disons v_i, transformant les prix de différents produits au cours du mois t en « prix ajustés en fonction de la qualité » $p_{i,t} / v_i$. Cette transformation implique

de convertir les quantités vendues $q_{i,t}$ pour chaque produit en quantités $v_i\, q_{i,t}$. À l'expression (3) ci-dessous, les v_i des produits correspondent aux prix déflatés moyens sur un intervalle de temps donné. Les v_i pourraient être interprétés comme des « prix de référence » et les $v_i\, q_{i,t}$ comme des quantités évaluées aux prix de référence des produits.

La transformation des prix et des quantités nous permet de définir et de calculer une « valeur unitaire ajustée en fonction de la qualité » $\tilde{p}_t$ pour un ensemble de produits G_t au cours du mois t :

$$\tilde{p}_t = \frac{\sum_{i\in G_t}\left(p_{i,t} / v_i\right)\left(v_i q_{i,t}\right)}{\sum_{i\in G_t} v_i q_{i,t}} = \frac{\sum_{i\in G_t} p_{i,t} q_{i,t}}{\sum_{i\in G_t} v_i q_{i,t}} \qquad (1)$$

À noter que $\sum_{i\in G_t} p_{i,t} q_{i,t}$, la dépense totale, n'est pas affectée par la transformation.

L'expression (1) peut servir à définir un indice de prix en divisant les valeurs unitaires ajustées en fonction de la qualité en deux mois :

$$P_t = \frac{\tilde{p}_t}{\tilde{p}_0} = \frac{\sum_{i\in G_t} p_{i,t} q_{i,t} \,/\, \sum_{i\in G_0} p_{i,0} q_{i,0}}{\sum_{i\in G_t} v_i q_{i,t} \,/\, \sum_{i\in G_0} v_i q_{i,0}} \qquad (2)$$

Le numérateur à droite de l'expression (2) est un indice mesurant l'évolution du chiffre d'affaires ou des dépenses entre deux mois. Le dénominateur est un indice de quantité pondéré. L'expression (2) montre clairement pourquoi l'indice de prix est transitif : l'indice de chiffre d'affaires comme l'indice de quantité pondéré sont transitifs.

Les poids v_i sont définis comme suit sur un intervalle de temps donné $[0,T]$:

$$v_i = \sum_{z=0}^{T} \frac{q_{i,z}}{\sum_{s=0}^{T} q_{i,s}} \frac{p_{i,z}}{P_z} \qquad (3)$$

L'expression (3) revient à dire que les v_i sont également des valeurs unitaires. Pour chaque produit, les dépenses sont additionnées sur la période concernée $[0,T]$ puis divisées par les quantités vendues pour le produit durant ce même intervalle de temps. Afin d'exclure les variations de prix des v_i et de l'indice de quantité pondéré, les prix en vigueur pour les produits durant des mois différents sont déflatés par l'indice de prix de la catégorie de produit concernée. Les v_i sont également appelés « prix de référence » (ou « prix internationaux » dans un contexte géographique). L'expression (3) représente le choix effectué sur ces prix dans la méthode Geary-Khamis (GK).

Les prix déflatés moyens d'une période donnée sont donc utilisés pour calculer les quantités transformées $v_i\, q_{i,t}$. Les prix des produits en vigueur durant tous les mois d'un intervalle de temps donné $[0,T]$ sont utilisés, comme cela est d'usage, ainsi que dans d'autres méthodes multilatérales. Toutefois, il pourrait être utile d'envisager certaines améliorations pour l'expression (3) : par exemple, les prix réduits pourraient être exclus des v_i afin que les valeurs obtenues représentent mieux la qualité. Cela pourrait faire l'objet de recherches ultérieures.

Le choix des prix pour définir les v_i est courant dans la théorie de l'indice. La méthode Q-U peut être considérée comme une famille de méthodes d'indice, dans la mesure où différents choix effectués pour les v_i donnent lieu à des formules d'indice différentes. Afin de produire plusieurs exemples à titre d'illustration, nous examinons tout simplement l'ensemble de produits vendus durant les deux mois, à savoir $G_0 \cap G_t$. Si nous définissons $v_i = p_{i,0}$ pour chaque produit $i \in G_0 \cap G_t$, alors l'expression (2) devient un indice de prix de Paasche. Si nous définissons $v_i = p_{i,t}$ pour chaque produit i, alors la formule (2) devient un indice de prix de Laspeyres. Si les v_i sont égaux pour tous les produits, alors (2) se simplifie et devient un indice de valeur unitaire. C'est précisément ce à quoi nous pourrions nous attendre pour des produits de même qualité, puisque leurs quantités vendues peuvent être additionnées sans être transformées.

Dans la mesure où l'indice de prix sert de déflateur dans (3), les équations (2) et (3) doivent être résolues simultanément. Chessa (2016a) décrit un algorithme itératif commençant par des valeurs initiales arbitraires pour les indices de prix $P_1,\ldots,P_T$, avec $P_0 = 1$ (voir également Maddison & Rao, 1996). Ces indices de prix sont substitués dans l'expression (3), de sorte que les valeurs initiales puissent être calculées pour chaque v_i. Ces valeurs sont ajoutées à l'expression (2) afin d'actualiser les indices de prix initiaux. Ces deux étapes sont répétées jusqu'à ce que les différences entre les indices de prix des deux dernières étapes d'itération respectent un critère d'arrêt défini par l'utilisateur. Geary (1958), Khamis (1972), Auer (2014) et Chessa (2016a) fournissent des informations plus détaillées sur les méthodes Q-U et GK.

Avant d'appliquer la méthode, plusieurs questions doivent être traitées, et en premier lieu la longueur de la fenêtre temporelle, l'intervalle de temps $[0,T]$, ainsi que la façon dont des données

supplémentaires peuvent être ajoutées à mesure que de nouvelles données deviennent disponibles chaque mois. Nous traitons ensuite la question de la définition des produits inclus dans l'ensemble de biens G_t.

Longueur de la fenêtre temporelle

Pour choisir l'intervalle de temps, nous avons utilisé un mois de base fixe (le mois de décembre de l'année précédente), conformément à la réglementation sur les indices des prix à la consommation harmonisés. L'IPC néerlandais utilise un intervalle de 13 mois, que nous avons aussi retenu ici.

L'impact de tout changement de l'intervalle temporel sur les indices de prix est étudié dans Chessa *et al.* (2017) et, de façon plus poussée, dans Chessa (2017a). La première étude compare les intervalles de 13 mois et la période entière de 50 mois pour quatre catégories de produits. Des différences significatives ont été identifiées pour l'une des catégories. Dans Chessa (2017a), les différences ont également été quantifiées au niveau de la nomenclature COICOP. Les différences entre les intervalles de 13 mois et de 4 ans sont de l'ordre de dixièmes de point de pourcentage pour des indices en glissement annuel, et négligeables dans un nombre assez grand de catégories COICOP. Pour le détaillant de la grande chaîne de supermarchés néerlandaise, il n'apparaît aucune différence entre les deux intervalles.

Actualisation des poids et calcul de l'indice

À mesure que de nouvelles données deviennent disponibles chaque mois, l'inclusion de données supplémentaires peut engendrer des valeurs différentes pour les v_i, et les indices de prix calculés jusqu'au mois précédent peuvent eux aussi changer. Toutefois, sauf circonstances exceptionnelles, il est impossible de réviser les indices de prix dans l'IPC. Compte tenu de ce « problème de révision », comment pouvons-nous calculer un indice de prix pour le mois suivant ?

En théorie, la solution des équations (2) et (3) nous donne un ensemble de 13 indices transitifs pour toute année $[0,T]$, sachant que le mois de base 0 est le mois de décembre de l'année précédente et que $T = 12$ représente le mois de décembre de l'année en cours. Les indices de prix et les poids des produits ou les prix de référence v_i sont calculés pour les 13 mois de l'année de façon simultanée, de sorte que les v_i ont la même valeur chaque mois. Il serait possible de publier les indices qui en découlent s'il était possible de réviser les indices de prix des mois précédents à chaque fois que de nouvelles données sont incluses dans le calcul de l'indice pour un mois ultérieur. Les v_i calculés pour le mois de décembre de l'année en cours finissent par donner les valeurs souhaitées pour les poids des produits, ce qui pourrait servir à produire les indices transitifs chaque mois.

Dans la pratique, comme on ne peut pas prévoir les prix des mois à venir, la construction d'indices transitifs restera, au mieux, une référence théorique idéale. L'inclusion de données d'un mois ultérieur modifie la valeur des v_i et, en conséquence, les indices de prix des mois précédents. Comme en règle générale, il est impossible de réviser les indices de prix de mois précédents dans l'IPC, cela soulève la question du mode de calcul d'un indice de prix pour un mois ultérieur.

Différentes méthodes ont été proposées pour mettre les v_i à jour et calculer les indices de prix d'un mois ultérieur. Ces méthodes sont basées sur des choix relatifs à trois aspects[4] :

- l'utilisation d'un mois de base fixe ou d'un mois de référence mobile ;

- l'adoption d'une fenêtre glissante ou d'un intervalle à expansion mensuelle, sachant que ce dernier ne peut s'utiliser qu'avec un mois de base fixe ;

- l'utilisation d'une méthode d'indice directe, de chaînage mensuel ou de raccordement.

Chessa (2016a) propose une méthode de mois de base fixe, un intervalle à expansion mensuelle et une méthode directe pour calculer l'indice de prix d'un mois ultérieur. Cette méthode utilise des données tirées de nombres de mois différents tout au long d'une année (deux mois en janvier, trois en février et ainsi de suite jusqu'à un nombre maximal de 13 mois en décembre) et ne requiert pas de données historiques. La méthode directe calcule les indices de prix du mois en cours par rapport au mois de base, en utilisant l'ensemble de valeurs le plus récent pour les v_i.

Elle fait en sorte que les indices de prix de décembre correspondent aux indices de prix transitifs qui auraient été obtenus en utilisant les données complètes des 13 mois pour chaque mois de l'année. Ainsi, la méthode de « fenêtre d'expansion mensuelle à base fixe » (ou FBEW de l'anglais *fixed*

4. À noter que ces choix, et en conséquence le type de méthode d'actualisation, peuvent être appliqués en combinaison avec toute méthode multilatérale. Un exemple en est fourni dans Chessa et al. *(2017).*

base monthly expanding window) permet d'éviter le biais de chaînage. Les séries d'indices de plus d'un an sont construites en chaînant la série de l'année en cours à l'indice du mois de décembre de l'année précédente, de sorte qu'une forme de chaînage est finalement réalisée. Mais c'est une forme de chaînage moins fréquente et, par ailleurs, compte tenu de l'utilisation d'un intervalle de 13 mois, la valeur théorique des v_i peut varier d'une année à l'autre pour chaque produit. Il s'agit d'un choix explicite, qui pourrait être fait pour refléter la variation progressive de la qualité au fil du temps.

La fenêtre à expansion mensuelle pourrait également être remplacée par un intervalle glissant de 13 mois, tout en maintenant le calcul des indices de prix à l'aide d'une méthode directe par rapport à un mois de base fixe. Cette méthode alternative est comparée à la méthode FBEW dans Chessa (2017a) et dans Lamboray (2017). Les différences entre les deux méthodes se sont avérées peu importantes ou négligeables. Les indices calculés à l'aide de méthodes d'actualisation et les « indices de référence » transitifs se sont avérés quasiment ou complètement égaux dans chacun des cas étudiés (Chessa, 2016a ; 2017a ; 2017b). Des différences importantes se sont occasionnellement manifestées, pour la plupart sur de courtes durées.

Une autre catégorie de méthodes consiste à utiliser un mois de référence mobile au lieu d'un mois de base fixe. L'un des choix les plus naturels est de combiner un mois de référence mobile avec une fenêtre glissante d'une durée fixe, ce qui permet d'inclure les données d'un mois ultérieur de façon élégante. Différentes méthodes peuvent être envisagées pour calculer un indice de prix pour le mois en cours, que l'on appelle « méthodes de raccordement ». Voir de Haan *et al.* (2016) pour une vue d'ensemble et Chessa *et al.* (2017) et Krsinich (2014) pour des applications précises.

La méthode du « raccordement des variations » chaîne l'indice en glissement mensuel de l'intervalle glissant le plus récent à l'indice du mois précédent, tandis que la méthode du « raccordement des intervalles » de Krsinich (2014) chaîne l'indice en glissement annuel de l'intervalle complet le plus récent à l'indice d'il y a douze mois. La méthode du raccordement des variations est une méthode de chaînage mensuel qui, par nature, subit le biais de chaînage. Bien que la méthode du raccordement relève d'une sorte de méthode directe, c'est aussi une méthode de chaînage à haute fréquence. Les résultats empiriques indiquent un biais de chaînage potentiel, qui pourrait être important (Chessa, 2016b).

Indices de prix pour les données moissonnées sur le Web et les données de caisse

Préparation des données et choix méthodologiques

Nous avons calculé les indices de prix à l'aide de la méthode Q-U pour les vêtements pour hommes et pour femmes du magasin en ligne néerlandais, sur la base des données de caisse et exclusivement à l'aide des données moissonnées sur le Web. Pour établir des comparaisons pertinentes, nous avons complété les données de caisse par les métadonnées tirées des données moissonnées. Pour cela, nous avons relié les deux tableaux de données en utilisant les codes d'article spécifiques au détaillant en guise de clé de couplage. Nous avons calculé les indices de prix de huit catégories de produits dans le segment des vêtements pour hommes (pantalons et jeans, manteaux et vestes, sous-vêtements et pyjamas, chemises, chaussures, vêtements de sport, pulls et gilets, tee-shirts et polos) et dans celui des vêtements pour femmes (pantalons et jeans, manteaux et vestes, robes et jupes, lingerie, chaussures, vêtements de sport, pulls et gilets, tee-shirts et hauts).

Les huit catégories couvrent environ 85 % des dépenses totales en vêtements pour hommes pour la période allant de mars 2014 à décembre 2016, et environ 80 % des dépenses totales en vêtements pour femmes. Les articles en promotion et haut de gamme ont également été inclus[5].

La première étape préalable au calcul de l'indice de prix consiste à définir le produit. Bien que cela ne soit pas la priorité de cette étude (qui vise à comparer les données de caisse et les données moissonnées sur le Web), il est évident que cette démarche doit être effectuée avec soin, puisque les indices peuvent être très sensibles à toute variation du degré de différenciation entre les produits (Chessa, 2016a; 2017b).

Le secteur de l'habillement affiche habituellement un taux de renouvellement élevé, qui se remarque également à un niveau moins détaillé que le niveau de l'article ou du GTIN (voir la figure IV). Les articles sortants et les nouveaux articles de qualité identique ou similaire doivent être reliés afin d'éviter que les indices ne subissent des biais à la baisse, dont l'ampleur peut être extrêmement importante si les articles sortent de l'assortiment

5. Les articles hors secteur de l'habillement inclus dans ces deux groupes ont été exclus durant l'extraction des données pour chacune des catégories susmentionnées.

à des prix de liquidation (Chessa, 2016a). Les articles sortants et les nouveaux articles peuvent être reliés par des caractéristiques communes, ici le nom de la marque et le « type », c'est-à-dire le niveau le plus détaillé de classification d'article.

Les articles sont donc rassemblés dans le même groupe lorsqu'ils appartiennent aux mêmes groupes [marque×type], que nous appelons « produits ». Les produits doivent être homogènes, c'est-à-dire que les articles d'un groupe donné doivent être de qualité identique ou comparable. Cette question devrait être étudiée de façon plus détaillée dans des travaux ultérieurs, notamment lorsque l'on envisage d'inclure les données d'un magasin en ligne dans l'IPC. La taille moyenne des « produits » va de 7 à 16 articles ; sachant que les codes d'article et les GTIN sont habituellement différents pour les différentes tailles de vêtements, et que ces vêtements sont considérés comme étant de même qualité, cette fourchette suggère que les définitions des produits sont assez étroites.

Les choix suivants ont été faits pour appliquer la méthode Q-U aux données de caisse et aux données du Web.

- Pour les données de caisse, les valeurs unitaires ont été calculées pour chaque produit lors de chaque mois durant lequel il a été vendu. Les dépenses et les quantités d'articles vendus pour un produit donné ont été additionnées, et le chiffre d'affaires correspondant aux articles retournés a été exclu.

- Pour les données moissonnées, les prix mensuels moyens ont été calculés pour chaque produit. Les quantités vendues ont été remplacées par le nombre total de prix moissonnés pour un produit donné au cours d'un mois, puis additionnées pour tous les articles. Les articles peuvent être moissonnés plus d'une fois et les nombres multiples sont conservés dans les quantités et les prix moyens.

- La méthode Q-U a été appliquée avec un mois de base fixe, à savoir le mois de décembre de chaque année comme dans l'IPC néerlandais. Le mois de base de 2014 est le mois de mars, car il s'agit du premier mois de la période choisie pour cette étude. Des intervalles de temps de 13 mois ont été utilisés (sauf bien sûr en 2014). Nous n'avons pas appliqué de mise à jour mais avons calculé les poids v_i et les indices de prix à l'aide des données complètes de tous les mois d'une année.

Le tableau 1 donne un exemple (pour des vêtements) de la façon dont les prix et les quantités des produits ont été calculés.

Le prix du produit est calculé à partir des données de caisse comme valeur unitaire, c'est-à-dire qu'il s'agit du rapport entre la somme des dépenses des six articles et la somme des quantités. Les dépenses et les quantités des articles retournés sont exclues, de sorte que ces valeurs sont additionnées avec

Tableau 1
Calcul des prix et des quantités de produits

Article	N° 1	N° 2	N° 3	N° 4	N° 5	N° 6	Produit
A – Données de caisse							
Dépense nette	0	118	13 201	2 711	25 108	13 009	-
Dépenses retournées	75	3 377	7 174	2 257	7 481	15 004	-
Quantité nette	0	0	899	186	1 643	986	-
Quantités retournées	5	198	372	124	434	812	-
Dépense	75	3 495	20 375	4 968	32 589	28 013	89 515
Quantité	5	198	1 271	310	2 077	1 798	5 659
Prix	14.95	17.65	16.03	16.03	15.69	15.58	15.82
B – Données moissonnées sur le Web							
Nombre de prix moissonnés	5	22	31	31	31	29	149
Somme des prix moissonnés	74.75	392.21	523.22	626.02	523.22	557.57	2 696.99
Prix	14.95	17.83	16.88	20.19	16.88	19.23	18.10

Note : pour les données de caisse : dépenses et quantités, tant pour les valeurs nettes que pour les retours de tee-shirts à manches courtes de la même marque. Les six articles ont des codes d'article différents (N° 1 à 6), qui sont combinés dans le même produit en fonction de caractéristiques communes. Les dépenses totales, la quantité totale et le prix (valeur unitaire en euros) du produit sont également indiqués. Les valeurs sont tirées des données de caisse du magasin en ligne et concernent un mois. Pour les données moissonnées sur le Web : nombres de prix moissonnés et somme de ces prix pour les mêmes articles et le même mois que pour les données de caisse. Ces valeurs sont également indiquées pour le produit ; elles correspondent à la somme des six articles.
Source : données de caisse et données moissonnées sur le Web.

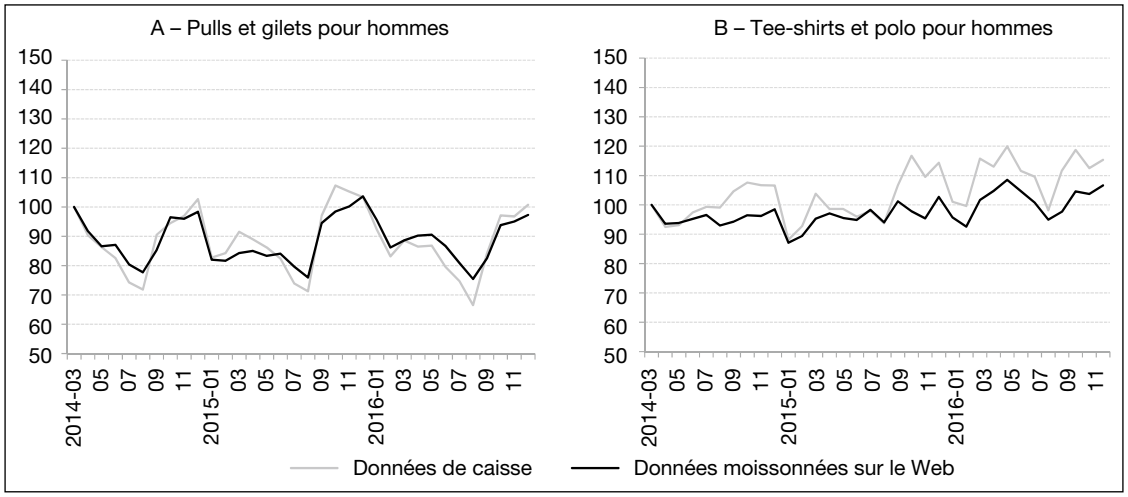
les quantités et les dépenses nettes. Pour le prix à partir des données moissonnées sur le Web (cf. tableau 1, B), il s'agit du rapport entre la somme des prix moissonnés durant les jours du mois et le nombre total des prix des articles moissonnés, additionnés pour les six articles (cf. dernière rangée du tableau 1, B).

Indices de prix

Les figures V et VI montrent les indices de prix calculés à partir de chaque source de données pour deux catégories de vêtements pour hommes et pour femmes. Les indices de prix calculés à partir des données moissonnées sur le Web suivent plus ou moins ceux calculés à partir des données de caisse, même les pics et creux des indices de données de caisse. La forte corrélation des prix comme des quantités entre données de caisse et données du Web se retrouve dans la comparaison des indices de prix. La correspondance étroite entre les indices de prix basés sur les deux jeux de données se retrouve également dans la totalité des 16 catégories de produits (voir en annexe les indices de prix de toutes les catégories de produits).

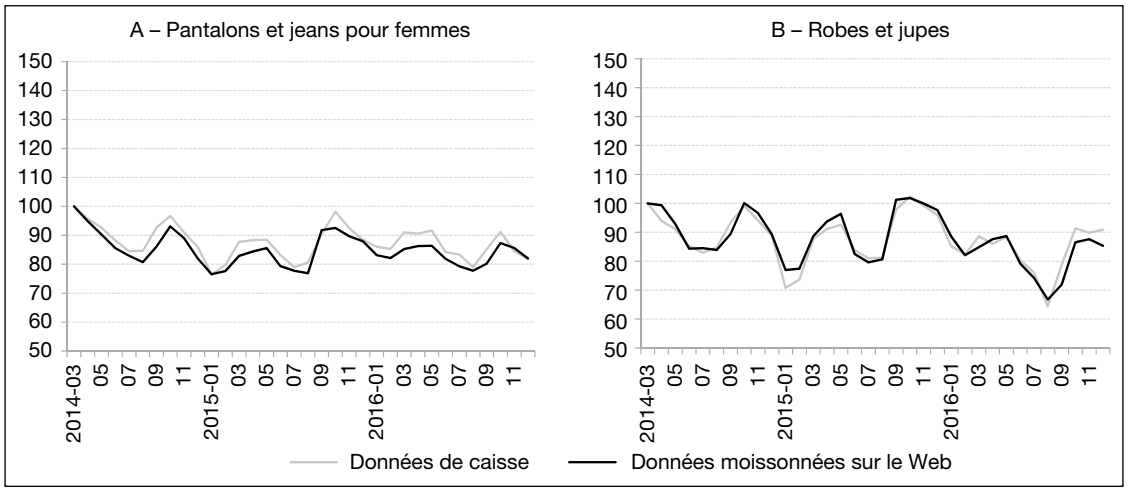
Les indices de prix des catégories de produits ont été combinés en appliquant l'habituelle méthode de Laspeyres, pour les vêtements pour hommes

Figure V
Indices Q-U pour deux catégories de vêtements pour hommes, pour les données de caisse et les données moissonnées sur le Web (mars 2014 = 100)



Source : données de caisse et données moissonnées sur le Web pour des vêtements.

Figure VI
Indices Q-U pour deux catégories de vêtements pour femmes, pour les données de caisse et les données moissonnées sur le Web (mars 2014 = 100)



Source : données de caisse et données moissonnées sur le Web pour des vêtements.

et les vêtements pour femmes classés en fonction de la nomenclature COICOP, et sont indiqués figure VII. S'agissant des données de caisse, on utilise des poids fixes annuels pour les catégories de produits. Les poids des catégories ont été fixés à un niveau égal à la part annuelle des dépenses de la catégorie concernée pour l'année précédente, excepté pour l'année 2014, première de la série, pour laquelle nous avons pris la part annuelle des dépenses de 2014.

Avec les données du Web, nous avons remplacé les dépenses par le prix moyen multiplié par le nombre de prix de produits moissonnés, additionné pour tous les produits au sein d'une catégorie donnée durant une année. Les différences entre les indices basés sur des données de caisse et ceux basés sur des données du Web sont très faibles pour les deux catégories COICOP. Entre les indices en glissement annuel, elles sont en moyenne seulement de 0.3 point de pourcentage, pour les deux catégories COICOP.

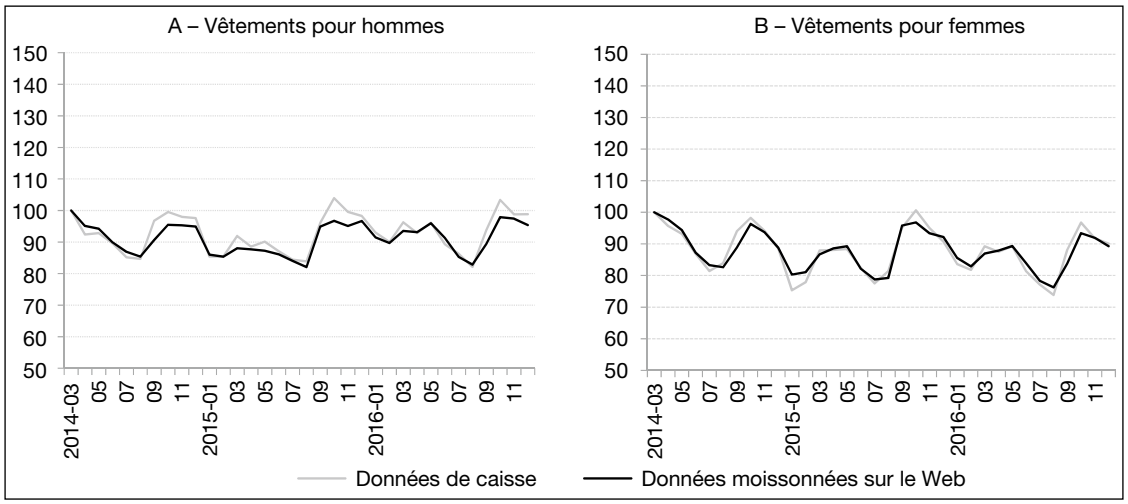
Analyse de sensibilité

Les résultats indiquent que, en utilisant le nombre de prix de produits moissonnés sur le Web au lieu du nombre de produits vendus, on obtient des indices de prix fiables. Cette conclusion correspond aux résultats de l'analyse des données fournie en première partie de cet article. Pour aller plus loin, nous avons cherché à déterminer si le fait de remplacer le nombre de prix de produits moissonnés par un nombre excluant toute corrélation avec le nombre de produits vendus était susceptible d'affecter les indices de prix. Nous avons remplacé le nombre de prix moissonnés par 0 ou 1, 0 indiquant que l'outil de moissonnage n'avait trouvé aucun prix pour un produit donné durant un mois donné et 1 indiquant que des prix avaient été trouvés mais que les nombres exacts avaient été ignorés. L'impact de ce changement sur les indices de prix est illustré figure VIII. Les résultats sont montrés uniquement au niveau de la nomenclature COICOP.

Le fait de remplacer le nombre de prix de produits moissonnés par 0 ou par 1 a un impact considérable sur les indices basés sur des données moissonnées sur le Web, ce qui est clairement visible au niveau de la nomenclature COICOP. Les résultats ne sont pas présentés pour les 16 catégories de produits, mais des différences du même ordre ont été identifiées pour 13 d'entre elles. Chacun de ces cas montre une tendance à la baisse de l'indice (comme dans la figure VIII).

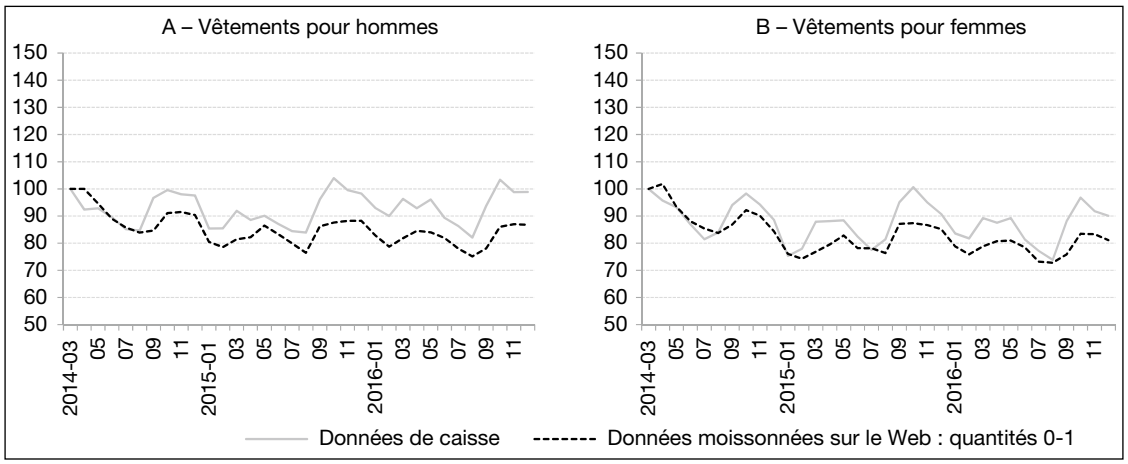
Les différences entre les indices d'une année à l'autre sont beaucoup plus importantes que celles obtenues avec les nombres initiaux de prix moissonnés. La différence moyenne avec les indices basés sur des données de caisse passe à près de 5 points de pourcentage pour les vêtements pour hommes et à près de 4 points de pourcentage pour les vêtements pour femmes. Ces résultats suggèrent que les nombres initiaux de prix moissonnés sur le Web devraient être utilisés pour calculer des indices de prix à partir de données moissonnées sur le Web. La manipulation de ces nombres, comme par exemple le retrait des prix dupliqués, est à éviter.

Figure VII
Indices de prix des vêtements pour hommes et pour femmes (mars 2014=100)



Source : données de caisse et données moissonnées sur le Web pour des vêtements.

Figure VIII
Indices de prix pour des vêtements pour hommes et pour femmes, avec le nombre de prix de produits moissonnés sur le Web remplacé par une valeur binaire (mars 2014 = 100)



Source : données de caisse et données moissonnées sur le Web pour des vêtements.

* *
*

À notre connaissance, l'étude ici présentée est la première à comparer des indices de prix calculés à partir de données de caisse et à partir de données moissonnées sur le Web. Cette comparaison a été possible parce que les deux sources de données sont disponibles pour le même détaillant. Ces premiers résultats semblent très prometteurs car les indices basés sur des données moissonnées sont d'une exactitude remarquable, surtout au niveau de la nomenclature COICOP. C'est d'autant plus utile que le recours au moissonnage du Web est de plus en plus envisagé pour les statistiques publiques. Les données de caisse restent l'option privilégiée car elles contiennent des données de transaction, mais tous les INS n'y ont pas nécessairement accès.

Ces résultats positifs et intéressants mettent encore plus l'accent sur la nécessité de comprendre pourquoi les indices de prix calculés uniquement à partir de données moissonnées sont si proches de ceux obtenus à partir de données de caisse. À ce stade, nous ne pouvons que spéculer sur les raisons possibles. Cela pourrait par exemple découler du fait que le détaillant est un magasin en ligne dépourvu de point de vente physique et qu'il pourrait donc être plus disposé à promouvoir les articles les plus vendus de son assortiment. Ces articles peuvent être agencés de façon à être rapidement identifiés par le consommateur sur le site Web, en les plaçant dans différents groupes principaux ou catégories. Par exemple, le même article peut être placé dans le groupe principal « En promotion » et dans l'un des autres groupes principaux classiques. Cela expliquerait, du moins en partie, la forte corrélation entre le nombre de produits vendus et le prix des produits moissonnés. Pour vérifier si un détaillant est plus susceptible de promouvoir les articles les plus vendus, il faudrait prendre contact avec lui afin qu'il explique la stratégie d'agencement de son site Web.

De façon plus générale, nous pouvons tirer plusieurs enseignements de cette étude.

- La méthode d'échantillonnage des prix sur un site Web est indéniablement importante. Cette étude montre, tout au moins pour le détaillant étudié dans cet article, que le moissonnage d'un site Web entier favorise l'exactitude des indices de prix calculés à partir de données du Web. Certes, il faut du temps pour moissonner un site Web entier, mais les instituts de statistique pourraient envisager de procéder à l'échantillonnage seulement certains jours et non quotidiennement.

- Le site Web traité dans cette étude a été moissonné par navigation, grâce à un outil de moissonnage de première génération développé par CBS. Cette technique prend elle aussi du temps, ce qui explique en partie la décision de moissonner la nuit. Les magasins en ligne utilisent une tarification dynamique. Les prix peuvent diminuer durant les heures d'ouverture, et tout prix manquant peut donc expliquer en partie les différences constatées entre les prix moissonnés et les prix tirés des données de caisse. Parallèlement, CBS a développé une deuxième génération

d'outils de moissonnage permettant d'extraire les prix et les métadonnées du code indiqué sur la page d'aperçu du produit. Cette technique est beaucoup plus rapide, ce qui permet de moissonner des sites Web de grande taille à différents moments de la journée. À l'avenir, cela permettra d'étudier l'impact de la tarification dynamique sur les indices de prix et de nous concentrer sur de nouvelles applications, comme par exemple la construction d'indices en temps réel. L'impact de la tarification dynamique sur les indices de prix est bien sûr impossible à quantifier ici. Toutefois, les différences constatées entre les indices de prix basés sur des données de caisse et ceux basés sur des données du Web sont faibles, ce qui suggère que cet impact est peu important dans le cas étudié ici.

- Cette étude suggère également d'utiliser le nombre initial de prix moissonnés pour calculer un indice de prix avec des données moissonnées sur le Web. La déduplication des prix est à éviter. Les résultats indiquent que les indices basés sur des données moissonnées perdent de leur exactitude lorsque l'on supprime les prix multiples (cf. figure VIII) : la différence avec les indices d'une année sur l'autre basés sur des données de caisse augmente alors à cinq points de pourcentage par an. En outre, tous les indices affichant un écart montrent une dérive à la baisse. Cela dit, nous admettons que la suppression des prix multiples a été assez extrême, dans la mesure où nous n'avons laissé qu'une seule observation par produit par mois. Néanmoins, les résultats indiquent que le nombre initial de prix moissonnés devrait être géré avec prudence.

- Il est toujours intéressant de demander aux détaillants de fournir des données sur les dépenses, même s'ils ne peuvent – ou ne veulent – pas fournir des données de caisse complètes.

Malgré tout, nous devons formuler des conclusions prudentes. En effet, les données moissonnées sur le Web ne sont pas des données de transaction et les résultats de cette étude ne concernent qu'un seul détaillant. L'analyse développée ici pourrait donc être complétée dans deux directions.

Tout d'abord, elle pourrait être élargie à d'autres magasins en ligne dont le site Web présente une structure semblable à celle analysée dans cet article, c'est-à-dire où les articles à prix réduit sont mis en avant plus souvent que les autres articles et où les articles les plus populaires sont plus faciles à trouver. L'unité IPC de CBS développe actuellement des outils de moissonnage du Web destinés aux détaillants vendant des appareils électroniques dont les données de caisse sont disponibles. Cela serait un cas d'étude intéressant, d'autant plus que ces détaillants ont des boutiques physiques. Promeuvent-ils les articles les plus vendus plus souvent que les articles moins populaires sur leur site Web ? Ou bien appliquent-ils une stratégie différente, par exemple en faisant la publicité des nouveaux articles ?

Le moissonnage du Web est un moyen utile pour compléter les informations collectées sur les articles dans les données de caisse, qui peuvent être limitées. En combinant les deux sources de données, on peut profiter de tous leurs avantages : des données de transaction fournies par les données de caisse et des informations supplémentaires sur les caractéristiques des articles fournies par les données moissonnées. En principe, ces conditions sont idéales pour appliquer et tester des méthodes visant à sélectionner les caractéristiques des articles et à définir des produits homogènes, permettant donc de traiter les relances. Toutefois, lorsque l'on utilise des métadonnées moissonnées pour compléter celles tirées des données de caisse fournies par des magasins physiques, il peut s'avérer impossible de compléter tous les GTIN inclus dans les données de caisse par des données moissonnées sur le Web. Les assortiments des magasins physiques et en ligne peuvent être différents, par exemple si les détaillants ne veulent inclure sur leur site Web qu'une partie des articles proposés dans les magasins physiques.

Pour finir, nous sommes conscients du fait que les études comparatives telles que celle présentée ici peuvent être difficiles à répliquer, car il est assez rare de disposer à la fois les données de caisse et les données moissonnées sur le Web pour un même détaillant. C'est encore plus difficile pour les INS rencontrant des problèmes dans l'acquisition des données de caisse. Pour cette raison, nous encourageons les INS ayant la chance de posséder des données de caisse à investir dans la recherche statistique sur données de caisse. Est-il possible, grâce à des analyses et tests statistiques, de caractériser les données de caisse ? Est-il possible d'identifier des tendances spécifiques, par exemple sur les corrélations entre les prix et les quantités au fil du temps ? En appliquant les mêmes analyses aux données moissonnées sur le Web, on pourrait évaluer leurs similarités avec les données de caisse et leur pertinence pour le calcul des indices de prix. Nous suggérons donc de donner une plus grande importance à l'analyse des séries temporelles et à d'autres analyses des données de caisse. □

BIBLIOGRAPHIE

Auer, L. von (2014). The Generalized Unit Value Index Family. *Review of Income and Wealth*, 60, 843–861. https://doi.org/10.1111/roiw.12042

Breton, R., Flower, T., Mayhew, M., Metcalfe, E., Milliken, M., Payne, C., Smith, T., Winton, J. & Woods, A. (2016). Research indices using web scraped data: May 2016 update. Office for National Statistics, internal report, 23 May 2016. https://www.ons.gov.uk/releases/researchindicesusingwebscrapedpricedatamay2016update

Cavallo, A. F. (2016). Are online and offline prices similar? Evidence from large multi-channel retailers. NBER, *Working Paper* N° 22142. http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2016/Session_2_MIT_are_online_and_offline_prices_similar.pdf

Chessa, A. G. (2016a). A new methodology for processing scanner data in the Dutch CPI. *Eurostat Review on National Accounts and Macroeconomic Indicators*, 1/2016, 49–69. https://ec.europa.eu/eurostat/cros/content/new-methodology-processing-scanner-data-dutch-cpi-antonio-g-chessa_en

Chessa, A. G. (2016b). Comparisons of the QU-method with other index methods for scanner data. Paper prepared for the first meeting on multilateral methods organised by Eurostat, Luxembourg, 7-8 December 2016. Statistics Netherlands, Internal paper.

Chessa, A. G. (2017a). Comparisons of QU-GK indices for different lengths of the time window and updating methods. Paper prepared for the second meeting on multilateral methods organised by Eurostat, Luxembourg, 14-15 March 2017. Statistics Netherlands, Internal paper.

Chessa, A. G. (2017b). The QU-method: A new methodology for processing scanner data. *Statistics Canada International Symposium Series : Proceedings*. https://www150.statcan.gc.ca/n1/en/catalogue/11-522-X201700014752

Chessa, A. G., Verburg, J. & Willenborg, L. (2017). A comparison of price index methods for scanner data. Paper presented at the *15th Meeting of the Ottawa Group on Price Indices*, Eltville am Rhein, Germany, 10-12 May 2017. http://www.ottawagroup.org/Ottawa/ottawagroup.nsf/4a256353001af3ed4b2562bb00121564/1ab31c25da944ff5ca25822c00757f87/$FILE/A comparison of price index methods for scanner data -Antonio Chessa, Johan Verburg, Leon Willenborg -Paper.pdf

Daas, P. J. H. & van Nederpelt, P. W. M. (2010). Application of the object oriented quality management model to secondary data sources. Statistics Netherlands, the Hague/Heerlen, The Netherlands, *Discussion paper* N° 10012.

Daas, P. J. H. & Ossen, S. J. L. (2010). In search of the composition of data quality in statistics and other research areas. Statistics Netherlands, *Discussion paper*.

Diewert, W. E. & Fox, K. J. (2017). Substitution bias in multilateral methods for CPI construction using scanner data. Vancouver School of Economics, The University of British Columbia, *Discussion paper* N° 17-02. https://irs.princeton.edu/sites/irs/files/Diewert and Fox Substitution Bias and MultilateralMethodsForCPI_DP17-02_March23.pdf

Eurostat (2017). *Practical Guide for Processing Supermarket Scanner Data*. September 2017. https://circabc.europa.eu/sd/a/8e1333df-ca16-40fc-bc6a-1ce1be37247c/Practical-Guide-Supermarket-Scanner-Data-September-2017.pdf

Geary, R. C. (1958). A note on the comparison of exchange rates and purchasing power between countries. *Journal of the Royal Statistical Society A*, 121, 97–99. https://doi.org/10.2307/2342991

Griffioen, A. R. & ten Bosch, O. (2016). On the use of internet data for the Dutch CPI. Paper presented at the *UNECE-ILO Meeting of the Group of Experts on Consumer Price Indices*, Geneva, Switzerland, 2-4 May 2016. https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2016/Session_2_Netherlands_on_the_use_of_internet_data_for_the_Dutch_CPI.pdf

Griffioen, A. R., ten Bosch, O. & Hoogteijling, E. H. J. (2016). Challenges and solutions to the use of internet data in the Dutch CPI. Paper presented at the *UNECE Workshop on Statistical Data Collection,* The Hague, The Netherlands, 3-5 October 2016. https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2016/mtg1/WP2-3_Netherlands_-_Griffioen_ap.pdf

de Haan, J., Willenborg, L. & Chessa, A. G. (2016). An overview of price index methods for scanner data. Paper presented at the *UNECE-ILO Meeting of the Group of Experts on Consumer Price Indices*, Geneva, Switzerland, 2-4 May 2016. https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2016/Session_1_room_doc_Netherlands_an_overview_of_price_index_methods.pdf

ILO/IMF/OECD/UNECE/Eurostat/The World Bank (2004). *Consumer Price Index Manual: Theory and Practice*. Geneva: ILO Publications. https://doi.org/10.5089/9787509510148.069

Khamis, S. H. (1972). A new system of index numbers for national and international purposes. *Journal of the Royal Statistical Society A*, 135, 96–121. https://doi.org/10.2307/2345041

Krsinich, F. (2014). The FEWS Index: Fixed Effects with a Window Splice – Non-revisable quality-adjusted price indexes with no characteristic information. Paper presented at the *UNECE-ILO Meeting of the group of experts on consumer price indices*, Geneva, Switzerland, 26-28 May 2014. https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2014/New_Zealand_-_FEWS.pdf

Lamboray, C. (2017). The Geary Khamis index and the Lehr index: how much do they differ? Paper presented at the *15th Meeting of the Ottawa Group on Price Indices*, Eltville am Rhein, Germany, 10-12 May 2017. http://www.ottawagroup.org/Ottawa/ottawagroup.nsf/4a256353001af3ed4b2562bb00121564/1ab31c25da944ff5ca25822c00757f87/$FILE/The Geary Khamis index and the Lehr index how much do they differ - Claude Lamboray -Paper.pdf

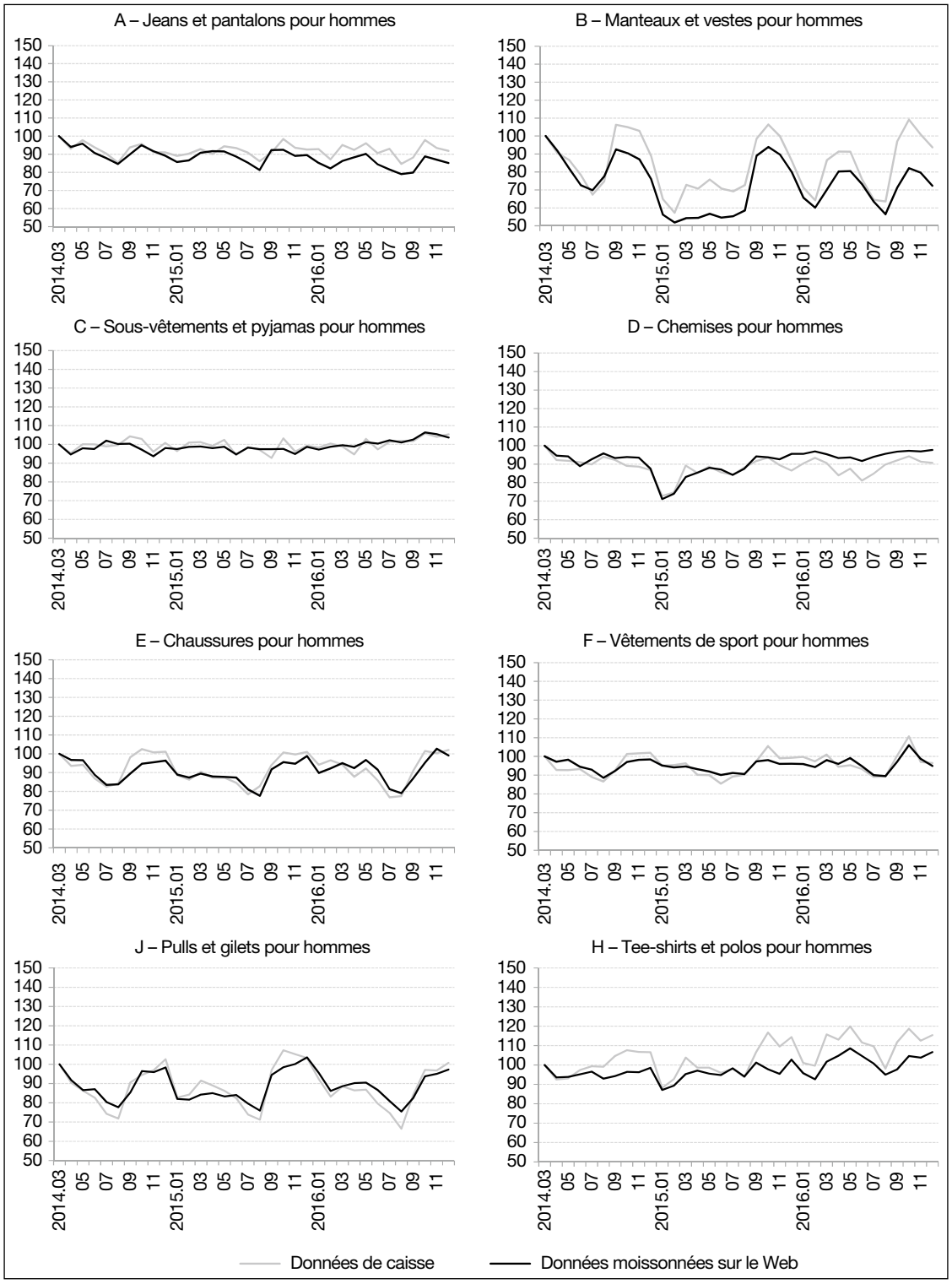
Maddison, A. & Rao, D. S. P. (1996). A generalized approach to international comparison of agricultural output and productivity. Groningen Growth and Development Centre, Research memorandum GD-27. https://www.rug.nl/research/portal/files/3258249/GD-27.pdf

Willenborg, L. (2017). Quantifying the dynamics of populations of articles. Statistics Netherlands, *Discussion Paper* N° 2017/10. https://www.cbs.nl/en-gb/background/2017/25/quantifying-the-dynamics-of-populations-of-articles

ANNEXE

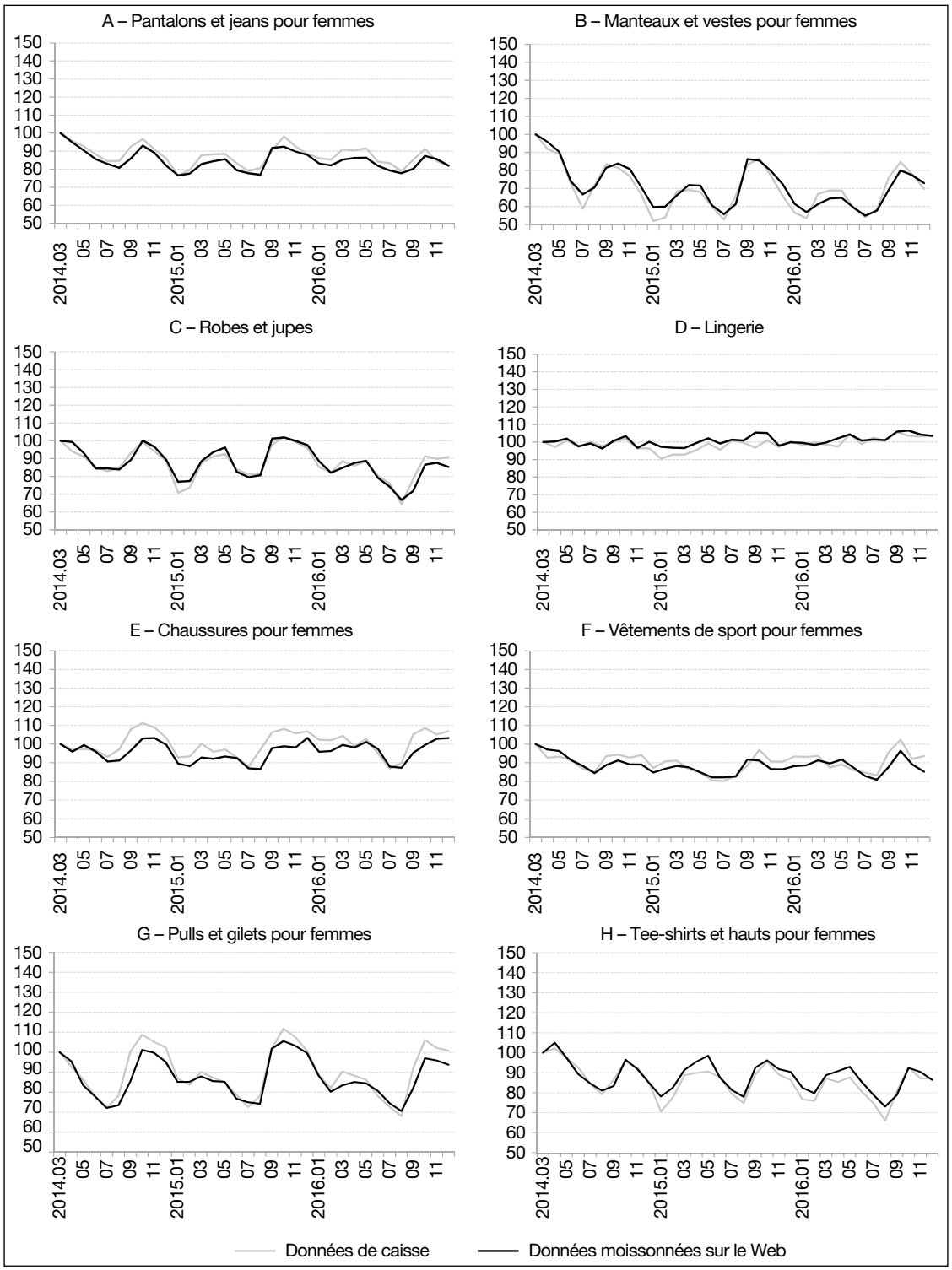
DONNÉES DE CAISSE ET DONNÉES MOISSONNÉES SUR LE WEB : INDICES DE PRIX POUR 16 CATÉGORIES DE PRODUITS DANS LE SECTEUR DES VÊTEMENTS POUR HOMMES ET POUR FEMMES

Figure A-I
Vêtements pour hommes (mars 2014 = 100)



Source : données de caisse et données moissonnées sur le Web pour des vêtements.

Figure A-II
Vêtements pour femmes (mars 2014 = 100)



Source : données de caisse et données moissonnées sur le Web pour des vêtements.