

Mesure de l'inflation avec des données de caisse et un panier fixe évolutif

Inflation Measurement with Scanner Data and an Ever-Changing Fixed Basket

Can Tongur*

Résumé – Statistics Sweden a introduit les données de caisse dans certaines parties de l'indice des prix à la consommation (IPC) il y a plusieurs années, avec la préoccupation de s'assurer de la comparabilité au fil du temps et entre pays. Cet article traite de la préservation de l'approche du panier fixe et interroge la pertinence de la méthode classique du remplacement manuel d'articles, accompagné d'ajustements de la qualité et de la quantité, pour assurer la comparabilité malgré le changement de mode de collecte et l'étendue des données. Les biais de l'IPC dus à des ajustements erronés de la quantité sont analysés et illustrés avec des exemples numériques basés sur des évolutions réelles du marché suédois des produits de consommation courante. Les ajustements manuels de la qualité et de la quantité sont mis en œuvre sur un échantillon aléatoire restreint d'articles représentatifs, c'est-à-dire un panier fixe, ce qui engendre de l'imprécision ou de la variance dans l'IPC. Cette approche pourrait être mise en question compte tenu du caractère de quasi-recensement des données de caisse, et le compromis biais-variance est donc analysé dans cette étude. La variance liée à la taille de l'échantillon est estimée avec une méthode *jackknife* et comparée avec des ajustements de la qualité/quantité.

Abstract – Statistics Sweden introduced scanner data into parts of the consumer price index several years ago, with the concern to ensure comparability over time and between countries. In this article, we discuss the issue of preserving the fixed basket approach and whether the traditional manual item replacement strategy, with quality and quantity adjustments, is still a relevant method to ensure comparability despite the change in data collection mode and extensiveness of data. Biases from improper quantity adjustments are discussed and illustrated through numeric examples based on real changes in the Swedish market of daily necessity products. Manual adjustments of quality and quantity are implemented by following a small random sample of representative items, i.e. a fixed basket, which therefore leads to imprecision or variance in the consumer price index. This may be a questionable approach given the availability of census-like scanner data, thus the bias-variance trade off is addressed. The sample size related variance is estimated through a jackknife method and contrasted with quality/quantity adjustments.

Codes JEL / JEL Classification : E31, C15, C83, C80

Mots-clés : données de caisse, indice des prix à la consommation, IPC, panier fixe, inflation cachée, variance *jackknife*

Keywords: scanner data, consumer price index, CPI, fixed basket, hidden inflation, jackknife variance

* Statistics Sweden (Can.Tongur@scb.se)

L'auteur remercie Anders Norberg, consultant en chef auprès de Statistics Sweden, pour sa contribution. Cet article a été largement amélioré grâce aux suggestions des rapporteurs anonymes de la revue.

Reçu le 31 juillet 2017, accepté après révisions le 4 juillet 2018

Traduit de la version originale en anglais

Rappel :

Les jugements et opinions exprimés par les auteurs n'engagent qu'eux mêmes, et non les institutions auxquelles ils appartiennent, ni a fortiori l'Insee.

Les données de caisse du commerce de détail ont été introduites dans l'indice des prix à la consommation (IPC) suédois en 2012 et concernaient initialement les produits de consommation courante. À cette époque, le Statistiska centralbyrån (Statistics Sweden) ne se posait pas de question conceptuelle sur le volume de données à utiliser. Le processus a commencé par le remplacement, un à un, des prix collectés manuellement par les données de caisse du seul détaillant à les fournir, en conservant la structure de l'échantillon pour les magasins et les articles. Avant que les données de caisse n'entrent dans la production de l'IPC, plusieurs études internes avaient été menées afin de s'assurer que cette nouvelle source de données était conforme à une attente de base, l'absence d'impact défavorable sur l'IPC.

Au fil du temps, le volume des données de caisse incluses dans l'indice et le nombre de détaillants fournissant gracieusement ces données ont augmenté. En termes de chiffre d'affaires, les données de caisse couvrent désormais plus de 80 % du marché suédois des produits de consommation courante¹. Conséquence positive de cette expérience menée dans le secteur des produits de consommation courante, d'autres parties de l'IPC suédois sont maintenant produites à l'aide des nouvelles sources de données de transactions. Malgré la hausse des volumes de données pouvant être utilisés, surtout dans le secteur des produits de consommation courante, la production de l'IPC suédois se poursuit conformément à une stratégie bien établie d'échantillonnage des produits et des magasins. La stratégie d'échantillonnage, pour l'essentiel indépendante du mode de collecte des données, a été adaptée et l'introduction de cette nouvelle source de données alternative, extrêmement prometteuse, n'a entraîné que des changements méthodologiques mineurs – et de simples petites divergences.

À l'ère des *Big Data*, dont la popularité se répercute sur la méthodologie statistique, cette position quelque peu prudente de Statistics Sweden peut être questionnée : pourquoi ne pas utiliser en continu la totalité, ou le plus grand nombre possible, de données, ce qui semblerait intéressant et plus au goût du jour ? On s'intéresse dans cet article à la question de la préservation de la méthodologie conventionnelle de production de l'IPC en présence des données de caisse. Statistics Sweden a choisi d'utiliser ces données tout en maintenant l'approche conventionnelle, afin de traiter de façon satisfaisante le

phénomène des relances commerciales. En effet, lors de ces relances, certaines caractéristiques d'un produit changent (par exemple sa taille). Le prix du nouveau produit, bien que quasiment semblable au produit antérieur à la relance, doit être ajusté en fonction du changement de quantité afin de préserver la comparabilité au fil du temps. L'impact d'une évaluation erronée des ajustements qualité/quantité sera examiné, dans le cadre de l'utilisation de paniers « automatiques » (*infra*) avec les données de caisse.

Cet article vise à étudier le compromis entre la précision de l'inflation mesurée et le biais qui survient lorsque les ajustements de la quantité sont ignorés. Bien qu'il se concentre sur les produits de consommation courante, l'analyse est également valide pour l'IPC global.

L'article s'articule comme suit. La prochaine section donne une vue d'ensemble de l'utilisation des données de caisse par Statistics Sweden dans la production de l'IPC. Il s'agit d'une section descriptive de ce mode de collecte des données relativement nouveau, et elle s'adresse principalement aux lecteurs peu familiers du sujet. Dans la section suivante, un estimateur de variance *jackknife* est appliqué dans un cadre simplifié afin d'évaluer la variance des indices de prix à la consommation avec des données de caisse. Nous examinons ensuite la question des ajustements qualité/quantité, qui est décrite et illustrée par des exemples basés sur des changements réellement survenus sur le marché suédois des produits de consommation courante. L'article se termine par des remarques générales et une mise en perspective des résultats présentés.

Les données de caisse pour les produits de consommation courante dans l'IPC suédois

Cette section présente certaines des questions méthodologiques qui ont dû être traitées avant de mettre en œuvre les données de caisse. Mais elle fait d'abord un petit détour par la terminologie et quelques éléments sur l'arrivée des données de caisse il y a une vingtaine d'années.

1. Les statistiques de marché peuvent être obtenues auprès de l'Institut de recherche commerciale suédois Handels Utredningsinstitut (HUI Research, 2017).

Données de caisse, données de transactions et données massives

Dans le contexte de la vente de biens de consommation, les « données de caisse » sont une expression, peut-être un peu maladroite, désignant les « données de transactions » dans le cadre des ventes au consommateur². Le terme « caisse » vient des codes-barres³ collés sur les emballages, qui sont scannés à la caisse du magasin afin d'enregistrer les articles au point d'achat. Le terme plus général de « données de transactions » peut être utilisé de façon interchangeable dans la mesure du possible, car il a une portée plus large : il couvre les données numériques des ventes et de la consommation de biens et de services. Les données de transactions relatives aux ventes sont, pour la plupart, des données bien structurées découlant d'un système commercial, et ne doivent pas être confondues avec, par exemple, les « données en masse » non structurées. Les données de transactions peuvent être de grande envergure, à haute fréquence (et elles peuvent être obtenues quasiment en temps réel), et ont un caractère proche des données administratives, en ce sens qu'elles ne sont pas destinées à établir de statistiques publiques, mais plutôt destinées à la gestion des stocks ou au suivi des ventes ou des bénéfices.

L'arrivée des données de caisse dans l'IPC suédois

Les données de caisse ne sont pas totalement nouvelles pour Statistics Sweden. Au milieu des années 1990, alors que les données numériques faisaient leur apparition, des contacts avaient été pris avec des analystes du marché suédois afin d'avoir un premier aperçu sur cette nouvelle source de données prometteuses – dont les avantages potentiels pour l'IPC étaient évidents et attrayants. Néanmoins, ces données étaient très chères et donc hors de portée pour une agence gouvernementale dans le contexte de la crise économique nationale la plus sévère depuis la guerre (pour plus de détails économiques et politiques, voir par exemple Bäckström, 1997, ou Englund, 2015). Aujourd'hui, une bonne vingtaine d'années plus tard, cette source de données fait partie intégrante et constitue une étape naturelle de la collecte de données mensuelle effectuée pour l'IPC suédois, et Statistics Sweden reçoit à titre gracieux les données de nombreux détaillants sur la base d'accords bilatéraux à but non lucratif. Cela ne concerne que l'échantillon de magasins inclus dans l'IPC durant une année donnée. Dans la mesure où les enseignes fournissent leurs données gratuitement, Statistics Sweden a limité

ses demandes, ce qui est par ailleurs un facteur de confiance car les détaillants ne fournissent pas d'informations commerciales exhaustives à haute fréquence.

Panier de l'IPC, données relatives aux transactions et exceptions

Le panier de l'IPC

Le panier de l'IPC est présenté au tableau 1 selon la nomenclature internationale COICOP⁴ à deux chiffres. Les prix sont collectés pour des produits spécifiques au sein de ces catégories de consommation. Plusieurs étapes de calcul interviennent entre la valeur totale de l'IPC et les produits définis – l'IPC est tout simplement une hiérarchie dans laquelle les données sur les prix sont agrégées par étape.

Un produit donné chez un détaillant donné, qui fait l'objet de la mesure du prix, est dénommé « offre de produit ». Les prix observés sont regroupés grâce à des formules d'indice et selon des poids infra-annuels fixes pour les groupes de produits, qui sont souvent des indices de premier niveau, c'est-à-dire des agrégats élémentaires. Un exemple de groupe de produits est le lait : les prix des différentes variétés de toutes les marques, tous les magasins et tous les types de teneur en matière grasse sont regroupés dans un même groupe de produits. Il en va de même pour les boissons gazeuses, avec ou sans sucre et toutes tailles confondues.

Le poids d'un groupe de produits reflète la part que ce groupe occupait dans la consommation privée l'année complète précédant l'année de base de l'indice, considérée comme période de référence. L'année de base de l'indice est décembre de l'année $y-1$ et les mois en cours utilisés pour relever les prix correspondent à l'année y , de sorte que les poids s'entendent (normalement) à partir de l'année $y-2$ pour l'indice mensuel. L'IPC est une série d'indices chaînés s'étalant sur plusieurs années, et une partie de la discussion présentée dans cet article concerne le chaînage mensuel (infra-annuel).

2. *Le Manuel de l'IPC fait la différence entre les données de caisse et les données recueillies au point de vente électronique (§6.117, ILO 2004), ce que cet article ne fait pas.*

3. *Le code-barres relie l'article, par le biais de son emballage, à un numéro d'article distinct fourni par un acteur du marché international conformément à la norme EAN (European Article Number) /GTIN (Global Trade Item Number).*

4. *Nomenclature COICOP (Classification des fonctions de consommation des ménages). Voir la page du site Web des Nations Unies dédiée à ce sujet (UN, 2017).*

Données de transactions au sein du panier

Les données de transactions sont utilisées pour mesurer les prix de certaines catégories de produits et sont également une source d'information pour le calcul des poids. Dans le cadre des produits de consommation courante, elles comprennent le chiffre d'affaires hebdomadaire au niveau de l'article et du magasin, il s'agit donc d'informations sur la consommation réelle. Certains produits, comme par exemple les boissons alcoolisées, les médicaments vendus en pharmacie et les soins dentaires, ne sont couverts qu'à une fréquence mensuelle par des données exhaustives. En outre, Statistics Sweden dispose des données de caisse agrégées par année pour l'ensemble de la Suède depuis le milieu des années 1990 ; elles ont servi à construire le panier.

Comme on peut le voir au tableau 1, si les données de transactions sont utilisées, elles ne le sont pas dans toutes les parties du panier (les principales exceptions sont indiquées à l'encadré 1).

Mise en œuvre des données de caisse dans l'IPC suédois

Dans la mesure où Statistics Sweden traite les données de caisse depuis plus de cinq ans pour produire l'IPC mensuel, nous proposons de commencer par expliquer certains des choix faits par le passé.

Options d'utilisation des données de caisse

Statistics Sweden et le Conseil de l'IPC ont décidé en 2011 de conserver l'approche du panier fixe

Tableau 1
Poids dans le panier de l'IPC pour l'année 2016

Code	Titre	Poids dans le panier (%)	Données relatives aux transactions
01	Produits alimentaires et boissons non alcoolisées	139	Oui
02	Boissons alcoolisées, tabac et stupéfiants	39	Oui
03	Articles d'habillement et chaussures	53	Non
04	Logement, eau, gaz, électricité et autres combustibles	251	Oui
05	Meubles, articles de ménage et entretien courant du foyer	55	Non
06	Santé	38	Oui
07	Transports	135	Non
08	Communications	35	Non
09	Loisirs et culture	120	Non
10	Enseignement	5	Non
11	Restaurants et hôtels	67	Non
12	Biens et services divers	63	Oui
Total	IPC	1 000	

Note : conformément aux divisions de la nomenclature COICOP (deux chiffres) pour la consommation des ménages. Les données de transactions sont indiquées dès lors qu'elles sont utilisées dans les relevés de prix. Deux autres divisions COICOP (codes 13 et 14) couvrent la consommation autre que celle des ménages et sont donc hors du champ d'application de l'IPC.

ENCADRÉ 1 – Exceptions dans les données de caisse des produits de consommation courante : détaillants ne fournissant pas de données et produits frais

Dans les deux premières divisions de la nomenclature COICOP (01 et 02), les données de transactions sont utilisées de façon quasi exclusive, à deux exceptions près. Premièrement, certains détaillants inclus dans la division 01 (Produits alimentaires et boissons non alcoolisées) ne fournissent pas de données de transactions ce qui nécessite donc une collecte manuelle des prix. Deuxièmement, la collecte manuelle des prix se poursuit dans le segment des fruits et légumes frais, de la viande fraîche et des fromages. Ces articles sont habituellement vendus au poids, ou parfois à la pièce (par exemple, unités d'avocats ou de citrons).

En 2017, les données de caisse ont fait leur apparition pour les produits frais, d'abord pour un seul détaillant (Tongur & Sandén, 2016). En 2018, la double collecte des données (manuelle et numérique) a pris fin et les données de caisse ont été entièrement adoptées pour les détaillants fournissant des données de caisse (Bilius *et al.*, 2017). Des sujets connexes sont traités dans les publications de l'office des statistiques Norvégien (Statistisk sentralbyrå), voir Nygaard (2010) ou Rodriguez & Haraldsen (2005), ainsi que dans celles de CBS (Pays-Bas), voir van der Grient & de Haan (2010).

(voir encadré 2), estimant que c'était la façon la moins intrusive d'inclure les données de caisse⁵. Sa mise en œuvre a été immédiate, dès 2012, et a plus ou moins consisté à simplement changer le mode de collecte des données, ce qui était censé minimiser l'impact sur la production de l'IPC global et les systèmes informatiques associés. Cette décision a été prise sur la base de plusieurs études et analyses de données, ainsi que sur des comparaisons avec la collecte de prix manuelle (Norberg *et al.*, 2011). À côté de

la question de savoir comment utiliser les données de caisse dans la pratique, il fallait également décider si les données devaient effectivement être utilisées. Norberg *et al.* (2011) ont identifié quatre façons fondamentalement différentes d'utiliser les données de caisse, en se limitant au champ des produits de consommation courante. Elles sont présentées dans l'encadré 3.

5. Cette décision a été prise avec l'approbation du Conseil de l'IPC, qui avait un mandat réglementaire à ce moment.

ENCADRÉ 2 – Le Conseil de l'IPC suédois

Le Conseil de l'IPC suédois (*Nämnden för Konsumentprisindex* en suédois) est un organe consultatif scientifique et interdisciplinaire externe dédié aux méthodologies de production de l'IPC. L'IPC suédois est une statistique produite chaque mois qui n'est pas révisable. Le Conseil se réunit habituellement deux fois par an dans les locaux de l'institut (Statistiska centralbyrån – Statistics Sweden).

Le Conseil a été créé il y a plusieurs dizaines d'années et joue depuis 2017 un rôle de consultation, sans force d'obligation, pour les questions ayant une importance substantielle dans le cadre de l'IPC. Ses membres sont nommés par Statistics Sweden et représentent les établissements publics concernés par l'IPC, comme la Riksbanken (banque centrale de Suède) et d'autres agences gouvernementales ou organismes publics. L'IPC norvégien y est aussi représenté, afin de partager les expériences et de renforcer la collaboration entre les pays nordiques. Ces collaborations ont été utiles lors de l'introduction des données de caisse, puisque l'institut norvégien (Statistisk

sentralbyrå) fait partie des premiers instituts nationaux à les avoir utilisées. Des experts externes de portée internationale sont également membres du Conseil.

Avant 2017, le mandat du Conseil avait force d'obligation. Par ses décisions, le Conseil pouvait influencer sur les réponses apportées à toute question relative à l'IPC. En outre, ses décisions ne pouvaient pas être officiellement remises en question, conformément aux instructions légales applicables à Statistics Sweden. Parallèlement, le Conseil comptait un membre permanent du ministère de tutelle. Toutefois, en 2012, le système de la statistique officielle en Suède, ainsi que le rôle de Statistics Sweden en tant que grande agence gouvernementale ont été réexaminés (SOU, 2012). Cet examen avait été demandé par le gouvernement et, s'agissant de l'IPC, a été recommandé de supprimer la force obligatoire du mandat du Conseil de l'IPC, qui semblait discutable en termes d'objectivité de l'agence et contraire au Code des bonnes pratiques de la statistique européenne.

ENCADRÉ 3 – Quatre façons d'utiliser les données de caisse dans la production de l'IPC

A – Remplacement des données de prix collectées manuellement par des données de caisse pour les échantillons ordinaires de magasins et de produits : cela n'engendrerait que des changements/ajustements mineurs dans le mode actuel de production de l'IPC, et la réglementation relative aux indices des prix à la consommation harmonisés^(a) serait entièrement respectée.

B – Utilisation des données de caisse en tant qu'information accessoire : cela exigerait de choisir entre deux approches possibles et de continuer à échantillonner manuellement les prix cotés. L'échantillon serait calibré selon (i) les données de caisse des périodes correspondantes, ou (ii) les données de caisse seraient calibrées selon la collecte manuelle correspondante.

C – Calcul de l'indice à partir d'un recensement de tous les produits pour lesquels des données de caisse sont

disponibles : soit l'approche du panier fixe est appliquée à grande échelle, avec l'attrition correspondante du panier durant l'année, soit une méthodologie totalement différente est introduite, probablement en adaptant les méthodes néerlandaise ou norvégienne^(b) avec un chaînage mensuel.

D – Utilisation des données de caisse pour l'audit et le contrôle qualité : il s'agit de l'usage le plus minimaliste des données de caisse dans la production de l'IPC. Toutefois, si elles n'étaient utilisées que pour cet usage, cela représenterait également un énorme gaspillage de ressources.

(a) Voir la réglementation sur les indices des prix à la consommation harmonisés (Eurostat, 2013).

(b) Comme le soulignent van der Grient & de Haan (2010), Nygaard (2010) et les premières discussions avec le Statistisk sentralbyrå (institut de la statistique de Norvège).

Les quatre options présentées à l'encadré 3 donnent de possibles utilisations de ces données, pour des usages allant au-delà du simple contrôle qualité des prix collectés manuellement qui correspond à l'option D. L'option B a été envisagée mais elle est apparue peu optimale par rapport aux autres. Les données de caisse ayant été obtenues et mises en œuvre de façon progressive, la première option (option A) maintenait le statu quo pour la production de l'IPC s'agissant du calcul de l'indice et de la conception des échantillons. L'opportunité présentée par les données massives et l'émergence de nouvelles méthodes dans ce domaine, dont CBS (Bureau central de la statistique aux Pays-Bas) et Statistics Norway étaient les pionniers rendaient l'option C attractive. Toutefois, face à des contraintes temporelles et économiques et compte tenu du besoin d'acquiescer de l'expérience avec cette nouvelle source de données, l'option A semblait préférable pour commencer la transition vers ces nouvelles données. Utiliser l'ensemble des données (option C) n'était pas la solution préférable pour la première étape, mais elle demeure le but.

Panier fixe / panier dynamique

Les données de caisse, en 2012, ont été pour la première fois mises en œuvre dans le cadre d'un panier fixe standard. D'autres pays appliquent une approche moins conventionnelle, à savoir le panier dynamique. Les principales caractéristiques des deux approches sont présentées dans les consignes d'Eurostat pour le traitement des données de caisse de supermarchés (Eurostat, 2017a). Ces consignes ont été formulées à partir des informations fournies par les pays participants, afin de formaliser les approches appliquées par les différents pays et de s'efforcer d'harmoniser les indices des prix à la consommation dans les pays commençant à utiliser des données de caisse. Les principales différences, et les principaux avantages et inconvénients des deux approches sont présentés ci-dessous.

L'approche du panier fixe

L'approche du panier fixe signifie que, chaque mois t (ou trimestre) d'une année donnée y , le panier est maintenu dans une configuration aussi constante que possible. Les prix des articles inclus dans le panier sont observés (si cela est possible) puis comparés de façon directe chaque mois avec les prix concernés du mois de base des mesures annuelles, qui correspond habituellement au mois de décembre $y-1$.

L'évolution constante du panier et le problème des remplacements

L'inconvénient de cette approche prudente est de ne pas tirer parti de l'abondance des données, ni de la mise à jour des informations de marché. Elle repose sur un panier qui pour être maintenu dans la même configuration, est limité – cette contrainte venant de sa révision mensuelle, c'est-à-dire du remplacement de produits disparus. Le problème des remplacements est central pour la préservation de la comparabilité dans le temps, et constitue sans doute l'argument le plus convaincant pour maintenir l'approche classique : les changements de qualité et de quantité intervenant lors des remplacements sont traités de façon explicite. Lorsque les données relatives à certains articles ne sont pas observables, il faut faire un choix entre le remplacement par un autre article comparable (ce qui, dans le meilleur des cas, passe par une relance du même article) ou, si cela n'est pas possible, l'arrêt de l'article. Dans les cas extrêmes, l'attrition peut engendrer un panier non représentatif⁶ basé sur les articles restants. Le problème peut être contourné (mais pas résolu) grâce à une solution plus automatisée pour les données de caisse : le panier dynamique.

L'approche du panier dynamique

Dans le cadre de l'utilisation dynamique des données de caisse, les prix sont mesurés pour un panier mis à jour en continu. Pour cela, un indice mensuel des articles appariés est calculé pour les ratios de prix des articles exactement appariés entre des mois adjacents (t et $t-1$ de l'année y), et ce chaînage mensuel est ensuite relié au mois de base de l'indice (décembre de l'année $y-1$). Cette approche revient à un panier fixe si tous les articles (et les poids) sont identiques durant toutes les périodes – voir par exemple Eurostat (2017b, formules 8.11 et 8.14), Eurostat (2017a) et Fisher (1922).

L'approche dynamique retient dans le panier l'univers d'articles le plus récent, c'est-à-dire qu'il s'agit d'un échantillon à jour, et cette couverture est la plus représentative et complète qui soit. Comme Boskin *et al.* (1997), entre autres, le soulignent, une telle approche devrait être utilisée pour réduire les coûts de collecte et pour élargir l'assortiment des biens et services inclus dans l'IPC.

6. Dans ce cas, le panier ne couvre pas toute la consommation cible et n'est donc pas représentatif.

À des fins de régularité, c'est-à-dire en termes de 1) stabilité, 2) représentativité au fil du temps et 3) parcimonie des données visant à éviter le bruit, il est nécessaire d'exclure du panier les produits dont la part dans la consommation au cours du mois concerné est trop faible, comme expliqué par Eurostat (2017a) et par van der Grient & de Haan (2010), ou des produits en fin de cycle qui peuvent être bradés. Même avec ces précautions, un biais de chaînage peut découler des variations de prix, hausses ou baisses significatives durant certaines périodes, tirant ainsi l'indice vers le haut/bas durant la période concernée. Quand les prix des produits reviennent à leur niveau initial sans que l'indice ne retrouve son niveau précédent, il y a une dérive de l'indice.

Le problème peut être illustré comme suit. Supposons, par exemple, qu'un filtre de taille soit appliqué de sorte que les dix premiers articles en termes de chiffre d'affaires soient choisis pour un mois donné (lesquels étaient déjà retenus dans le panier la période précédente). Si le chiffre d'affaires d'un article est élevé un mois donné, cela peut être simplement parce qu'il est régulièrement très consommé, mais cela peut aussi être « temporaire » en raison, par exemple, d'une campagne publicitaire ou de ventes saisonnières (par exemple à Noël). Le mois suivant, ces articles « temporaires » sont susceptibles de ne plus être vendus au même prix, ou d'être arrêtés. En conséquence, les mêmes articles ne seront pas inclus dans les dix premiers ou le seront à des niveaux de prix différents, et l'indice chaîné ne retrouvera pas son niveau précédent, c'est-à-dire qu'il subira une dérive.

Cette dérive est encore plus marquée quand les quantités vendues, connues grâce aux données de caisse, sont utilisées dans la formule d'indice pour agréger les prix. Le biais de chaînage, une question à part entière, a été examiné de façon approfondie (voir Johansen & Nygaard, 2011 ; Nygaard, 2010 ; van der Grient & de Haan, 2011).

L'approche dynamique et les remplacements/relances : un problème qui n'en est pas un

Le plus gros inconvénient de l'approche dynamique est qu'elle ne retient pour le calcul de l'indice d'un mois donné que les produits présents deux mois successifs. Or une relance commerciale s'accompagne en général d'une augmentation de prix (soit le prix est inchangé pour une quantité vendue plus faible, soit le prix augmente sans amélioration tangible de la qualité). Cette augmentation est ignorée, ou « masquée », si la relance

n'est pas traitée explicitement. En effet, avec l'approche dynamique, aucun ajustement qualité n'est pratiqué car tous les articles du panier dynamique sont par définition présents deux mois adjacents. Or, comme le note Eurostat, « *les relances et les remplacements sont potentiellement problématiques dans cette méthode, car le système ne relie pas automatiquement un code d'article sortant avec le code de la relance ou du remplacement.* » (Eurostat, 2017a, p. 28).

Les données hebdomadaires dans un indice mensuel : comment les agréger ?

Disposer de données à plus haute fréquence pose la question des données multiples : faut-il combiner des données collectées à des dates différentes et si oui, comment ? La collecte manuelle des prix était effectuée une fois par mois et par magasin, ce qui est toujours le cas. Les prix collectés sont donc des prix ponctuels uniques. Comme stipulé dans les consignes sur les indices des prix à la consommation harmonisés (Eurostat, 2013), la procédure opérationnelle standard consiste à relever les prix durant la semaine correspondant au 15 du mois, ou également une semaine avant/après le milieu de semaine. Les relevés de prix (c'est-à-dire ceux des magasins inclus dans l'échantillon) sont habituellement attribués *a priori* durant trois semaines, pour une plus grande précision sur le mois.

Les données de caisse offrent la possibilité de calculer la consommation hebdomadaire, c'est-à-dire le chiffre d'affaires et les quantités achetées chaque semaine. Les données suivent les semaines civiles (du lundi au dimanche), ce qui empêche une utilisation régulière sur plus de trois semaines en raison des semaines chevauchant deux mois. En utilisant le milieu de semaine et les deux semaines adjacentes, on obtient trois points de données par offre de produit, dans le meilleur des cas. L'échantillon est donc plus précis, mais dans une dimension peu utilisée dans la méthodologie standard en raison de la nature des statistiques économiques : mesure discrète de données temporelles continues (voir le Manuel de l'IPC § 15.70, ILO, 2004).

Deux possibilités intuitives pour combiner les points de données hebdomadaires en un seul prix unique par offre de produit et par mois sont la moyenne géométrique et la moyenne arithmétique, qui semblent toutes les deux appropriées. Lors de la toute première mise en œuvre, le Conseil de l'IPC a décidé qu'une moyenne géométrique non pondérée

sur la période (maximale) de trois semaines serait appropriée pour calculer le prix mensuel de chaque offre de produit à partir des données de caisse. Ainsi, les données de caisse fournies par un même détaillant correspondraient aux données autres que les données de caisse des autres offres de produit. L'idée était que les trois semaines de données de caisse pouvaient être considérées comme trois cycles de collecte de données plutôt que comme une collecte unique, comme les autres offres de produit. L'agrégation par moyenne géométrique non pondérée était également conforme à la construction de l'indice, qui passe par une moyenne géométrique (indice de Jevons).

La question de l'agrégation semaine/mois a été réexaminée après obtention de données d'un plus grand nombre d'enseignes, à partir de 2013, et le Conseil de l'IPC a été consulté (Sammar & Norberg, 2012). Cette fois-ci, en raison de l'élargissement de la couverture, le Conseil a décidé d'utiliser une moyenne arithmétique pondérée sur trois semaines, qui refléterait mieux les prix unitaires mensuels, en ligne avec les données réelles (hebdomadaires). « Pondéré » signifie que les chiffres d'affaires de trois semaines (au plus) sont additionnés puis divisés par la somme des quantités de ces semaines, donnant ainsi un prix unitaire mensuel moyen.

L'évolution des deux moyennes envisagées a été étudiée dans le contexte d'un indice de prix (cf. Norberg *et al.*, 2012), ce qui a permis d'identifier des comportements différents dans certaines situations. Pour plus de 90 % des observations, les deux moyennes ne différaient que très peu, mais les différences s'accroissaient lorsque la pondération jouait un rôle important, par exemple en période de vacances lorsque les prix étaient bas. L'étude a également permis de réaliser que tout impact sur la période de base affectait par la suite les rapports de prix agrégés (c'est-à-dire l'indice) tout au long de l'année, même si les deux moyennes coïncidaient durant le mois concerné.

Maintenance des échantillons

Suite au passage aux données de caisse, les remplacements/substitutions d'articles correspondant aux articles du panier devenus obsolètes devaient être faits en interne par les responsables de l'IPC, à partir de la surveillance de l'attrition au sein du panier. Afin de limiter l'épuisement potentiel des échantillons, un système de maintenance du panier extrêmement simple a été mis en œuvre, permettant de comparer les ventes du mois en cours t avec la période de base décembre $y-1$. Ce suivi se fait tant

au niveau du nombre de magasins dans lesquels le produit a été vendu qu'au niveau du nombre de lots vendus, c'est-à-dire qu'il s'agit d'une analyse bidimensionnelle. Cela se déroule *a posteriori* pour chaque mois achevé. Ainsi, l'échantillon utilisé pour l'IPC reste (vraisemblablement) représentatif, ce qui ne demande qu'une journée de travail chaque mois, au plus, pour rechercher les articles substitués dans les données de caisse. Les prix manquants ne font l'objet d'aucun calcul et les magasins ne sont pas remplacés s'ils ferment avant la prochaine date de mise à jour de l'échantillon. Toutefois, les cas de non-réponse sont rares, surtout s'il s'agit de magasins bien établis ou dont le chiffre d'affaires est élevé.

Estimation de la variance de l'indice

On s'intéresse ici à la contribution d'un article à la variance de l'indice des prix dans le cas d'un panier fixe, en utilisant les données de caisse dans leur intégralité ou partiellement. Après une brève description de l'échantillonnage, la construction de l'indice dans le cadre des agrégats élémentaires est présentée, suivie par une estimation *jackknife* de la variance. Cette section se termine par une discussion sur les propriétés de population finie de l'échantillon des produits de consommation courante.

Échantillonnage des articles et des magasins pour les produits de consommation courante

La conception de l'échantillon porte sur deux dimensions : l'endroit et le produit (articles pouvant être achetés). Par « endroit » il faut entendre le magasin dans lequel les produits destinés à une consommation privée sont achetés. Les articles sont sélectionnés par le biais d'un échantillonnage annuel, quel que soit le mode de collecte des données. Pour les données de caisse comme pour la collecte manuelle des prix, une probabilité d'échantillonnage proportionnelle à la taille est appliquée dans les deux dimensions (voir Ohlsson, 1990 ; Rosén, 2000).

Échantillonnage des articles

À partir de chacune des enseignes couvertes par des données de caisse, environ 800 articles sont inclus dans l'échantillon annuel. Les bases de sondage sont définies chaque année en fonction des données de caisse annuelles agrégées de l'année précédant le mois de base. Les codes identifiant les articles dans les données de caisse, les codes EAN/GTIN et les niveaux plus détaillés

de la nomenclature COICOP sont ensuite reliés de façon précise. En faisant les appariements nécessaires avec les données de caisse hebdomadaires, on obtient l'échantillon souhaité. Les échantillons d'articles des enseignes sont tirés avec une coordination négative entre les échantillons des différentes chaînes. Toutefois, de nombreux articles de marques répandues sont vendus par tous les détaillants et représentent de gros volumes de vente. Ces articles sont souvent inclus dans plusieurs des échantillons spécifiques aux détaillants.

Échantillonnage des magasins

L'échantillon des magasins de produits de consommation courante rassemble environ 60 magasins, répartis dans l'ensemble du pays. La probabilité d'inclusion dans l'échantillon est proportionnelle à la taille (*Poisson sampling*, voir Ohlsson, 1990). Ainsi, il est possible d'obtenir des rotations des points de vente tirés. Toutefois, le système de rotation standard de Statistics Sweden (20 % chaque année) n'est pas strictement appliqué ici. La rotation est appliquée si elle se justifie d'un point de vue probabiliste (c'est-à-dire en termes de représentativité), afin d'éviter une trop forte pression sur les fournisseurs de données pour qu'ils changent le contenu des données communiquées. Pour des raisons techniques, les magasins sont rééchantillonnés chaque année, mais ne sont remplacés que si leur importance relative est extrêmement affectée par rapport à l'échantillonnage des années précédentes.

Configuration de l'estimation

Il est très difficile d'estimer la variance dans un indice des prix à la consommation. La variance découle de l'échantillonnage bidimensionnel, magasins et articles ; des évaluations formelles de la variance sont fournies par Balk (1989, 1991), Dalén & Ohlsson (1995) et Norberg (2004).

L'indice au niveau le plus détaillé : agrégats élémentaires

Les agrégats élémentaires sont calculés comme la moyenne géométrique⁷ des rapports de prix d'articles appartenant à un groupe de produits donné, dans tous les magasins. Les rapports de prix de la période d'observation t de l'année en

cours y par rapport aux prix du mois de base 0, $P_{t,i}$ et $P_{0,i}$ constituent la formule de l'indice $I_g^{0,t}$:

$$I_g^{0,t} = \prod_{i=1}^{k_g} \left(\frac{P_{t,i}}{P_{0,i}} \right)^{w_i} \quad (1)$$

où la somme est calculée pour les k_g offres de produit i d'un groupe de produits g dans lequel chaque offre de produit peut avoir un poids distinct w_i . Dans le cas suédois, les poids w_i sont calculés en fonction des probabilités relatives aux magasins et aux articles. Il s'agit pour la plupart de poids unitaires, c'est-à-dire égaux ($w_i = 1$), mais certains peuvent être plus importants pour refléter un article largement vendu dans un hypermarché, par exemple une marque de café.

Si tous les poids sont égaux (ce qui revient à ne pas pondérer), alors l'équation (1) correspond à l'indice de Jevons non pondéré. Si les éléments inclus dans l'échantillon résultent d'un échantillonnage proportionnel à la taille, les probabilités d'inclusion et les poids se compensent mutuellement, c'est-à-dire que l'on obtient une pondération implicite. Lorsque les poids reflètent la part de la consommation respective des articles, l'expression correspond à l'indice de Young géométrique (voir le Manuel de l'IPC, formule 1.9, ILO, 2004).

La méthode du jackknife pour l'échantillonnage stratifié

La méthode du *jackknife* suggérée ici permet une approximation de la contribution à la variance du $n^{\text{ème}}$ élément d'un échantillon existant. La méthode est expliquée dans Wolter (1985), et une analyse similaire sur données de caisse est disponible dans l'étude de Leaver & Larson (2001) pour l'IPC américain (Bureau of Labor Statistics).

La stratégie de calcul consiste à estimer le paramètre cible, ici une expression de l'indice agrégé pour les indices du groupe de produits (équation 1) en excluant, un par un, chaque élément de l'échantillon existant, c'est-à-dire en conservant $n-1$ éléments à chaque estimation et en calculant le paramètre cible en fonction des éléments restants. En appliquant cette procédure à tous les éléments n , on obtient la contribution moyenne à la variance. L'échantillon de magasins choisi est fixe, c'est-à-dire que l'échantillon d'articles dépend de l'échantillon existant de magasins. Cette approche est supposée suffire pour la validation de principe, à savoir le compromis entre la

7. Cette formule d'indice est l'une des deux méthodes explicitement recommandées pour les indices des prix à la consommation harmonisés (Eurostat, 2013) au plus bas niveau.

contribution à la variance d'un article et le biais qui survient lorsque les ajustements de la quantité sont négligés.

Mécanisme d'estimation jackknife

Les quelque 800 articles échantillonnés pour lesquels des données de caisse sont disponibles dans chacune des trois enseignes constituent ensemble environ 90 groupes de produits du secteur des produits de consommation courante dans la nomenclature COICOP. Les groupes de produits sont, par définition, les agrégats élémentaires pour lesquels un indice de prix est calculé avec l'équation (1) pour tous les produits et les enseignes, c'est-à-dire qu'il y a un agrégat pour tous les articles d'un groupe de produits. Les articles sont classés et codés selon le groupe de produits auquel ils appartiennent, et donc un « article » correspond à un produit.

Le dispositif de stratification est présenté au tableau 2, et montre le mécanisme d'exclusion pour chaque cycle $n-1$. Dans ce dispositif, les groupes de produits sont croisés avec chaque enseigne pour définir les strates, ce qui donne environ 270 strates desquelles les articles sont exclus. L'équation (1) estimée pour tous les groupes de produits débouche sur le paramètre cible : l'indice agrégé des produits de consommation courante pour la division 01 de la nomenclature COICOP.

De par leur conception, les 90 groupes de produits croisés avec les trois enseignes au plus donnent environ $L = 270$ strates. Au total, les près de 800 produits échantillonnés au sein de chaque enseigne donnent 2 400 produits, avec des

variations dues à l'évolution des assortiments. Une strate de magasins h compte n_h articles/produits. Le n_h varie selon les strates au sein du même groupe de produits, qui compte donc k_g produits au total : $k_g = \sum_{h=1}^H n_h, h \in g$. Au sein de chaque k_g il peut y avoir $H = 3$ strates, où h se résume à $L = 270$ pour tous les $g : h \in (g, L)$.

Les quelques strates où un seul produit est identifié sont éliminées du calcul car la procédure $n-1$ donne un nombre nul de produits restants, ce qui veut dire qu'aucune variance ne peut être estimée dans la strate concernée. Les assortiments et les échantillons variant d'une enseigne à l'autre, parfois de manière substantielle, tous les groupes de produits ne sont pas nécessairement inclus dans les trois enseignes.

Chaque estimation exclut séquentiellement l'une des rangées indiquées au tableau 2, c'est-à-dire chaque produit d'une strate, de sorte qu'aucun élément aléatoire n'entre dans la procédure d'estimation. Par contre, le caractère aléatoire de l'échantillon initial est reflété entre les cycles par la modification de la composition de l'échantillon concerné.

Le paramètre d'intérêt

L'équation (1) peut être exprimée sous forme logarithmique, produisant la somme suivante pour chaque groupe de produits, suivie d'une exponentiation :

$$I_g^{0,t} = \prod_{i=1}^{k_g} \left(\frac{P_{t,i}}{P_{0,i}} \right)^{w_i} = \exp \left[\sum_{i=1}^{k_g} w_i (\ln(P_{t,i}) - \ln(P_{0,i})) \right] \quad (2)$$

Tableau 2
Configuration du mécanisme d'estimation jackknife

Cycle d'estimation	Groupe de produits	Code du produit	Strate h	Chaîne
1	1113	1113001	1	1
2		1113002	1	1
3		1113003	2	2
4		1113004	3	3
5		1113005	3	3
6		1113006	3	3
7	1114	1114001	4	1
.
.
$n = 2\ 400$.	.	$L = 270$.

Note : les nombres $n = 2\ 400$ et $L = 270$ sont approximatifs et sont fournis uniquement à titre d'illustration. Les nombres exacts sont indiqués dans la sous-section consacrée aux estimations. Les champs grisés illustrent la stratification en termes de chaîne.

La partie entre crochets à droite de l'équation (2) est une version linéarisée de l'équation (1), semblable à la formule utilisée par Leaver & Larson (2001). Cela constituera le paramètre d'intérêt lors de l'élimination des produits/articles, $n-1$, au sein de chaque strate h du groupe de produits g .

Pour les estimations de cette étude, l'indice de l'agrégat élémentaire (2) est calculé de façon légèrement différente de la pondération réelle⁸. La différence tient au fait que les observations et les rapports de prix au sein de chaque enseigne (=strate) sont ramenés à la moyenne et agrégés pour le groupe de produits en pondérant selon la part de marché moyenne de chaque détaillant pour obtenir (2) pour le groupe de produits complet. Cela remplace les poids des articles individuels w_i ; il est nécessaire de le faire car l'alternance du nombre de produits compense la pondération implicite découlant de l'échantillonnage proportionnel à la taille. Les poids sont normalisés, de sorte que, selon le nombre d'enseignes pour chaque groupe de produits, un poids connu *a priori* est attribué au rapport de prix moyen des détaillants⁹. Cela modifie l'équation (2) de la façon suivante :

$$I_g^{0,t} = \prod_{h=1}^H \left[\prod_{i=1}^{n_{h,g}} \left(\frac{P_{t,i}}{P_{0,i}} \right) \right]^{w_h} \quad (2')$$

$$= \exp \left[\sum_{h=1}^H w_h \sum_{i=1}^{n_{h,g}} \left(\ln(P_{t,i}) - \ln(P_{0,i}) \right) \right]$$

La dernière estimation de l'indice de prix des produits de consommation courante est une moyenne arithmétique pondérée calculée sur l'ensemble des indices des groupes de produits, selon la formule

$$I^{0,t} = \sum_{g=1}^G w_g I_g^{0,t} \quad (3)$$

où les poids du groupe de produits w_g sont normalisés de façon à sommer à 1 (leur part agrégée dans le panier total, cf. tableau 1).

Par analogie aux définitions de Wolter (1985) pour l'estimation par stratification, l'indice de prix de l'équation (3) est calculé lorsque la $(h, i)^{\text{ème}}$ observation est supprimée. Cela est fait pour toutes les suppressions au sein d'une strate et dans toutes les strates, soit environ $n = 2\,400$ fois, ce qui produit autant d'estimations qu'il y a d'articles/de produits. L'estimation de la variance est obtenue avec, au plus, environ $L = 270$ moyennes (strates) tirées des estimations (voir équation (5) ci-dessous). Ces L moyennes sont calculées, pour chaque strate h , comme l'estimation moyenne du paramètre pour les n_h paramètres estimés,

$$\hat{\theta}_{(h\cdot)} = \sum_{i=1}^{n_h} \hat{\theta}_{(hi)} / n_h \quad (4)$$

de sorte que chaque suppression ($n-1$) produise le paramètre $\hat{\theta}_{(hi)}$ de l'équation (4), c'est-à-dire une estimation de l'indice total du prix des produits de consommation courante de l'équation (3), $\hat{\theta} = I^{0,t}$, avec le $i^{\text{ème}}$ article supprimé.

L'estimateur de variance *jackknife* de l'indice finalement calculé pour tous les groupes de produits dans le secteur des produits de consommation courante est :

$$v(\hat{\theta}) = \sum_{h=1}^L \frac{w_h}{n_h} \sum_{i=1}^{n_h} \left(\hat{\theta}_{(hi)} - \hat{\theta}_{(h\cdot)} \right)^2 \quad (5)$$

Notons que w_h dans l'équation (5) est un facteur de correction en fonction des strates, $w_h = (n_h - 1) \left(1 - \frac{n_h}{N_h} \right)$, sans échantillonnage de remplacement.

Résultats de l'estimation

Sur la base de $n = 2\,066$ cycles des $L = 231$ strates complètes ($n > 1$), l'écart-type estimé de la variation de l'indice des prix des produits de consommation courante calculé avec des données de caisse est de 0.168 unités indicielles en moyenne sur les douze mois de l'année 2016, c'est-à-dire la variation mensuelle par rapport à la période de base. Ainsi, pour une valeur de l'indice de 102, l'incertitude d'un intervalle de confiance à 95 % devient [101.67 ; 102.33]. Les estimations de l'écart-type mensuel sont indiquées au tableau 3.

Les résultats du tableau 3 méritent d'être remis dans le contexte de la réalité pratique. Si les échantillons découlaient d'un échantillonnage aléatoire simple et si, au même moment, la consommation de biens était également distribuée entre tous les produits au sein de chaque groupe de produits (c'est-à-dire si les préférences des consommateurs étaient identiquement hétérogènes et la consommation répartie à parts égales sur tous les articles), les résultats obtenus pourraient facilement être étendus à l'ensemble de tous les produits. Dans

8. C'est actuellement le cas pour Statistics Sweden. D'autres options sont possibles : CBS (Pays-Bas) applique le calcul de l'indice, les agrégats élémentaires, aux enseignes individuelles, ce qui représente un niveau un peu plus détaillé qu'ici (van der Grient & de Haan, 2010).

9. En réalité, certains produits ont des poids individuels reflétant de gros volumes de consommation. Cela est ignoré ici afin d'éviter toute volatilité dans les estimations de variance qui ne serait due qu'à la pondération. Tous les produits de l'échantillon sont considérés comme le résultat d'un échantillonnage aléatoire simple.

Tableau 3
Estimations de l'écart-type

Mois en 2016	écart-type
Janvier	0.1725
Février	0.1464
Mars	0.1514
Avril	0.1668
Mai	0.1692
Juin	0.1705
Juillet	0.1825
Août	0.2047
Septembre	0.1651
Octobre	0.1684
Novembre	0.1805
Décembre	0.1426

Note : valeur en unités indicielles. Indice des produits de première nécessité basé sur des données de caisse. 2 066 produits et 231 strates.

un tel cas hypothétique, et sachant qu'un magasin de produits de consommation courante vend en moyenne plus de 10 000 articles, l'échantillon utilisé pour l'IPC suédois (800 articles) permettrait, avec une couverture de 8 % de calculer la variance totale, *via* l'ajustement pour une population finie, $(1 - (n/N))$. Si la taille de l'échantillon est $n = 800$ et que la taille de la population est $N = 10\ 000$, la correction en population finie est de $(1 - (800/10\ 000))$, appliquée au tableau 3.

Les écarts-type estimés peuvent être évalués relativement à l'écart-type total de l'IPC. La part des produits de consommation courante dans l'IPC est de 13.9 %, comme indiqué au tableau 1, tandis que l'écart-type total de l'IPC, pour le taux d'inflation annuel, est estimé à 0.12 unités d'indice (SCB, 2017). Si l'écart-type estimé pour les produits de consommation courante est relié à cet écart-type total en fonction de la pondération, alors seulement 4 % de la variance de l'IPC résulte des produits de consommation courante (le poids est au carré, de même que les écarts-type, pour obtenir les niveaux appropriés). En raison de cette faible contribution à la variance, une augmentation de la taille de l'échantillon ne permet pas d'obtenir une précision beaucoup plus importante de l'IPC global, même si les éléments inclus dans le panier découlent d'un échantillonnage aléatoire simple.

La pondération des articles, explicite ou implicite par le biais d'un échantillonnage proportionnel à la taille, compense ce calcul linéaire puisqu'il s'agit d'un effet du plan de sondage. Pour cette raison, prendre un échantillon des quelques

articles les plus vendus et de quelques articles représentatifs pour le reste engendre en pratique une contribution à la variance inférieure à celle qui découlerait d'une simple estimation de la variance, comme nous le faisons ici. Une approche alternative consisterait à utiliser le panier dynamique avec une limite pour les articles les plus vendus. Certes, l'application d'un tel seuil en termes de part de valeur par groupe de produits améliorerait la précision, mais cela n'est pas nécessairement mieux pour estimer l'inflation – c'est plus simple, mais probablement seulement un peu plus précis puisque la consommation n'est pas répartie à parts égales entre tous les articles.

Interactions et propriétés de population finie

Il peut y avoir des liens, en termes de niveau de prix, entre les articles et les magasins et, en conséquence, entre les marques. Il peut être pertinent de tenir compte de cette interaction dans le cadre de l'estimation de la variance de l'IPC, comme l'explique Norberg (2004). Toutefois, dans la mesure où on considère ici que l'échantillon de magasins est fixe, toute interaction potentielle est ignorée dans ce qui suit, probablement sans affecter les résultats.

Une autre caractéristique de l'échantillon d'articles existant est sa propriété de population finie. Comme nous l'avons indiqué, les échantillons d'articles sont obtenus à partir de bases de sondage complètes présentant une couverture quasi complète pour l'année concernée, $y-2$. Dans la mesure où la probabilité d'échantillonnage est proportionnelle à la taille, certains

articles/produits de l'échantillon sont les plus vendus et ont donc une probabilité d'inclusion égale à 1. Par conséquent, la variance effective est inférieure à celle estimée ici, car l'estimation jackknife traite tous les items avec la même probabilité d'inclusion alors qu'elle varie dans la réalité. Le compromis proportionnel correspond au scénario le plus défavorable, comme si tous les articles étaient échantillonnés avec une probabilité égale.

Changements de quantités dans les produits de consommation courante

On examine maintenant certains changements de quantités observés et notables sur le marché suédois des produits de consommation courante, et leur impact sur l'IPC si les produits concernés font partie de l'échantillon utilisé pour l'IPC¹⁰. Jusqu'à présent, à notre connaissance, les cas de lots dont la taille augmente ou de paquets de plus en plus pleins sont assez rares, et les questions soulevées ici concernent des diminutions de la quantité, qui ont été reprises par les médias suédois. En un sens, la quantité peut être considérée comme un aspect de qualité – d'ailleurs les deux termes sont parfois utilisés de façon interchangeable (voir le Manuel de l'IPC §7.77, ILO, 2004). Lorsque c'est nécessaire, des ajustements de quantité sont effectués pour les articles (de remplacement) entrants afin que leur prix soit exprimé pour des unités comparables à celles des produits précédents (ceux de la période de base).

Substitution d'articles et ajustements en unités comparables

Le plan de sondage et l'introduction d'articles de remplacements au sein du panier de l'IPC ont un intérêt particulier pour assurer la comparabilité durant l'année, comme souligné dans le Manuel de l'IPC (ILO, 2004, Ch. 8) qui examine aussi les données de caisse. Ainsi, il est noté que : « *Si peu de choses changent en termes de qualité et de gamme des produits disponibles, la méthode des modèles appariés présente de nombreux avantages. La méthode des modèles appariés permet de faire des comparaisons à périmètre constant dans des magasins semblables. [...] En cas de roulement très rapide des articles engendrant un épuisement rapide de l'échantillon, les remplacements ne sont pas fiables pour maintenir l'échantillon. Des mécanismes alternatifs, qui échantillonnent le double univers des articles dans chaque période ou qui utilisent ce double univers, doivent être*

employés. Citons par exemple les indices chaînés et les indices hédoniques [...] » (ibid., § 8.62).

Il apparaît clairement que, en cas d'attrition au sein du panier, ou plus précisément de diminution de sa représentativité, une procédure de mise à jour plus rapide du chaînage et du rééchantillonnage mensuel sera probablement plus efficace et plus appropriée pour les données de caisse. Mais reste la question des changements de quantités des produits : le début de la citation ci-dessus (« *Si peu de choses changent...* ») suggère que la méthode des modèles appariés ne fonctionne pas. La principale différence entre la formule d'un indice chaîné mensuellement et celle du panier fixe repose sur le fait que les changements de quantité, s'ils ne sont pas traités, affectent le panier fixe en fonction du calendrier – le biais dépend du nombre de mois restants jusqu'à la mise à jour annuelle de l'échantillon. Une procédure de chaînage mensuel ne fait que renvoyer le problème à plus tard.

La question connexe est celle des valeurs unitaires. Elle est examinée par von Auer (2011) qui discute des indices basés sur la valeur unitaire lorsque les produits sont similaires mais pas identiques, et des valeurs unitaires dans la durée. L'un des critères de similitude est la taille du lot – la commensurabilité – pour laquelle une stratégie de « valeur unitaire modifiée » est présentée. La valeur unitaire modifiée consiste à transformer/recalculer, de façon linéaire, la taille des lots en unités communes entre les produits similaires, afin de préserver la comparabilité avec la période de base¹¹. Bien que cela ne puisse pas être utilisé directement dans notre analyse, cette approche semblerait très pertinente : les vraies valeurs unitaires sont en quelque sorte reportées à la période de base. Cette approche revient à un panier, et non simplement l'indice, basé sur la valeur unitaire. Ici encore, l'enjeu méthodologique est de pouvoir faire des comparaisons pertinentes et non pas de contourner le problème¹². En particulier, qu'il s'agisse du concept de changement des niveaux de prix ou de la méthodologie conventionnelle de l'IPC, la linéarité du calcul de la valeur unitaire proportionnelle peut être mise en question. Des études menées en interne par Statistics Sweden

10. Pour des raisons de confidentialité, les produits spécifiques inclus dans le panier de l'IPC ne peuvent pas être communiqués. Toutefois, ces exemples sont de source publique et concernent ici l'impact potentiel de situations hypothétiques sur l'IPC.

11. von Auer (2011) s'intéresse au changement des niveaux de prix, notion qui s'éloigne de la moyenne des changements de prix de l'IPC.

12. Le chaînage et l'orgueil démesuré des statisticiens des prix ont fait l'objet d'une étude approfondie par feu le Professeur Peter von der Lippe. cf. www.von-der-lippe.org (19/07/2017).

montrent que la corrélation taille/prix n'est pas proportionnelle mais plutôt exponentielle, en-dessous du niveau unitaire, c'est-à-dire qu'un doublement de la taille engendre moins d'un doublement du prix.

Changements des quantités : quelques exemples sur le marché suédois

Ces dernières années, plusieurs changements d'emballage de certains produits de consommation courante sont intervenus sur le marché suédois. Certains de ces changements ont directement affecté le calcul de l'IPC en raison des ajustements de quantité correspondants aux prix de la période de base pour le panier fixe. Sans ces ajustements, il pourrait se produire un biais notable dans l'IPC en termes d'inflation cachée. Les exemples suivants en fournissent une illustration.

Le café : de nombreux paquets de café ont diminué en taille, passant d'une contenance auparavant « standard » de 500 grammes à maintenant 450 grammes, soit une baisse de 10 %. D'ailleurs, la plupart des paquets actuellement vendus contiennent moins de 500 grammes. Le prix du café peut être assez fluctuant et les ventes par lot sont courantes, par exemple « trois pour le prix de deux », faisant de cet article un cas intéressant. La variation de 10 % de la taille du paquet a été prise en compte manuellement, suivant la procédure standard de l'IPC. Toutefois, la question du changement réel de prix est discutable. De fait, la hausse de prix implicite suite au changement de la taille des paquets a fait la une des médias lors d'un conflit entre un grand distributeur suédois de produits de consommation courante et un producteur de café détenant une part importante du marché¹³. Avec un chaînage mensuel, ce changement ne serait pas détecté. Le poids du groupe de produit « Café » s'élève à 0.39 % dans le panier, de sorte que, si aucun ajustement n'était fait, une inflation de 0.039 (découlant du changement de 0.1 de la taille unitaire) serait ignorée. À noter toutefois qu'elle serait sans doute amalgamée avec les fluctuations du prix du café.

Le lait caillé : en 2015, au moins un producteur laitier a changé le contenu de la brique de lait caillé (le « *filmjolk* », produit spécifiquement suédois très populaire, proche du yaourt et décliné dans plusieurs variétés et teneurs en matière grasse), en passant des litres aux grammes. Ce changement est passé quasiment inaperçu jusqu'à

ce que la presse quotidienne et la radio publique¹⁴ l'annoncent dans un flash info. Dans la mesure où le litre est une mesure de volume et le gramme une mesure de poids, et sachant que la densité d'un produit laitier dépend de sa teneur exacte en matières grasses¹⁵ (FAO, 2012), il était donc difficile d'évaluer la variation de quantité. Des ajustements ont été effectués de façon pragmatique pour toutes les marques et les variétés de l'échantillon utilisé pour l'IPC.

Le groupe de produits correspondant, qui couvre à la fois les yaourts et le lait caillé, représente 0.419 % du panier. Suite à la réduction de quantité de 3 % (une mesure approximative du changement de volume), une brique de 1 000 millilitres ne contient plus que 970 millilitres. Puisque les prix n'ont pas changé au point de vente, cela engendre un biais de 0.03 unité pour plusieurs produits inclus dans l'IPC pour un poids agrégé de 0.419 % du panier. Si au moins un tiers du groupe de produits est constitué par ces produits, le biais se chiffre alors à 0.013 %. Pris de façon isolée, ce chiffre apparaît très faible. Mais dans un contexte global, l'addition (ou la multiplication) de biais pour divers les articles peut devenir conséquente au fil du temps, et modifier l'évolution de l'indice.

Le tabac : en raison de la réglementation européenne, la taille des paquets des produits inclus dans le groupe du tabac (qui comprend les cigarettes et le « *snus* », tabac humide spécifique à la Suède) a changé. Le nombre de cigarettes dans un paquet a oscillé entre 19 et 20 cigarettes. Ces changements de quantité doivent être comptabilisés dans le panier fixe lors des remplacements, sans quoi les prix en vigueur sur le marché ne correspondent plus à la même quantité de produit, et ce changement de 0.05 unité engendre un biais sur les articles de tabac dû aux cigarettes. Le poids des produits de tabac est de 1.545 % au sein du panier, et celui des cigarettes de 1.01, soit un biais de 0.05 uniquement dû aux cigarettes.

Au total, si les biais présentés dans ces trois exemples étaient masqués dans le chaînage, un biais total d'environ 0.1 % pourrait survenir ($\approx 0.039 + 0.013 + 0.05$ % des poids). Cela peut être comparé à l'écart-type de 0.168 unité d'indice dans le cas de l'échantillonnage aléatoire simple, soit une surestimation de l'écart-type effectif.

13. Cela a suscité une petite controverse, un producteur de café suédois estimant que les prix imposés aux consommateurs reflètent la politique de prix des détaillants et non pas celle des producteurs (Berge, 2016).

14. Voir l'expérience réalisée par la radio publique en Suède (Sveriges Radio) dans Bressler & Näslund (2015).

15. Des données sont disponibles sur Internet pour calculer la densité du lait. Nous n'avons pas de chiffre exact pour ce produit suédois spécifique.

* *
*

Avec l'arrivée de nouvelles sources de données, de nouvelles possibilités voient le jour. La couverture, une caractéristique des données massives telles les données de transactions, est incontestable en termes de contexte et de portée. Ces données sont de l'ordre des recensements, moins d'un siècle après l'introduction de la théorie de l'échantillonnage aléatoire, qui visait à préserver la représentativité par le biais d'échantillons petits et rentables (sur la théorie de l'échantillonnage aléatoire, voir Neyman, 1934 ; plus généralement sur les enquêtes par sondage, voir la passionnante anthologie de Betlehem, 2009).

Les données de caisse ont quelque peu mis en question la méthodologie traditionnelle de production de l'IPC, en particulier avec le développement de nouvelles méthodes empruntées à l'analyse des « big data » (par exemple l'apprentissage automatique). De ce point de vue, Statistics Sweden a avancé avec prudence, dans un premier temps à petite échelle, pour préserver la comparabilité dans le temps et avec d'autres pays aux fins de l'harmonisation des indices des prix à la consommation, et pour assurer la transparence.

Dans cet article, nous nous sommes concentrés sur l'inclusion dans l'IPC des données de caisse pour les produits quotidiens, et particulièrement sur la question de l'arbitrage entre la variance liée à l'item et le biais lié au fait de ne pas tenir compte explicitement des ajustements de quantités. Une hypothèse implicite est l'absence de changement technologique, c'est-à-dire que les développements technologiques n'ont pas d'impact direct sur les prix des aliments et des boissons à court terme, de sorte que l'approche traditionnelle du panier fixe peut être maintenue toute l'année. De plus, la collecte manuelle des prix demeure le moyen le plus courant de produire l'IPC, y compris avec des comparaisons directes et des ajustements quantitatifs en cas de remplacement d'un article. Nous avons vu, pour les produits de consommation courante, que la contribution à la variance ou à l'écart-type d'un élément échantillonné au hasard est plutôt faible et aurait tendance à diminuer avec un échantillonnage approprié. Étant donné que les échantillons sont fondés sur des stratégies d'échantillonnage proportionnelles à la taille, la précision est en fait plus élevée que ne le suggèrent les résultats de cet article – bien qu'inférieure à celle obtenue dans les approches dynamiques couvrant des volumes de ventes plus

importants. Cela doit être reconnu comme un avantage des méthodes dynamiques, mais l'ampleur de l'amélioration de la précision n'est pas certaine, notamment en raison de la dépendance entre les produits et les détaillants. Comme le montre l'article, les approches mécaniques sans contrôle peuvent être mises en question, non pas en termes de couverture, mais parce que l'indice qu'elles génèrent peut masquer l'inflation plutôt que la montrer si les changements de quantité sont ignorés.

Bien que l'article ait traité du cas des produits de consommation quotidienne, ce problème concerne l'IPC global, ce qui met en lumière un inconvénient possible de l'utilisation de données de caisse : des détails importants comme les ajustements quantitatifs peuvent maintenant être brouillés dans le déluge de données – comme si la couverture seule était la panacée pour obtenir des mesures précises de l'inflation.

Cependant, cela ne doit pas conduire à ignorer ou à nier les opportunités offertes par les données de caisse. De nombreux développements sont en cours dans d'autres pays, comme en témoignent les meetings du Groupe d'Ottawa, le plus important forum mondial sur les indices des prix. Il convient ici de noter les avancées de Statistics Netherlands (CBS) dans ce domaine, dont témoignent divers rapports de recherche publiés par l'institut néerlandais. Néanmoins, d'un point de vue comparatif, l'utilisation de données de caisse avec des méthodes isolées, qui ne peuvent être comparées mais qui modifient significativement la méthodologie de l'IPC, peut être discutable. L'effort pourrait également être disproportionné pour obtenir une légère augmentation de la précision globale : nous avons vu ici que les variations de prix de produits de nécessité quotidienne (à l'exclusion des fruits et légumes) sont faibles, ce qui peut être comparé à d'autres sources d'erreur qui peuvent affecter l'IPC.

Pour finir, l'arrivée des « big data » devrait nous inviter à garder à l'esprit que la production de statistiques exige une évaluation de la qualité du processus dans son ensemble, et pas seulement des données, comme le soulignent par exemple Biemer *et al.* (2014) et Biemer & Lyberg (2003). Il s'agit de penser en termes d'« erreur totale d'enquête » (Biemer *et al.*, 2017). Pour les données de caisse, et en particulier l'échantillonnage dynamique, cela implique un contrôle de qualité au niveau de la codification au sein de la nomenclature COICOP. Sinon, les données pourraient ne pas être cohérentes avec le panier.

S'assurer que les données sont cohérentes avec la méthodologie de l'enquête est une question de précaution, comme le note par exemple Couper (2013), qui souligne la nécessité d'adapter les données au sujet plutôt que de déformer le sujet pour l'adapter aux données.

Alors que d'autres pays ont commencé à mettre en œuvre des approches développées avec les « big data », Statistics Sweden s'est tenu pour l'instant à la méthodologie traditionnelle de l'IPC. Mais d'autres étapes dans l'utilisation des données de caisse sont probables dans un proche avenir. □

BIBLIOGRAPHIE

von Auer, L. (2011). The Generalized Unit Value Index. Universität Trier, *Research Papers in Economics* N° 12/11.

Balk, B. (1989). On calculating the precision of consumer price indices. *Contributed Papers 47th Session of the ISI*, Paris.

Balk, B. (1991). Estimating the precision of a consumer price index: some experiences from the Netherlands. *Contributed Papers 48th Session of the ISI*, Cairo. Also presented at the Joint ECE/ILO Meeting on Consumer Price Indices, 18-21 November, Geneva. Modified version in *Netherlands Official Statistics*, 7(1), 48–49.

Bäckström, U. (1997). What Lessons Can be Learned from Recent Financial Crises? The Swedish Experience. *Speech at the Federal Reserve Symposium* Jackson Hole, Wyoming, USA, August 29, 1997. www.riksbank.se/pagefolders/1722/970829e.pdf

Berge, A. (2016). Viktfiffel. *Råd & Rön*, 19 April 2016. www.radron.se/artiklar/viktfiffel/ (accessed on July 26th 2017)

Betlehem, J. (2009). The rise of survey sampling. Statistics Netherlands, *Discussion paper* N° 09015. <https://hdl.handle.net/11245/1.312955>

Biemer, P., Trewin, D., Bergdahl, H. & Japac, L. (2014). A System for Managing the Quality of Official Statistics. *Journal of Official Statistics*, 30(3), 381–415. <https://doi.org/10.2478/jos-2014-0022>

Biemer, P. P. & Lyberg, L. E. (2003). *Introduction to Survey Quality*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Biemer, P. P., de Leeuw, E., Eckman, S., Edwards, B., Kreuter, F., Lyberg, L. E., Tucker, C. N. & West, B. T., Eds (2017). *Total Survey Error in Practice*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Bilius, Å., Bubuioc, R. & Tongur, C. (2017). Bestämning av prisvariabeln vid utökad användning av kassaregisterdata för viktvaror. Paper prepared for the CPI Board at Statistics Sweden.

<https://www.scb.se/contentassets/1b48f2064ebd46a78eda4d68d51c0403/3/3-1-bestamning-av-prisvariabeln-for-viktvaror.pdf>

Boskin, M. J., Dulberger, E. R., Gordon, R. J., Grilliches, Z. & Jorgenson, D. W. (1997). The CPI Commission: Findings and Recommendations. *The American Economic Review*, 87(2), 78–83. <https://www.jstor.org/stable/i352631>

Bressler, P. & Näslund, N. (2015). Här mäter P4 Kalmar filjmjölken – bara 9 dl i paketet. Sveriges Radio (Swedish National Radio), 11 May 2015. sverigesradio.se/sida/artikel.aspx?programid=86&artikel=6162534 (accessed on July 26th 2017)

Couper, M. (2013). Is the Sky Falling? New Technology, Changing Media, and the Future of Surveys. *Notes from talk before ESRA, European Survey Research Association in Ljubljana*, Slovenia, July 15-19, 2013. www.europeansurveyresearch.org/sites/default/files/files/Couper%20keynote.pdf

Dalén, J. & Ohlsson, E. (1995). Variance Estimation in the Swedish Consumer Price Index. *Journal of Business & Economic Statistics*, 13(3), 347–356. <https://www.jstor.org/stable/1392194>

Eurostat (2013). Compendium of HICP reference documents. Eurostat, *Methodologies and Working Papers*. <https://ec.europa.eu/eurostat/documents/3859598/5926625/KS-RA-13-017-EN.PDF/59eb2c1c-dal1f-472c-b191-3d0c76521f9b>

Eurostat (2017a). *Practical Guide for Processing Supermarket Scanner Data*.

Eurostat (2017b). *HICP Methodological Manual*.

- Englund, P. (2015).** The Swedish 1990s banking crisis. A revisit in the light of recent experience. Paper for the *Riksbank Macroeconomic Conference*, Stockholm 23-24 June 2015. www.riksbank.se/Documents/Avdelningar/AFS/2015/Session%201%20-%20Englund.pdf
- Food and Agriculture Organization of the United Nations (2012).** FAO/INFOODS Density Database Version 2.0 (2012), prepared by Charrondiere, R. U., Haytowitz, D. & Stadlmayr, B. <http://www.fao.org/docrep/017/ap815e/ap815e.pdf>
- Fisher, I. (1922).** *The Making of Index Numbers*. Boston, MA: Houghton-Mifflin.
- van der Grient, H. & de Haan, J. (2010).** The use of supermarket scanner data in the Dutch CPI. www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2010/zip.6.e.pdf
- van der Grient, H. & de Haan, J. (2011).** Scanner Data Price Indexes: The “Dutch Method” versus Rolling Year GEKS. <http://m.stats.govt.nz/ottawa-group-2011/~media/Statistics/ottawa-group-2011/Ottawa-2011-Presentations/deHaan-2011-presentation-Dutch-scanner-method.pdf>
- Horrigan, M. W. (2013).** *Big Data: A Perspective from the BLS*. *Amstat News*. magazine.amstat.org/blog/2013/01/01/sci-policy-jan2013/ (accessed on July 7th 2017)
- HUI Research (2017).** Dagligvarukartan 2016. Handels Utredningsinstitut. www.hui.se/statistik-rapporter/index-och-barometrar/dagligvarukartan (accessed on July 7th 2017)
- Hull, I., Löf, M. & Tibblin, M. (2017).** Webbinsamlade prisuppgifter och kortsiktiga inflationsprognoser. *Ekonomisk kommentar*, Sveriges Riksbank. www.riksbank.se/Documents/Rapporter/Ekonomiska_kommentarer/2017/rap_ek_kom_nr2_170609_sve.pdf
- ILO, IMF, OECD, UNECE, Eurostat, The World Bank (2004).** *Consumer price index manual: Theory and practice*. Geneva: International Labour Office.
- Johansen, I. & Nygaard, R. (2011).** Dealing with bias in the Norwegian superlative price index of food and non-alcoholic beverages. Paper written for the 2011 *Ottawa Group Conference*, Wellington, New Zealand, 2011. [http://www.ottawagroup.org/Ottawa/ottawagroup.nsf/4a256353001af3ed4b2562bb00121564/b1ab2e631d34d9bbca2578a7007fa493/\\$-FILE/2011%2012th%20meeting%20-%20Ing-vild%20Johansen,%20Ragnhild%20Nygaard%20\(Statistics%20Norway\)%20Dealing%20with%20bias%20in%20the%20Norwegian%20superlative%20price%20index%20of%20food%20and%20non-alcoholic%20beverages.pdf](http://www.ottawagroup.org/Ottawa/ottawagroup.nsf/4a256353001af3ed4b2562bb00121564/b1ab2e631d34d9bbca2578a7007fa493/$-FILE/2011%2012th%20meeting%20-%20Ing-vild%20Johansen,%20Ragnhild%20Nygaard%20(Statistics%20Norway)%20Dealing%20with%20bias%20in%20the%20Norwegian%20superlative%20price%20index%20of%20food%20and%20non-alcoholic%20beverages.pdf)
- Leaver, S. G. & Larson, W. E. (2001).** Estimating Variances for a Scanner-Based Consumer Price Index. Bureau of Labor Statistics. <https://www.bls.gov/osmr/research-papers/2001/st010130.htm>
- Neyman, J. (1934).** On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. *Journal of the Royal Statistical Society*, 97, 558–625. <https://doi.org/10.2307/2342192>
- Norberg, A. (2004).** Comparison of Variance Estimators for the Consumer Price Index. Paper presented at the 8th *Ottawa Group Meeting*, Helsinki 2004. <http://www.stat.fi/og2004/norbergp.pdf>
- Norberg, A., Sammar, M. & Tongur, C. (2011).** A Study on Scanner Data in the Swedish Consumer Price Index. Paper presented at the *Twelfth meeting of the Ottawa Group*, Wellington, New Zealand, 2011. www.ottawagroup.org/Ottawa/ottawagroup.nsf/home/Papers
- Norberg, A., Sammar, M. & Tongur, C. (2012).** Scanner data – comparability issues. Paper prepared for the CPI Board at Statistics Sweden. www.scb.se/contentassets/1b48f2064ebd46a78eda4d68d51c0403/scanner-data-comparability-issues.pdf
- Nygaard, R. (2010).** Chain Drift in Monthly Chained Superlative Price Index. UNECE, Geneva 2010. www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2010/zip.7.e.pdf
- Ohlsson, E. (1990).** Sequential Poisson Sampling from a Business Register and its Application to the Swedish Consumer Price Index. Statistics Sweden, *R&D Report* N° 1990/6. <https://www.scb.se/contentassets/7d78fc7dc1e643729f7e8388cd3adf32/rnd-report-1990-06-green.pdf>
- Rodriguez, J. & Haraldsen, F. (2005).** The use of scanner data in constructing elementary aggregates for food and beverages – ideas and experiences from Statistics Norway. Statistics Norway, *Unpublished report*.
- Rosén, B. (2000).** A User’s guide to Pareto PPS Sampling. Statistics Sweden, *R&D Report* N° 2000/6. <https://www.scb.se/contentassets/14f5e346f4814dd0acd52d10b23286c6/rnd-report-2000-06-green.pdf>

Sammar, M. & Norberg, A. (2012). Sammanvägningsmetod över tre veckor för kassaregisterdata i KPI. Paper prepared for the CPI Board at Statistics Sweden. www.scb.se/contentassets/1b48f2064ebd46a78eda4d68d51c0403/sammanvagningsmetod-over-tre-veckor-for-kassaregisterdata-i-kpi.pdf

SCB (2017). Kvalitetsdeklaration för KPI. Quality declaration for the Swedish CPI. www.scb.se/contentassets/a1e257bb3a574420b9d3f2ff59851c0a/pr0101_kd_2017.pdf

SOU (2012). Vad är officiell statistik? En översyn av statistiksystemet och SCB. *SOU* 2012/83. [https://www.regeringen.se/contentassets/3521811df5b34bd0bed672bd5c71c7f0/vad-ar-officiell-statistik-](https://www.regeringen.se/contentassets/3521811df5b34bd0bed672bd5c71c7f0/vad-ar-officiell-statistik-en-oversyn-av-statistiksystemet-och-scb-hela-dokumentet-sou-201283)

[en-oversyn-av-statistiksystemet-och-scb-hela-dokumentet-sou-201283](https://www.regeringen.se/contentassets/3521811df5b34bd0bed672bd5c71c7f0/vad-ar-officiell-statistik-en-oversyn-av-statistiksystemet-och-scb-hela-dokumentet-sou-201283)

Tongur, C. & Sandén, B. (2016). Viktvaror från kassaregisterdata. Paper prepared for the CPI Board at Statistics Sweden. <https://www.scb.se/contentassets/1b48f2064ebd46a78eda4d68d51c0403/utokning-av-livsmedel-i-kassaregisterdata.pdf>

United Nations, web page (2017). Detailed structure and explanatory notes. COICOP.

Wolter, K. M. (1985). *Introduction to Variance Estimation*. New York: Springer Verlag.
