

## La méthode QU : une nouvelle méthodologie pour le traitement des données de lecteurs optiques

Antonio G. Chessa<sup>1</sup>

### Résumé

Le présent article décrit une nouvelle méthode de calcul des indices de prix pour le traitement des données issues de transactions électroniques (données de lecteurs optiques). Les indices de prix sont calculés sous forme du ratio d'un indice du chiffre d'affaires et d'un indice pondéré de quantité. Les pondérations de produits des quantités vendues sont calculées à partir des prix corrigés de l'inflation pour chaque mois de l'année de publication courante. Les nouveaux produits peuvent être incorporés en temps voulu, sans imputation de prix, de manière à ce que toutes les transactions puissent être traitées. Les pondérations des produits sont mises à jour mensuellement et sont utilisées afin de calculer les indices directs par rapport à un mois de base fixe. Les indices des prix construits de cette façon sont exempts du problème de « chain drift » (biais de chaînage). Les résultats sont robustes avec des écarts des choix méthodologiques. La méthode fait partie du programme de l'IPC des Pays-Bas depuis janvier 2016, quand elle a été appliquée pour la première fois aux téléphones mobiles.

Mots-clés : données de lecteurs optiques, GTIN, relancement, homogénéité des produits, indice des prix, « chain drift » (biais de chaînage).

### 1. Introduction

Les ensembles de données de lecteurs optiques contiennent les renseignements sur le chiffre d'affaires et le nombre d'unités vendues pour chaque article, identifié de manière unique par son code à barres (GTIN, Global Trade Item Number). Outre l'information sur les ventes, les données de lecteurs optiques devraient aussi contenir des renseignements descriptifs permettant de caractériser les articles. Les pays qui utilisent des données de lecteurs optiques dans leur IPC demandent habituellement aux détaillants de spécifier les transactions par semaine.

Les données de lecteurs optiques présentent d'importants avantages comparativement aux données collectées par les enquêtes traditionnelles. Elles permettent de calculer les prix moyens de transactions par GTIN, tandis que les données sur le chiffre d'affaires peuvent servir à calculer les parts de dépenses au niveau du GTIN. Les données d'enquête ne contiennent que les prix en rayon, qui ne sont pas nécessairement égaux aux prix éventuellement payés par les consommateurs. Les renseignements sur les nombres d'articles vendus et sur le chiffre d'affaires ne sont pas recueillis dans le cadre des enquêtes traditionnelles. Les données de lecteurs optiques offrent la possibilité de calculer des indices plus exacts, en utilisant une pondération fondée sur les dépenses au niveau du GTIN. Un autre avantage évident des données de lecteurs optiques tient au fait que l'on dispose des transactions effectuées sur tous les articles vendus. Par conséquent, les données de lecteurs optiques permettent de remplacer les méthodes basées sur des échantillons par des méthodes qui permettent le traitement intégral des données. En dépit du potentiel des données de lecteurs optiques, seul un petit nombre d'organismes statistiques les utilisent à l'heure actuelle dans leur IPC, mais ce nombre augmentera vraisemblablement dans les années à venir<sup>2</sup>.

Au moment de leur introduction dans l'IPC des Pays-Bas en 2002, les données de lecteurs optiques provenaient de deux chaînes de supermarchés. À l'heure actuelle, le programme utilise les données de lecteurs optiques de 10 chaînes et a cessé de réaliser des enquêtes auprès des supermarchés depuis janvier 2013. Outre les supermarchés, des données

---

<sup>1</sup>Statistics Netherlands, IPC ; Henri Faasdreef 312, 2492 JP La Haye, Pays-Bas ([ag.chessa@cbs.nl](mailto:ag.chessa@cbs.nl)). Les opinions exprimées dans le présent article sont celles de l'auteur et ne reflètent pas nécessairement les politiques de Statistics Netherlands.

<sup>2</sup> En Europe, six pays utilisent des données de lecteurs optiques depuis janvier 2016 (Belgique, Danemark, Pays-Bas, Norvège, Suède et Suisse). L'Islande commencera à utiliser des données de lecteurs optiques dans l'IPC en avril 2016. Les ateliers sur les données de lecteurs optiques qui ont eu lieu à Vienne (2014) et à Rome (2015) ont révélé que plusieurs pays attendent leurs premières données, tandis que d'autres ont pris des mesures concrètes en vue d'acquiescer leurs premières données de lecteurs optiques.

de lecteurs optiques et d'autres données issues de transactions électroniques sont utilisées dans l'IPC, entre autres pour les magasins de bricolage, les pharmacies et les téléphones mobiles. Plus de 20 % de l'IPC des Pays-Bas s'appuie aujourd'hui sur des données de transactions électroniques (sur la base des pondérations Coicop de 2015).

Les méthodes de calcul des indices qui s'appuient sur des données de lecteurs optiques dans l'IPC des Pays-Bas diffèrent selon les détaillants et les biens de consommation. Par exemple, pour les supermarchés, la méthode consiste en un indice de Jevons chaîné mensuellement pour les agrégats élémentaires. Les méthodes utilisées pour d'autres détaillants font appel à des indices de type Laspeyres pour des échantillons d'articles qui représentent une certaine proportion du chiffre d'affaires. Les méthodes adoptées pour différents détaillants découlent de perspectives historiques différentes.

Les méthodes utilisées à l'heure actuelle présentent un certain nombre de problèmes. La méthode appliquée pour les supermarchés utilise des pondérations égales au niveau du GTIN (un seuil de chiffre d'affaires est utilisé afin d'exclure les articles dont les parts du chiffre d'affaires sont inférieures au seuil). Les articles de suivi ne sont pas liés avec leurs prédécesseurs quand le GTIN change. Cela peut se produire dans le cas des « relancements », terme utilisé pour les articles qui réapparaissent dans les magasins après avoir subi des modifications, qui ont fréquemment trait à l'emballage. Le contenu de l'emballage (volume) et les ingrédients demeurent souvent les mêmes. Si l'article de suivi, repositionné, est assorti d'une hausse de prix, celle-ci sera manquée si l'ancien et le nouveau GTIN ne sont pas liés. Les prix de dumping des articles sortants sont supprimés dans la méthode appliquée pour les supermarchés afin de réduire un éventuel biais à la baisse de l'indice des prix (voir de Haan et van der Grient (2011) pour des détails). Les méthodes appliquées aux autres détaillants dans l'IPC des Pays-Bas comprennent l'appariement des GTIN faisant l'objet de relancements, mais ces méthodes s'appuient sur des échantillons d'articles. Certaines de ces méthodes comprennent le remplacement manuel des articles, ce qui peut parfois demander beaucoup de temps.

Les problèmes susmentionnés ont motivé la recherche d'une méthodologie générique pouvant être appliquée aux données de transactions électroniques de différents détaillants et biens de consommation. La section 2 attire l'attention sur l'un des problèmes essentiels du calcul des indices de prix, la définition de produits homogènes. La méthode de calcul des indices est décrite à la section 3. La section 4 résume certains résultats d'une étude comparative destinée à quantifier les effets qu'ont sur les indices de prix les écarts par rapport aux choix faits dans la méthode de calcul de l'indice. Enfin, la section 5 offre un résumé des premières expériences dans l'IPC et des plans pour l'avenir concernant la nouvelle méthodologie.

## **2. Classifications des produits et homogénéité**

Le calcul des indices de prix dans l'IPC suit une série d'étapes d'agrégation qui débute par le calcul des prix moyens au plus fin niveau de détail des produits tel que défini par les agences statistiques. Les prix des produits sont ensuite utilisés pour calculer les indices de prix à un premier niveau d'agrégation, indices qui sont par la suite combinés pour obtenir des indices de prix à des niveaux plus élevés d'une classification emboîtée de groupes de produits. Le plus haut niveau de cette classification est le Système de classification des fonctions de consommation des ménages (Coicop) bien connu. Le plus fin niveau de détail auquel sont publiés les chiffres de l'IPC est appelé « L-Coicop » dans l'IPC des Pays-Bas<sup>3</sup>.

Les données de lecteurs optiques contiennent des données sur les transactions spécifiées par GTIN. Les ensembles de données de lecteurs optiques peuvent contenir de très grands nombres de GTIN, parfois plus de 100 000 pour un seul détaillant. De tels grands nombres requièrent une procédure efficace pour assigner les GTIN aux codes de L-Coicop. Cela peut se faire en utilisant la classification des articles propre aux détaillants (appelée « ESBA » dans l'IPC des Pays-Bas). Les ESBA doivent donc faire partie d'un ensemble de données de lecteurs optiques. Habituellement, nous choisissons les ESBA au plus fin niveau de détail pour établir les liens entre les GTIN et la Coicop. Cependant, les ESBA les plus détaillées peuvent encore couvrir plus d'un code de L-Coicop. Par conséquent, nous créons un niveau inférieur à la L-Coicop, que nous appelons « segments de consommation ». Dans l'IPC des Pays-Bas, les segments de consommation représentent habituellement des classes d'articles similaires, tels que les T-shirts pour hommes, les chaussettes pour femmes, les téléphones mobiles ou le chocolat.

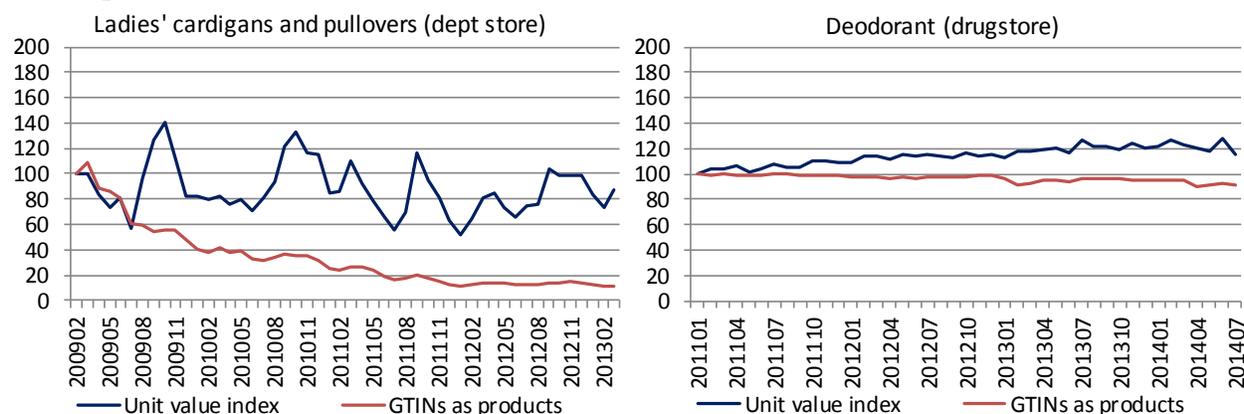
---

<sup>3</sup> Les codes L-Coicop sont spécifiés, au plus, au niveau à cinq chiffres (selon la division de la Coicop).

Les segments de consommation peuvent encore être trop hétérogènes pour pouvoir être considérés comme des produits individuels pour lesquels des prix significatifs peuvent être calculés. Un segment « chaussettes » peut englober différents types de chaussettes (de marche, thermiques, de sport) et les articles peuvent contenir des nombres différents de paires de chaussettes. Les GTIN représentent le niveau le plus détaillé de différenciation des articles, mais peuvent souffrir des relancements de produits. L'effet possible des relancements est illustré à la figure 2-1, en particulier dans le graphique de gauche. L'indice de prix calculé en utilisant les GTIN comme produits uniques tombe presque à zéro après trois ans. L'assortiment de cardigans et de pull-overs, dans le cas d'un grand magasin des Pays-Bas, est presque renouvelé chaque année. Les articles entrent dans le magasin à des prix d'introduction élevés, qui diminuent par la suite au cours de l'année.

**Figure 2-1**

**Indices de valeur unitaire et indices de prix avec les GTIN comme produits pour deux segments de consommation. L'axe des x représente l'année et le mois (aaaaamm). Les indices de prix (axe vertical) sont fixés à 100 le premier mois**



Le résultat est une cascade d'indices qui diminuent annuellement lorsque les anciens et les nouveaux GTIN ne sont pas liés.

La figure 2-1 révèle aussi des écarts importants entre les indices de prix établis en utilisant les GTIN comme produits et les indices de valeur unitaire. Étant donné qu'un aussi grand écart entre les deux niveaux de différenciation des produits peut avoir lieu, une question fondamentale que soulève l'exercice du calcul des indices de prix est celle de savoir comment définir les produits. Comment pourrait-on créer un niveau intermédiaire entre les segments de consommation et les GTIN, de manière que les GTIN combinés à ce niveau intermédiaire puissent être considérés comme étant homogènes? Les groupes de GTIN au niveau intermédiaire seront désignés comme étant des produits homogènes ou simplement des « produits ».

Le problème du relancement d'un article joue un rôle crucial dans la sélection d'une approche pour définir les produits. Les options qui suivent peuvent être identifiées :

1. En l'absence de relancements de produits, les GTIN seraient un choix naturel pour représenter les produits homogènes.
2. En présence de relancements, un niveau plus général de différenciation des produits auquel les GTIN sont combinés est nécessaire. Les options qui suivent peuvent alors être envisagées :
  - a. les anciens et les nouveaux GTIN pourraient être appariés au moyen des codes de produit internes ou numéros de référence d'article (SKU, *Stock Keeping Units*) du détaillant. Les détaillants attribuent habituellement le même SKU aux articles de suivi qu'aux articles qui sortent de l'assortiment<sup>4</sup>;

<sup>4</sup> Statistics Netherlands s'attend à recevoir un ensemble élargi de données de lecteurs optiques provenant d'une chaîne de magasins de bricolage auquel seront ajoutées des SKU. En outre, les données contiendront les dates exactes où l'ancien article sera remplacé par l'article de suivi.

- b. si les SKU ne sont pas disponibles, ou qu'ils ne peuvent pas être utilisés pour une quelconque raison, différents GTIN pourraient être appariés en se fondant sur un ensemble de caractéristiques des produits.

Une question évidente est celle de savoir comment vérifier si des relancements ont lieu ou non, ou s'ils ont lieu dans une mesure où les indices de prix sont à peine affectés. Il est important de calculer des statistiques de base, comme la durée de vie des GTIN, les parts des dépenses imputables aux GTIN qui forment la partie stable d'un assortiment et également aux GTIN sortants. Ces statistiques peuvent être utilisées pour effectuer une première évaluation de l'effet possible des augmentations de prix masquées par des relancements sur un indice de prix.

La disponibilité des SKU résout des problèmes importants : les hausses de prix au moment des relancements seront prises en compte, tandis que les techniques pour déterminer les caractéristiques des produits à partir de l'information obtenue sur les produits deviennent superflues. La détermination des caractéristiques n'est pas un problème lorsque celles-ci sont consignées dans des champs distincts dans les ensembles de données de lecteurs optiques. Cependant, si les caractéristiques des produits sont incluses dans des chaînes de texte dans les descriptions des articles, des techniques de forage de textes pourraient être utilisées afin de déterminer les caractéristiques. Des termes de recherche doivent être sauvegardés dans des listes, qui doivent être mises à jour chaque mois.

Si les SKU ne sont pas disponibles, les GTIN peuvent être appariés en fonction des caractéristiques des produits. Les GTIN sont appariés et combinés en un même produit quand ils ont en commun les mêmes caractéristiques. La question est alors de savoir quelles caractéristiques doivent être choisies. Chessa (2016) propose une analyse de sensibilité consistant à quantifier l'effet de l'ajout de caractéristiques sur un indice de prix. Ce genre d'analyse pourrait être utilisé pour établir des ensembles minimums de caractéristiques qui pourraient aussi permettre de restreindre le temps consacré à la mise à jour mensuelle de ces listes.

### 3. La méthode de calcul de l'indice

La méthode de calcul de l'indice décrite à la présente section est destinée à être appliquée aux segments de consommation. Les indices de prix calculés pour ces segments seront agrégés aux niveaux de la Coicop conformément à des méthodes conventionnelles de type Laspeyres<sup>5</sup>. La figure 3-1 résume les méthodes de calcul d'indice des prix et les concepts de prix utilisés à différents niveaux de groupe d'articles dans l'IPC.

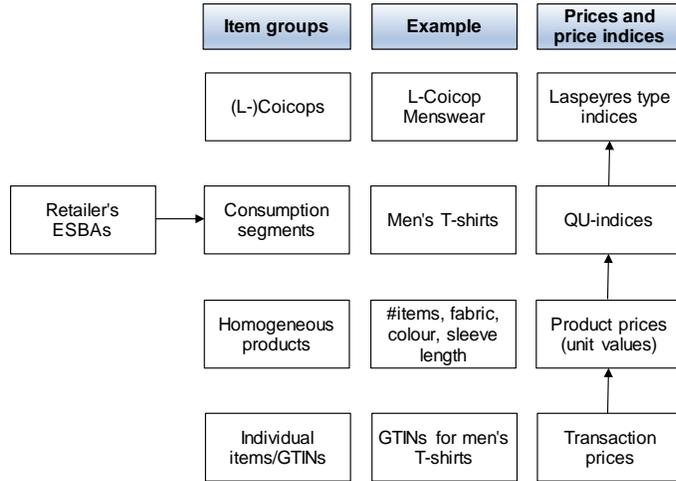
Le choix de la méthode de calcul de l'indice au niveau des segments de consommation part du cas simplifié où les produits dans un segment de consommation sont équivalents. Dans ce cas, nous pouvons calculer les valeurs unitaires pour les segments de consommation, c'est-à-dire le chiffre d'affaires total divisé par la somme des quantités vendues sur l'ensemble des produits compris dans les segments (voir le manuel de l'IPC (2004, p. xxii)). L'indice de prix est alors égal à un indice de valeur unitaire, qui peut s'écrire sous la forme du ratio d'un indice du chiffre d'affaires divisé par un indice de quantité non pondérée.

Quand les articles compris dans un segment de consommation ne sont pas homogènes, le segment doit être différencié en produits homogènes. L'indice de valeur unitaire ne peut pas être appliqué dans ce cas et doit également être perfectionné. Le perfectionnement s'applique simplement à l'indice de quantité (car évidemment, le chiffre d'affaires demeure inchangé). Nous préservons la forme additive des mesures de quantité, mais les quantités vendues  $q_{i,t}$  du produit  $i$  durant chaque mois  $t$  sont maintenant pondérées par un facteur  $v_i$ .

---

<sup>5</sup> Il est également possible d'étendre la méthode d'agrégation utilisée à l'intérieur des segments de consommation à des niveaux plus élevés. Cela aboutirait à une méthode plus cohérente. Cependant, nous réservons cette autre méthode d'agrégation pour de futurs travaux de recherche.

**Figure 3-1**  
**La nouvelle méthodologie (« QU ») dans l'IPC**



L'indice de prix  $P_t$  du mois  $t$  par rapport à un mois de base ou de référence 0 prend la forme suivante :

$$P_t = \frac{\sum_{i \in G_t} p_{i,t} q_{i,t} / \sum_{i \in G_0} p_{i,0} q_{i,0}}{\sum_{i \in G_t} v_i q_{i,t} / \sum_{i \in G_0} v_i q_{i,0}}, \quad (1)$$

où  $p_{i,t}$  désigne le prix (valeur unitaire) du produit  $i$  au mois  $t$  et  $G_t$  désigne l'ensemble de produits vendus au cours du mois  $t$  pour un segment de consommation  $G$ . Les ensembles  $G_t$  et  $G_0$  durant le mois  $t$  et le mois de référence 0 peuvent être différents.

La formule (1) peut être réécrite en divisant les numérateurs et les dénominateurs de l'indice du chiffre d'affaires et de l'indice de quantité pondérée par les sommes des quantités vendues sur l'ensemble des produits durant les mois  $t$  et 0 respectivement, de sorte que :

$$P_t = \frac{\bar{p}_t / \bar{p}_0}{\bar{v}_t / \bar{v}_0}. \quad (2)$$

Le numérateur de (2) est égal à l'indice de valeur unitaire. Il s'agit de l'indice de prix sous homogénéité du segment de consommation  $G$ . Sinon, l'indice de valeur unitaire doit être ajusté, et le terme d'ajustement est donné par le dénominateur de (2). Ce dernier est un ratio des valeurs moyennes des  $v_i$  aux mois  $t$  et 0, qui sont pondérées par les parts des quantités vendues de chaque produit. Le ratio  $\bar{v}_t / \bar{v}_0$  peut être considéré comme un indice de qualité. Ses valeurs sont déterminées par deux facteurs, à savoir les valeurs relatives des  $v_i$  des produits et les parts des quantités vendues de chaque produit durant les deux mois. Par exemple, un déplacement vers des produits ayant un  $v_i$  plus élevé (meilleure qualité) durant le mois  $t$  se traduit par un indice de qualité supérieur à 1, ce qui mène à un ajustement à la baisse de l'indice de valeur unitaire. Étant donné cette propriété de la formule d'indice (2), nous parlons d'un « indice de valeur unitaire ajusté pour la qualité », ou « indice QU » (de l'anglais *quality-adjusted unit value index*).

L'expression (1) peut être considérée comme une famille de méthodes de calcul d'indices. Différents choix pour les  $v_i$  mènent à différentes formules d'indice. Par exemple, si nous prenons les  $v_i$  égaux aux prix des produits du mois de référence, alors (1) devient un indice de Paasche; si nous prenons les  $v_i$  égaux aux prix du mois  $t$ , alors (1) devient un indice de Laspeyres; si nous prenons la moyenne des prix des mois 0 et  $t$ , nous obtenons l'indice de Fisher comme cas particulier (voir aussi Chessa (2016, p. 18)).

Les trois exemples correspondent tous à des indices bilatéraux. Les indices-chaînes mensuels ne permettent pas de traiter les variations de prix des produits qui ne sont pas disponibles temporairement, tandis que les indices directs bilatéraux ne permettent pas de traiter les nouveaux produits sans délai, à moins de procéder à une certaine forme

d'imputation de prix. Cependant, les imputations de prix ne sont pas nécessaires quand les prix des produits de chaque mois durant un intervalle de temps  $T$  sont utilisés pour construire les  $v_i$ . Cela nous permet d'établir une méthode de calcul de l'indice selon laquelle toutes les transactions peuvent être traitées. La dynamique d'un assortiment, dans lequel des articles peuvent disparaître, de nouveaux articles peuvent entrer, tandis que d'autres articles persistent, peut alors être reflétée complètement par le comportement des indices de prix et de quantité au fil du temps.

Puisque les  $v_i$  font partie de l'indice de quantité, les variations de prix doivent être éliminées des  $v_i$ . Les prix des produits dans les  $v_i$  sont par conséquent corrigés de l'inflation au moyen de l'indice de prix lui-même. De toute évidence, on peut imaginer de nombreux moyens de combiner les prix corrigés de l'inflation provenant de plusieurs mois pour obtenir les pondérations des produits  $v_i$ . Dans le présent article, nous considérons des moyennes arithmétiques pondérées :

$$v_i = \sum_{z \in T} \varphi_{i,z} \frac{p_{i,z}}{P_z}, \quad (3)$$

où les pondérations  $\varphi_{i,z}$  sont non-négatives et leur somme est égale à 1 sur l'ensemble des  $z \in T$ .

Différentes questions doivent être résolues avant de pouvoir calculer les indices de prix :

1. Comment faut-il choisir les pondérations  $\varphi_{i,z}$ ?
2. Comment faut-il choisir l'intervalle  $T$ ?
3. Comment faut-il mettre à jour les  $v_i$  chaque mois?

Notre choix de base pour les pondérations  $\varphi_{i,z}$  est la part des quantités vendues au cours du mois  $z$  dans la somme des quantités sur une période complète  $T$ , c'est-à-dire :

$$\varphi_{i,z} = \frac{q_{i,z}}{\sum_{s \in T} q_{i,s}}. \quad (4)$$

Si nous remplaçons la période par le pays, les expressions (1) à (4) deviennent équivalentes à la méthode de Geary-Khamis utilisée pour les comparaisons internationales de prix (Khamis, 1972). Cette méthode a suscité un certain débat dans le contexte des comparaisons internationales de prix, au sujet d'un phénomène appelé « biais de substitution » (Balk, 2001). Cependant, nous pensons que ce phénomène ne sera guère un problème dans les comparaisons entre périodes. Les consommateurs ont tendance à acheter de plus grandes quantités à plus faibles prix, par exemple, à rabais. L'expression (4) attribue des pondérations plus importantes aux prix correspondant à des cas de ce genre. À la section 4, d'autres choix pour les pondérations  $\varphi_{i,z}$  seront examinés afin de quantifier l'effet de différents choix sur les résultats.

L'intervalle de temps  $T$  représente l'année courante de publication, qui couvre une période de 13 mois englobant aussi décembre de l'année précédente comme mois de base. Ce choix implique que les  $v_i$  soient constants durant une année, du moins en théorie, et qu'ils puissent varier d'une année à l'autre. Le choix de l'année courante découle du fait que l'on vise à inclure également les nouveaux produits dans les calculs de l'indice. L'adoption d'une durée d'un an et de décembre comme mois de base concorde avec les pratiques liées à l'IPC.

L'utilisation des prix et des quantités de produits provenant de l'année courante introduit un problème supplémentaire, car les prix et les quantités pour 13 mois ne sont connus qu'à la fin de l'année. Les  $v_i$  doivent donc être mis à jour au moyen des prix et des quantités du mois suivant. L'utilisation d'une fenêtre de 13 mois glissante a donné des résultats volatiles et apparemment biaisés (Chessa, 2015). Nous utilisons donc plutôt une fenêtre temporelle qui est élargie chaque mois par rapport à un mois de base fixé annuellement et mettons à jour les  $v_i$  conformément aux expressions (3) et (4), en remplaçant  $T$  par l'intervalle  $[0, t]$ , où le mois 0 désigne le mois de base et le mois  $t$ , le mois courant. Les  $v_i$  mis à jour mensuellement sont utilisés pour calculer un indice de prix pour le mois  $t$  par rapport au mois de base. Les indices de prix durant le mois final d'une année sont fondés sur les  $v_i$  dans lesquels sont utilisés les prix et les quantités des produits pour l'ensemble des 13 mois. Si ces  $v_i$  étaient utilisés chaque mois d'une année, ce qui est impossible en pratique, les indices résultants seraient transitifs. Par conséquent, les « indices en temps réel » avec mise à jour mensuelle des  $v_i$  sont exempts du problème de biais de chaînage (*chain drift*) par construction. Une question intéressante est celle de savoir comment les « indices de référence » en temps réel et transitifs se comparent au cours d'une année. Nous examinerons cette question à la section 4 également.

Comme l'indice de prix figure dans l'expression (3) pour les  $v_i$ , la question est de savoir comment les indices de prix peuvent être calculés. Une méthode qui est simple à appliquer s'appuie sur un schéma itératif qui débute par des valeurs initiales arbitraires pour les indices de prix pour chaque mois entre le mois de base et le mois courant. Les valeurs initiales sont ensuite utilisées pour calculer les valeurs des  $v_i$  selon l'expression (3), qui sont par la suite utilisées pour recalculer les indices de prix selon la formule (1). Cette procédure converge vers une solution unique, quand elle existe<sup>6</sup>, et est répétée jusqu'à ce que les écarts entre les valeurs des indices de prix des deux dernières étapes d'itération soient « suffisamment petits » (l'utilisateur doit établir un critère d'arrêt). Il est important de noter que les indices de prix des mois précédents sont recalculés, mais ne sont pas utilisés pour réviser les indices de prix qui sont déjà publiés.

Au lieu d'utiliser des valeurs initiales arbitraires pour les indices de prix, on peut construire des valeurs initiales en utilisant une méthode en deux étapes qui fait abstraction des prix et des quantités du mois courant à la première étape. L'indice de prix « de première étape » résultant est alors utilisé pour corriger les prix et les quantités du mois courant, qui sont utilisés pour mettre à jour les  $v_i$  et l'indice des prix. Cette méthode a donné des résultats très précis (Chessa, 2015), ce qui permet de réduire la durée des calculs. Contrairement à la méthode itérative complète, les indices des prix des mois précédents ne sont pas recalculés à la phase initiale de la méthode en deux étapes.

#### 4. Effet de différents choix sur les indices de prix

À la présente section, nous comparons différents choix méthodologiques comme il a été mentionné à la section précédente, et nous quantifions leur effet sur les indices de prix. À cette fin, nous avons utilisé des ensembles de données de lecteurs optiques fournis par un grand magasin et par une chaîne de pharmacies des Pays-Bas. Pour le calcul des indices, nous avons utilisé des données historiques couvrant une période de 4 ans pour le grand magasin et de 3 ans et 7 mois pour les pharmacies.

Les deux ensembles de données de lecteurs optiques contiennent les prix hebdomadaires et les quantités hebdomadaires vendues au niveau du GTIN pour différents groupes de produits. Le programme de l'IPC des Pays-Bas utilise les trois premières semaines complètes d'un mois pour calculer les prix mensuels moyens des produits (chiffre d'affaires divisé par les quantités vendues sur les trois semaines). Les ensembles de données contiennent aussi les caractéristiques de chaque article. Les numéros de référence d'article (SKU) ne sont pas disponibles dans ces deux ensembles de données. Pour le grand magasin, les produits sont définis conformément à un ensemble de caractéristiques pour les vêtements (type d'article, tissu, couleur, nombre d'articles dans un emballage), tandis que les GTIN ont été considérés comme étant les produits pour le reste de l'assortiment. Les GTIN n'ont pas été utilisés pour caractériser les produits des pharmacies, en raison de la portée des relancements de produit dans l'assortiment. Les caractéristiques ont par conséquent été utilisées pour différencier les produits pour l'assortiment complet.

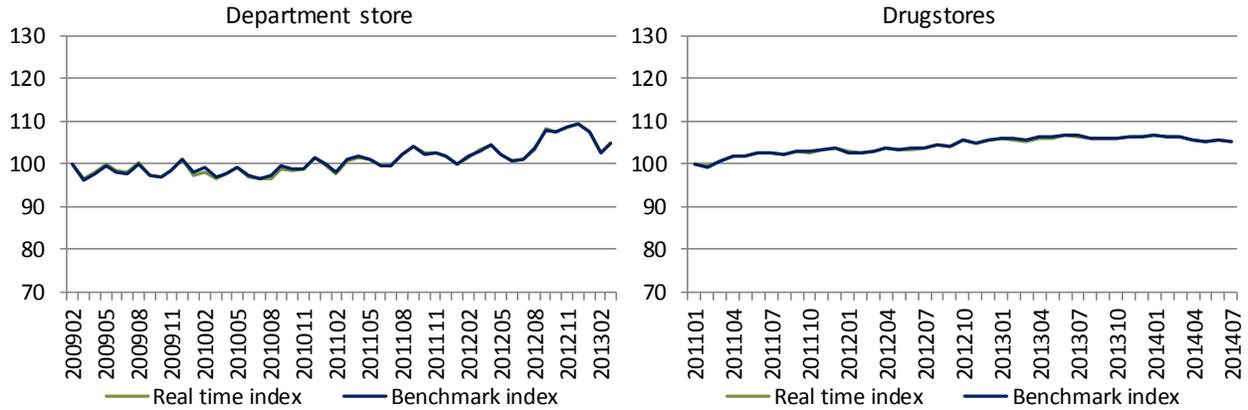
Ci-dessous, nous comparons les indices de prix par rapport à différents aspects méthodologiques. Étant donné la brièveté du présent article, il a été décidé de montrer les indices de prix à un niveau agrégé pour le grand magasin ainsi que pour la chaîne de pharmacies. La figure 4-1 donne une comparaison des indices en temps réel avec les indices de référence transitifs pour le grand magasin et la chaîne de pharmacies. Les écarts sont négligeables, et sont également petits ou négligeables à de plus faibles niveaux d'agrégation (Chessa, 2016). Les résultats montrent que les indices en temps réel se comportent presque comme s'ils étaient des indices transitifs.

---

<sup>6</sup> Dans le contexte des pratiques de calcul de l'IPC, cela revient à vérifier pour chaque mois de publication s'il existe un produit qui a été vendu à la fois dans le mois courant et dans un des mois précédents. Si ce n'est pas le cas, l'indice de prix du segment de consommation sera imputé pour le mois de publication (p. ex., en partant du niveau L-Coicop correspondant).

**Figure 4-1**

**Indices en temps réel et indices de référence pour un grand magasin (février 2009 = 100) et une chaîne de pharmacies (janvier 2011 = 100) des Pays-Bas au niveau global**

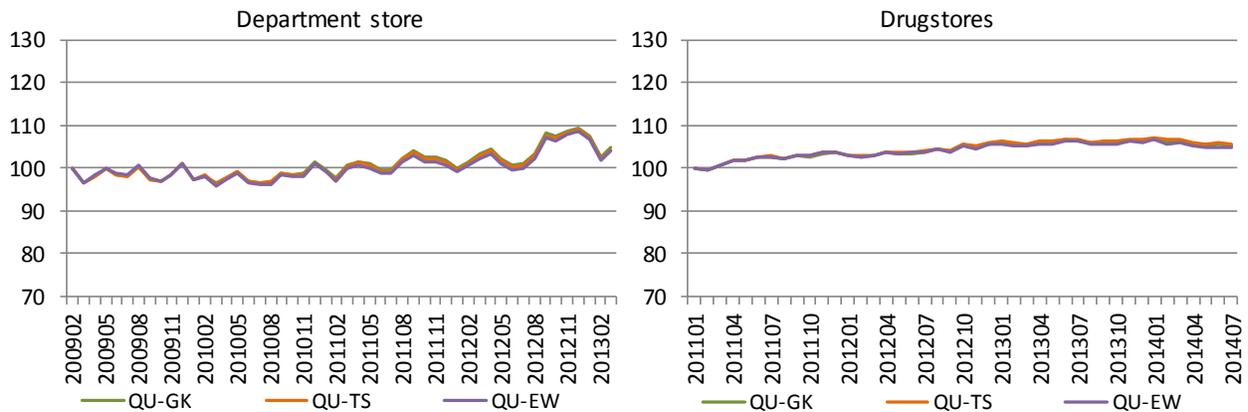


À la figure 4-2, les indices en temps réel sont comparés aux indices QU qui sont calculés en utilisant deux schémas de pondération différents pour les prix corrigés de l'inflation dans les  $v_i$ . L'un des schémas consiste à utiliser des pondérations égales (QU-EW) pour les  $\varphi_{i,z}$  dans l'expression (3). Les prix sont écartés quand les quantités diminuent de plus de 90 %. Le second schéma consiste à utiliser les parts du chiffre d'affaires (QU-TS) au lieu des quantités vendues comme dans (4). Ces deux variantes résultent en des indices de prix qui diffèrent à peine des indices de la figure 4-1 (que nous désignons par QU-GK). Les résultats semblent donc être robustes sous différents schémas de pondération dans les  $v_i$ . Cette observation tient également à de plus faibles niveaux d'agrégation (Chessa, 2016).

La méthode QU-GK a également été comparée à ce qu'il est convenu d'appeler la « time product dummy method » (méthode des variables indicatrices période-produit) (de Haan et Krsinich, 2014). Comme la méthode QU-GK, la time product dummy method (méthode TPD en abrégé) est une adaptation d'une méthode appliquée pour les comparaisons internationales de prix (« country product dummy method ») aux comparaisons entre périodes. Les indices de prix selon la méthode TPD peuvent être exprimés sous la forme d'un rapport de moyennes géométriques pondérées des « prix ajustés pour la qualité »  $p_{i,t}/v_i$  et  $p_{i,0}/v_i$  pour deux mois  $t$  et  $0$ , lesquels sont pondérés par les parts du chiffre d'affaires de chaque produit dans les deux mois. La méthode QU-GK peut s'écrire de manière similaire, ce qui donne un ratio de moyennes harmoniques pondérées des prix ajustés pour la qualité dans deux périodes, en prenant aussi pour pondération les parts du chiffre d'affaires de chaque produit.

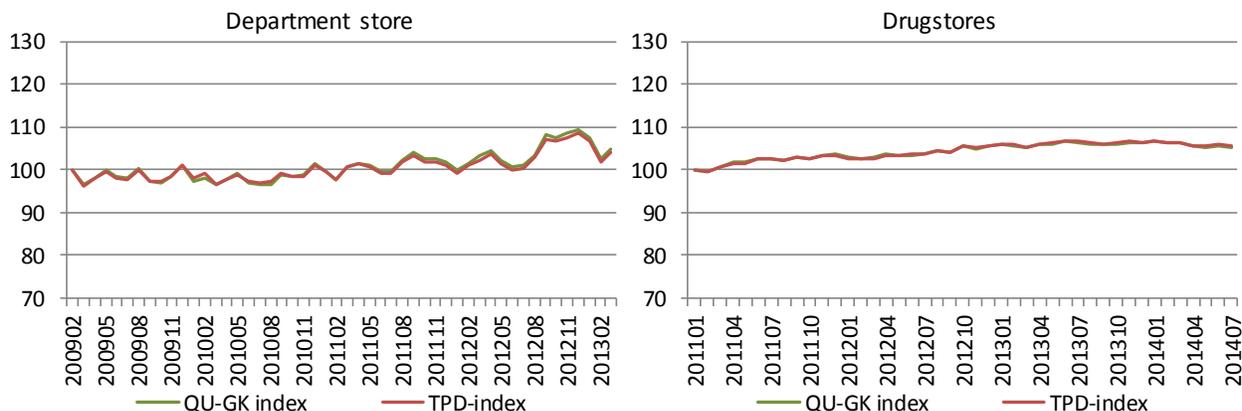
**Figure 4-2**

**Indices en temps réel de la figure 4-1 (QU-GK) comparés aux variantes QU-TS et QU-EW de la méthode QU**

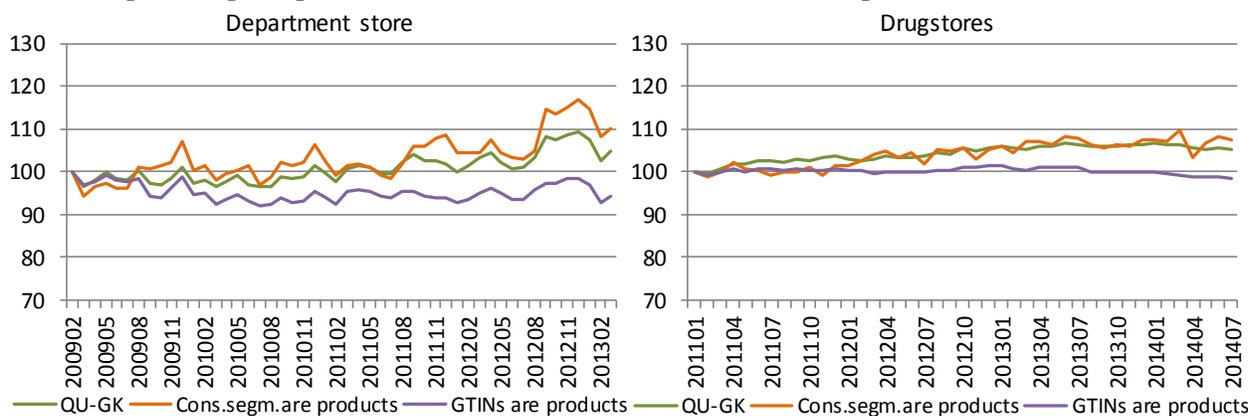


Les indices QU-GK et TPD sont illustrés à la figure 4-3 sous forme d'indices de prix globaux pour le grand magasin et la chaîne de pharmacies des Pays-Bas. Les indices de prix obtenus par les deux méthodes diffèrent à peine, constatation qui s'applique également à la plupart des agrégats de la Coicop et aux segments de consommation sous-jacents. La figure 4-4 compare les indices de prix obtenus pour les choix que nous avons faits concernant la différenciation des produits pour le grand magasin et les pharmacies en utilisant deux niveaux extrêmes de différenciation de produits. D'une part, les segments de consommation sont considérés comme des produits homogènes pour lesquels des valeurs unitaires sont calculées. À l'autre extrême, chaque GTIN est considéré comme un produit distinct.

**Figure 4-3**  
**La méthode QU-GK comparée à la méthode TPD**



**Figure 4-4**  
**Indices de prix comparés pour deux niveaux extrêmes de différenciation de produits**



La figure 4-4 révèle des écarts importants entre les trois indices. Les indices de prix en prenant les GTIN comme des produits sont ceux qui s'écartent le plus des indices QU-GK. Considérer les GTIN comme représentants des produits donne des indices qui se situent en dessous des indices QU-GK pour les deux détaillants, comme prévu, puisque les augmentations de prix après les relancements ne sont pas prises en compte. L'indice basé sur les valeurs unitaires se situe au-dessus de l'indice QU-GK pour le grand magasin. Cela implique que le dénominateur de l'expression (2) a augmenté, ce qui signifie que les habitudes de consommation ont évolué vers des produits de plus grande qualité au cours du temps.

## 5. Premières expériences concernant l'IPC et remarques finales

Une nouvelle méthodologie de traitement des ensembles de données issues de transactions électroniques et de calcul des indices de prix est proposée dans le présent article. L'approche offre plusieurs avantages. Elle permet de traiter

toutes les transactions, en incluant sans délai les nouveaux produits au cours d'une année. Par conséquent, la méthode de calcul de l'indice ne nécessite pas d'imputations de prix. Des filtres artificiels pour les prix de dumping et les faibles parts du chiffre d'affaires ne sont pas nécessaires. La méthode de calcul de l'indice s'appuie sur les prix et les quantités des articles vendus pour calculer les pondérations des produits  $v_i$ , et les anciens et nouveaux GTIN sont reliés, en se basant sur les caractéristiques de l'article ou sur les numéros de référence d'article (SKU).

Les résultats présentés à la section 4 montrent que les indices de prix sont remarquablement robustes sous divers choix de pondérations dans les  $v_i$  et de forme de la formule d'indice des prix. L'indice en temps réel, avec les  $v_i$  mis à jour mensuellement, diffère à peine d'un indice de référence transitoire. La méthode proposée est exempte de biais de chaînage (*chain drift*) et l'utilisation des données sur les prix et les quantités provenant d'un petit nombre de mois au début de chaque année ne pose pas problème.

L'étude comparative de la section 4 montre que le problème de la définition de produits homogènes est, de loin, l'aspect le plus délicat et mérite donc une attention particulière. Les résultats indiquent à quel point il est important que les ensembles de données de lecteurs optiques contiennent de l'information au sujet des caractéristiques pertinentes des produits (de préférence dans des colonnes distinctes afin de limiter les techniques de forage de texte qui prennent beaucoup de temps). Les organismes statistiques devraient demander aux détaillants s'ils peuvent fournir leurs propres codes de produit internes. Ces codes peuvent offrir un moyen efficace de relier les GTIN sortants aux articles de suivi, donc saisir les augmentations de prix après les relancements sans utiliser les caractéristiques du produit (section 2).

La nouvelle méthodologie fait maintenant partie de l'IPC des Pays-Bas. En janvier 2016, la nouvelle approche a été appliquée pour la première fois aux téléphones mobiles. La production mensuelle se déroule efficacement : cela prend de 30 à 45 minutes pour préparer les données d'entrée pour les calculs de l'indice (principalement pour recueillir les données sur les caractéristiques des appareils qui manquent dans les données). La nouvelle approche a permis de réaliser une énorme économie de temps, comparativement aux 2 à 3 jours que nécessitaient les méthodes antérieures.

Les données de lecteurs optiques fournies par le grand magasin seront traitées par la méthode QU dans l'IPC d'ici quelques mois. Les données des pharmacies et celles des magasins de bricolage suivront à une étape ultérieure. Au cours de 2016, un projet de recherche portant sur les données de lecteurs optiques des supermarchés débutera en vue de comparer la méthode QU à la méthode utilisée à l'heure actuelle pour les supermarchés dans l'IPC des Pays-Bas.

## Bibliographie

- Balk, B.M. (2001), "Aggregation Methods in International Comparisons: What Have We Learned?", paper originally prepared for the Joint World Bank - OECD Seminar on Purchasing Power Parities, 30 January - 2 February 2001, Washington DC.
- Chessa, A.G. (2015), "Towards a Generic Price Index Method for Scanner Data in the Dutch CPI", Ottawa Group Meeting, 20-22 May 2015, Urayasu City, Japan.
- Chessa, A.G. (2016), "Processing Scanner Data in the Dutch CPI: A New Methodology and First Experiences", paper presented at the Meeting of the Group of Experts on Consumer Price Indices, 2-4 May 2016, Geneva, Switzerland. (To be published in *EURONA*, June 2016.)
- CPI-manual (2004), *Consumer price index manual: Theory and practice*, Geneva, Switzerland: ILO/IMF/OECD/UNECE/Eurostat/The World Bank.
- de Haan, J., and H.A. van der Grient (2011), "Eliminating Chain Drift in Price Indexes Based on Scanner Data", *Journal of Econometrics*, 161, pp. 36-46.
- de Haan, J., and F. Krsinich (2014), "Time Dummy Hedonic and Quality-Adjusted Unit Value Indexes: Do They Really Differ?", paper presented at the Society for Economic Measurement Conference, 18-20 August 2014, Chicago, U.S.
- Khamis, S. H. (1972), "A New System of Index Numbers for National and International Purposes", *Journal of the Royal Statistical Society A*, 135, pp. 96-121.