

## Défis reliés à l'utilisation de données de lecteurs optiques pour l'Indice des prix à la consommation

Catherine Deshaies-Moreault et Nelson Emond<sup>1</sup>

### Résumé

Pratiquement tous les grands détaillants utilisent des lecteurs optiques pour enregistrer de l'information sur les transactions avec leurs clients (consommateurs). Ces données comprennent habituellement un code de produits, une brève description, un prix et une quantité vendue. Il s'agit d'une source hautement pertinente pour les programmes statistiques comme l'Indice des prix à la consommation (IPC) de Statistique Canada, l'un des plus importants indicateurs économiques au pays. L'utilisation des données de lecteurs optiques pourrait améliorer la qualité de l'IPC par, entre autres, l'augmentation du nombre de prix utilisés dans les calculs, une plus grande couverture géographique et l'utilisation des quantités vendues, tout en réduisant les coûts de collecte. Cependant, l'utilisation de ces données présente de nombreux défis. L'exploration de données de lecteurs optiques obtenues auprès d'un premier détaillant a démontré un taux de changement élevé au cours d'une année parmi les codes identifiant les produits. Les conséquences de ces changements posent des défis d'un point de vue de la classification des produits et de la qualité des estimations. Cet article se concentrera sur les enjeux reliés à l'acquisition, la classification et l'exploration de ces données afin d'évaluer la qualité pour leur utilisation dans l'IPC.

Mots clés : Indice des prix à la consommation; données de lecteurs optiques; classification.

### 1. Introduction

Diverses sources de données riches en information sont devenues accessibles au cours des dernières décennies avec l'évolution de la technologie. Les agences statistiques, désireuses de profiter de ces sources de données alternatives, innovent afin de continuer de produire des statistiques de haute qualité, sans augmenter indument les coûts. L'une de ces sources de données, déjà utilisée dans quelques pays, provient des lecteurs optiques. Les données de lecteurs optiques sont des enregistrements électroniques de transactions collectées en lisant le code-barres d'un produit. Elles contiennent de l'information, entre autres, sur les revenus totaux provenant des ventes ainsi que les quantités vendues des produits contenus dans le fichier, identifiés avec le Code Universel de Produits (CUP). Une telle source d'information pourrait profiter à divers programmes statistiques, notamment l'Indice des prix à la consommation (IPC). L'IPC est construit par classes d'agrégats élémentaires mutuellement exclusifs qui regroupent différents produits ayant des changements de prix homogènes. Le défi sera de faire le pont entre chaque produit provenant des données de lecteurs optiques et la classe au niveau de l'agrégat élémentaire. Un survol de l'IPC et un aperçu des données de lecteurs optiques peuvent être trouvés dans la section 2 et la section 3 respectivement. Quelques concepts clés à considérer lors de l'introduction des données de lecteurs optiques dans l'IPC sont revus dans la section 4. Un projet en 5 phases développé à Statistique Canada pour analyser et inclure des données de lecteurs optiques dans l'IPC publié mensuellement est décrit dans la section 5. Finalement, une brève conclusion est présentée dans la section 6.

### 2. Indice des prix à la consommation

L'IPC est un indicateur des variations dans les prix à la consommation subies par les ménages. L'IPC canadien mesure les variations de prix en comparant, à travers le temps, le coût d'un panier fixe de biens et services. L'IPC d'ensemble à l'échelle du Canada est fondé sur un échantillon annuel de plus de 950 000 relevés de prix. Certains prix sont collectés

---

<sup>1</sup>Catherine Deshaies-Moreault, Statistique Canada, 100 promenade Tunney's Pasture, Canada, K1A 0T6 ([catherine.deshaies-moreault@canada.ca](mailto:catherine.deshaies-moreault@canada.ca)); Nelson Emond, Statistique Canada, 100 promenade Tunney's Pasture, Canada, K1A 0T6 ([nelson.emond@canada.ca](mailto:nelson.emond@canada.ca)).

via des entrevues téléphoniques, proviennent de sources alternatives (telle qu'Internet), sont collectés par questionnaires ou proviennent d'autres programmes de Statistique Canada. Par contre, puisqu'environ 55% du panier provient de prix collectés en personne, beaucoup de ressources sont allouées à la collecte.

L'échantillonnage pour l'IPC est effectué à différents niveaux. Tout d'abord, certaines zones géographiques avec une activité économique importante sont sélectionnées. Ensuite, des points de vente sont échantillonnés à l'intérieur de ces zones. Afin d'obtenir un échantillon de prix par catégories de produits, les points de vente sont choisis en fonction du type de produits devant être collectés. Ceci permet d'obtenir un échantillon représentatif non seulement des produits et services que les ménages achètent, mais aussi de l'endroit où ils les achètent. Tous les produits ne pouvant pas être sélectionnés, une offre de produit est choisie à partir d'une définition de produits; le produit représentatif. Ce dernier est élaboré afin de représenter le mouvement de prix de plusieurs produits similaires. Le prix relevé pour le produit choisi est le prix affiché. Il n'y a ainsi aucune garantie qu'il y a effectivement eu des transactions pour ce produit lors du mois de collecte.

L'IPC suit la structure d'agrégation suivante : 8 composantes principales se divisant en divers agrégats intermédiaires, suivi des agrégats élémentaires. Le niveau d'agrégats élémentaires représente la classification la plus ventilée pour laquelle un estimé fixe des dépenses (quantités) est disponible. Le lecteur pourra consulter le document de référence sur l'indice des prix à la consommation canadien (Statistique Canada 2015) pour de plus amples informations.

### **3. Données de lecteurs optiques**

#### **3.1 Information disponible**

Tel que mentionné précédemment, les données de lecteurs optiques sont des enregistrements électroniques de transactions collectées en lisant le code-barres d'un produit. Une importante quantité d'information est ainsi maintenue et conservée par les détaillants dans les fichiers de lecteurs optiques. Voici quelques variables qu'on y retrouve typiquement: identifiant du magasin (incluant information d'adresse), date de la fin de la semaine, variables de classification du magasin, CUP, description du CUP, revenu des ventes et quantités vendues.

Notons que la date de la fin de la semaine indique que le fichier transmis ne contient pas toutes les transactions individuelles, mais plutôt un sommaire de toutes les ventes de chaque CUP, par semaine. Si un CUP n'a pas été vendu pour une semaine donnée, alors il n'y aura aucun enregistrement pour ce CUP durant cette semaine-là.

#### **3.2 Intérêt envers les données de lecteurs optiques**

Divers aspects de la qualité de l'IPC pourraient être améliorés avec l'utilisation des données de lecteurs optiques, notamment via la taille de l'échantillon, la représentativité, la couverture géographique et l'utilisation de valeurs unitaires, construites à partir de transactions et non de prix affichés. Tel que mentionné à la section 2, la collecte actuelle de l'IPC implique la sélection de produits et de points de vente dans certaines zones géographiques spécifiques. L'univers des produits d'un magasin donné sera potentiellement disponible ainsi que l'ensemble approximatif des magasins, avec les données de lecteurs optiques. Les magasins utilisant un système de lecteurs optiques et ceux qui fourniront leurs données à Statistique Canada représenteront l'univers approximatif des magasins. La taille de l'échantillon obtenue sera ainsi nettement supérieure à celle pouvant être collectée en personne, tout en réduisant de façon importante les coûts de collecte, avec l'arrêt de la collecte en personne.

Puisque des données pour potentiellement tous les produits vendus d'un magasin en particulier pourraient être obtenues, un gain du côté de la représentativité serait remarqué. En effet, au lieu de sélectionner une ou quelques offres de produits par agrégat élémentaire, plusieurs ou tous les produits pourraient être utilisés. De plus, ayant l'information sur les quantités vendues, les produits pourraient être pondérés selon les quantités vendues afin que les produits les plus achetés par les consommateurs apportent une juste contribution à l'indice.

De même, il n'y a aucune limite de collecte physique avec les données de lecteurs optiques, ce qui permettra d'augmenter la couverture géographique. Aussi, les données de lecteurs optiques remplaçant en partie la collecte en

personne, les erreurs non dues à l'échantillonnage, par exemple des erreurs de saisie, s'en trouveront grandement diminuées. Il serait toutefois irréaliste de supposer une absence totale d'erreurs dans les fichiers de lecteurs optiques.

Finalement, il pourrait être d'intérêt pour d'autres programmes de Statistique Canada d'utiliser des données de lecteurs optiques. Pensons notamment à l'enquête sur les dépenses des ménages, les enquêtes portant sur le commerce de détail et le système de comptabilité nationale.

## **4. Revue des concepts**

Afin de vérifier si la forme des données de lecteurs optiques répond toujours aux principes conceptuels de l'IPC canadien, certains points importants doivent être comparés, notamment en ce qui a trait aux prix et produits collectés, au panier fixe, ainsi que le délai afin d'obtenir les données.

La collecte actuelle de l'IPC implique de collecter des prix de certains produits représentatifs. En effet, les interviewers se rendent dans des points de vente pour noter les prix inscrits de produits donnés, à un jour donné. Bref, il s'agit d'un prix ponctuel saisi au même moment d'un mois à l'autre, par exemple le mardi de la semaine 2. Avec les données de lecteurs optiques, les prix reçus sont dérivés à partir des revenus des ventes de la semaine et des quantités vendues – ce sont les prix unitaires hebdomadaires. Cette différence est importante dans l'univers des prix et de nombreuses formules ont été développées afin d'accommoder les prix unitaires. Ce concept n'est pas le point central de cet article, mais le lecteur pourra se référer, notamment, à Richardson (2000). Aussi, les prix pour tous les CUP ayant des ventes non nulles sont reçus, éliminant ainsi la contrainte de se restreindre à une liste de produits représentatifs. Par contre, étant donné la réception des données hebdomadaires, un enjeu d'agrégation des CUP est maintenant présent. À la section 5.2.4, divers scénarios d'identification des produits à partir des CUP sont explorés. Il y a aussi un enjeu quant aux ventes nulles – sont-elles temporaires, justifiant ainsi un traitement pour valeur manquante, ou permanente, justifiant un remplacement de produits et les ajustements subséquents (dits « ajustements de qualité », voir Statistique Canada 2015)?

Les quantités implicites des dépenses au niveau des agrégats élémentaires sont fixes dans la définition actuelle de l'IPC canadien. Étant donné que l'intention est de classer les CUP sous les agrégats élémentaires, le concept de panier fixe sera respecté. À long terme, étant donné que de l'information détaillée sur les dépenses sera disponible avec les données de lecteurs optiques, il y a possibilité de revoir les pondérations (basées sur les quantités), utilisées pour combiner les différents types de produits, ou même de restructurer le panier.

Notons que les données seront transmises à chaque semaine; ainsi, les délais pour la publication mensuelle pourront être respectés. Il n'est pas déterminé si les données d'un mois complet seront utilisées ou si, pour des contraintes de temps de traitement et d'analyse, seules les deux ou trois premières semaines serviront à la compilation de l'IPC.

## **5. Projet en 5 phases**

Afin d'inclure les données de lecteurs optiques dans l'IPC, un projet en 5 phases a été développé par une équipe multidisciplinaire de Statistique Canada. Les sous-sections suivantes couvriront ces phases ainsi que les divers enjeux rencontrés.

### **5.1 Phase 1 – recherche exploratoire**

D'autres agences statistiques utilisent déjà les données de lecteurs optiques dans leur IPC national. Citons notamment les Pays-Bas (Schut, 2001 et van der Grient et de Haan, 2010) qui ont déjà publiés de nombreux articles à ce sujet. D'une agence à l'autre, la stratégie pour utiliser les données de lecteurs optiques, la provenance de ces données ou la méthode de collecte varient. Certaines agences achètent des données de lecteurs optiques traitées par un tiers, ou encore reçoivent seulement un échantillon de données spécifié aux détaillants. Quelques agences utilisent notamment les variables de classification des magasins au lieu d'aligner les CUP à leur classification traditionnelle pour l'IPC. Le contexte canadien est différent sur bien des aspects : les produits de l'échantillon sont confidentiels (donc

Statistique Canada ne peut pas fournir une liste de produits désirés aux détaillants) et le Canada a une très grande superficie, menant naturellement à une ample variété dans les détaillants – utiliser la classification des magasins plutôt que la classification actuelle de l’IPC pourrait devenir incontrôlable.

## **5.2 Phase 2 – Acquisition, analyse, étude de faisabilité et études pilotes**

Une première entente a été conclue avec un important détaillant dans l’industrie alimentaire canadienne afin de recevoir des données de lecteurs optiques pour un échantillon de 6 magasins, sur 2 ans. Les analyses et études de faisabilités décrites ci-dessous sont basées sur ces données.

### **5.2.1 Enjeu d’acquisition des données**

Un premier enjeu relié aux données de lecteurs optiques est l’acquisition des données. Dans un premier temps, Statistique Canada a l’autorité de collecter des données (Statistique Canada, 2016) selon la Loi sur la Statistique, mais a aussi le devoir de se soucier du fardeau des répondants. L’organisation veut s’assurer de maintenir des relations positives avec ses répondants, et désire établir un cadre propice à une collaboration fructueuse dès le début du projet d’intégration de données de lecteurs optiques. Ensuite, il s’agit d’identifier quelles données l’agence aimerait acquérir. Bien que simple à première vue, cette question se complexifie étant donné la diversité possible des standards d’entreposage de telles données dans l’industrie (où chaque répondant pourrait utiliser son propre dictionnaire de données). L’expérience des autres pays apporte ici un bon point de départ pour identifier quelles variables sont habituellement présentes sur un fichier de données de lecteurs optiques et pertinentes pour les besoins d’un programme comme l’IPC. De plus, en raison du volume important de ces fichiers, des enjeux informatiques sont rencontrés tels que des enjeux de transmission, d’entreposage et de traitement des données.

### **5.2.2 Enjeu d’exploration des données**

Les fichiers reçus sont assez imposants, soit 9 million d’enregistrements en un an, incluant 190 000 CUP uniques. En moyenne, 85 000 CUP uniques sont présents mensuellement. La distribution des CUP est fortement asymétrique. Effectivement, seulement 3,5% des CUP sont observés à chaque mois dans les 6 magasins, et ils représentent 41% des ventes totales à eux seuls. Ceci indique qu’une quantité importante des CUP pourrait être retirée des fichiers avec un impact minimal sur les ventes totales. Il faut cependant réduire la taille du fichier de façon prudente. Certains CUP pourraient avoir un impact minime sur les ventes totales, mais être importants à l’intérieur de leur catégorie de produits. Cet enjeu mène à un questionnement du type « l’œuf ou la poule », dans le sens où la taille du fichier doit être réduite pour faciliter le travail de classification, mais les CUP doivent être classifiés afin de déterminer lesquels peuvent être retirés sans impact néfaste à la qualité de l’indice. Notons aussi que les CUP présentent un taux de renouvellement élevé, avec seulement 33% des CUP présents deux années consécutives dans un magasin donné. Ceci suggère que le processus de classification devra être effectué en continu.

### **5.2.3 Enjeu de classification des données**

Afin d’être utilisables dans la construction d’un IPC, les données de lecteurs optiques doivent être classifiées au niveau le plus fin de la structure d’agrégation, ce qui équivaut aux agrégats élémentaires dans le cas présent. Deux types de variables retrouvés dans les fichiers de données de lecteurs optiques peuvent être utiles à la classification, soient : les variables de classification du magasin et la description du CUP. Les techniques traitant les chaînes de caractères sont donc appropriées pour répondre à cet enjeu. Il serait intéressant d’utiliser un système généralisé afin d’effectuer la tâche de classification. De tels systèmes existent à Statistique Canada afin d’effectuer de l’appariement et du codage de données. Ceci étant dit, la classification est un enjeu différent, requérant un outil alternatif. Certains systèmes de classification existent aussi, mais sont plutôt adaptés pour traiter des classifications stables, et non pas révisées continuellement. L’idée d’un dictionnaire semble plutôt prometteuse, mais représenterait des efforts de maintenance considérables. Ainsi, le processus doit rester assez générique pour s’adapter aisément aux changements. Il n’est pas si simple de développer une solution générique s’adaptant à tous les types de produits. Par exemple, utiliser les marques de produits pour aider à la classification semble une bonne idée à première vue. Bien que cela soit le cas pour certaines marques associées à des produits très spécifiques, l’idée échoue dans le cas de marques rattachées à un large éventail de produits différents, alors classifiés à des endroits distincts dans la structure de l’IPC.

Le processus de classification actuellement en développement fera l'utilisation des variables de classification du détaillant et de la description des CUP, afin de les diriger vers l'agrégat élémentaire de l'IPC approprié. De plus, étant donné que les liens entre les données de lecteurs optiques et les agrégats de l'IPC se font sur des mots, il pourrait y avoir plusieurs mots qui sont liés. Afin de prioriser les mots liés, une pondération sur la fréquence d'apparition des mots sera effectuée; moins un mot est fréquent, plus il aura un poids important s'il est lié. Cette idée diminuera l'impact des mots fréquemment retrouvés dans le vocabulaire comme adjectifs, par exemple, des saveurs (« framboise », « orange », etc.) ou un état (« frais », « congelé », etc.). Des méthodes plus complexes, telles que la classification par apprentissage automatique, ont brièvement été considérées et seront évaluées en profondeur dans le futur.

## 5.2.4 Simulation d'indices

Tel que mentionné précédemment, les données de lecteurs optiques sont reçues de façon hebdomadaire. Afin d'avoir un prix mensuel, il faut donc agréger les prix hebdomadaires d'une certaine manière. De plus, puisque le concept de produit est différent avec les données de lecteurs optiques, la question se pose à savoir quelle est la nouvelle définition de produit : est-ce un CUP en particulier, ou un regroupement de CUP? Divers scénarios pour la définition du produit et la façon d'agréger les données hebdomadaires sont présentés ci-après. Mais tout d'abord, la formule de Jevons servant à calculer l'indice de prix de la majorité des agrégats élémentaires est introduite à la formule 5.1.

### Formule 5.1

#### Indice de Jevons

$$I_j^{0:t} = \prod_{i=1}^n \left( \frac{p_i^t}{p_i^0} \right)^{1/n}$$

Où :

$I_j^{0:t}$  est l'indice de l'agrégat élémentaire  $j$  entre les périodes 0 et  $t$ ;

$p_i^t$  est le prix du produit  $i$  au temps  $t$ ;

$p_i^0$  est le prix du produit  $i$  au temps 0;

$n$  est le nombre de produits dans l'agrégat  $j$ .

Le tableau 5.1 illustre les différents scénarios proposés concernant la définition de produit et la façon d'agréger les prix hebdomadaires. Les définitions de  $p_i$  et  $n$  font référence à la formule de Jevons définie précédemment. On cherche à obtenir un  $p_i$  mensuel ; ainsi, l'agrégation est toujours effectuée sur les semaines, et aussi sur les CUP si la définition de produits l'indique.

Tableau 5.1

#### Scénarios proposés

	Formule	Définitions	Produit
1	$p_i^t = \sum_{y=1}^r \sum_{k=1}^{m_y} \text{prix unitaire}_{y,k,t} / \sum_y m_y$	$r$ : nombre de CUP au mois $t$ $m_y$ : nombre de semaines dans le mois $t$ avec ventes pour le CUP $y$ $n = 1$	Tous les CUP similaires représentent le même produit ( $n=1$ )
2	$p_i^t = \frac{\sum_{k=1}^m \text{prix unitaire}_{i,k,t} * \pi_i^k}{\sum_k \pi_i^k}$	$m$ : nombre de semaines dans le mois $t$ avec des ventes pour le CUP $i$ $n$ : nombre de CUP au mois $t$ $\pi_i^k$ : quantités vendues pour le CUP $i$ à la semaine $k$	Chaque CUP est un produit différent
3	$p_i^t = \prod_{k=1}^m (\text{prix unitaire}_{i,k,t})^{1/m}$	$m$ : nombre de semaines dans le mois $t$ avec des ventes pour le CUP $i$ $n$ : nombre de CUP au mois $t$	Chaque CUP est un produit différent
4	$p_i^t = \text{prix unitaire}_{i,k,t}$	$n = 1$	Le prix unitaire d'une seule semaine donnée pour un seul CUP choisi est le produit représentatif

Les différents scénarios seront expliqués en s'appuyant sur l'exemple fictif d'extrait de fichier de données de lecteurs optiques, contenu dans le tableau 5.2. Certaines variables typiquement trouvées dans les fichiers de lecteurs optiques y sont représentées, incluant le prix unitaire qui est dérivé à partir du revenu des ventes divisé par les quantités vendues.

**Tableau 5.2**  
**Exemple de données de lecteurs optiques**

Date de fin de la semaine	Classification du magasin	CUP	Description du CUP	Revenu des ventes	Quantités vendues	Prix unitaire
6FEV15	SOUPE EN CONSERVE	123	CRM DE BROCOLI	12\$	12	1,00\$
13FEV15	SOUPE EN CONSERVE	123	CRM DE BROCOLI	10\$	10	1,00\$
20FEV15	SOUPE EN CONSERVE	123	CRM DE BROCOLI	11,25\$	15	0,75\$
27FEV15	SOUPE EN CONSERVE	123	CRM DE BROCOLI	10\$	8	1,25\$
6FEV15	SOUPE EN CONSERVE	456	BOUILLON DE POULET	9\$	6	1,50\$
13FEV15	SOUPE EN CONSERVE	456	BOUILLON DE POULET	9,80\$	20	0,49\$
27FEV15	SOUPE EN CONSERVE	456	BOUILLON DE POULET	9,95\$	5	1,99\$

Les scénarios relatifs au concept de produit et à l'agrégation des prix hebdomadaires proposés sous le tableau 5.1 sont décrits plus en détails ci-après.

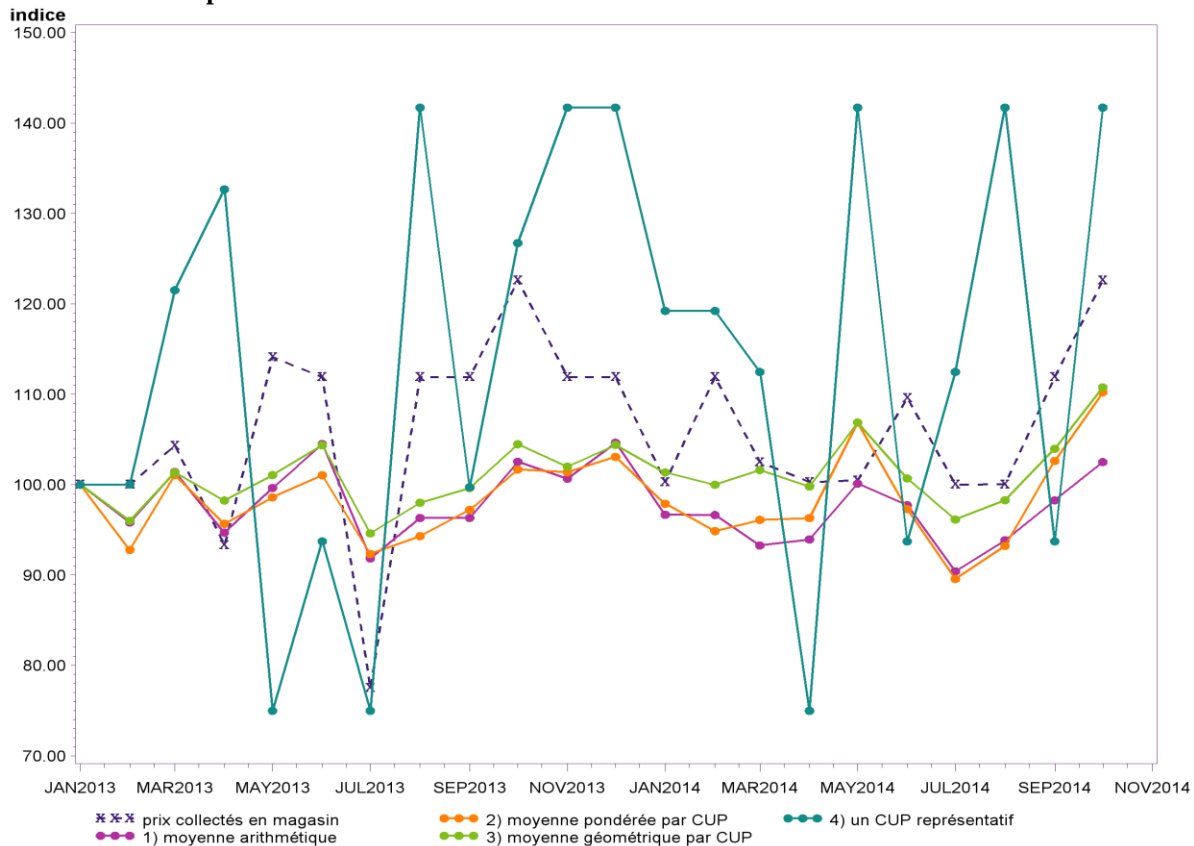
Le scénario 1 considère que le produit est un regroupement de CUP semblables. Si l'on se réfère à l'exemple du tableau 5.2, les *soupes en conserve* seraient le produit ici. Le prix mensuel du produit est obtenu en effectuant une moyenne sur tous les prix unitaires.

Dans les scénarios 2 et 3, chaque CUP est un produit différent. Dans le tableau 5.2, il y aurait ainsi deux produits distincts, soient les CUP 123 et 456. Le prix mensuel de chaque CUP du scénario 2 est obtenu en calculant une moyenne pondérée sur les quantités vendues, afin d'obtenir un prix représentatif du comportement des acheteurs. Le prix mensuel de chaque CUP sous le scénario 3 est dérivé à partir d'une moyenne géométrique des prix unitaires, pour respecter l'esprit de la formule de Jevons.

Finalement, sous le scénario 4, le prix unitaire d'un seul CUP à une semaine donnée est choisi pour représenter le prix pour le mois d'un groupe de produits similaires. Dans le tableau 5.2, ce pourrait être par exemple le prix unitaire du CUP 456 du 13 février 2015 qui représenterait le prix du mois de février pour les *soupes en conserve*. Ce scénario se rapproche de la collecte terrain actuelle, où le prix d'un produit représentatif donné est collecté.

Les différents scénarios énumérés ci-haut pour les données de lecteurs optiques sont comparés à l'indice actuel de données collectées en personne pour l'IPC dans le graphique 5.1. Les données représentées correspondent à un produit donné, dans un magasin en particulier. Pour le scénario 4, l'UPC ayant le plus de ventes la première semaine est suivi de mois en mois (en conservant toujours le prix de la première semaine du mois, afin d'être cohérent avec la fréquence de collecte de l'IPC pour les produits alimentaires).

**Graphique 5.1**  
**Indices de Jevons pour différents scénarios**



Tel que constaté à la figure précédente, les indices utilisant un seul produit (produit représentatif en magasin et scénario 4) sont beaucoup plus volatiles que ceux dérivés à partir des autres scénarios. Mentionnons que l'IPC actuel n'est pas aussi volatile que l'exemple illustré ci-haut, puisqu'il combine plusieurs produits différents collectés à travers de nombreux points de ventes.

### 5.3 Mise en œuvre et développement des systèmes

Les données de lecteurs optiques seront introduites graduellement dans le calcul de l'IPC à travers différentes phases de mise en œuvre. Les phases 3, 4 et 5 portant sur la mise en œuvre et le développement des systèmes sont décrites dans cette section.

#### 5.3.1 Phase 3 – Acquisition de données de production et mise en œuvre simple et systématique

La phase 3 du projet consiste en l'acquisition de données de production, le développement des systèmes nécessaires et l'intégration des données de lecteurs optique. À court terme, les données de lecteurs optiques d'un premier détaillant pourraient être introduites dans le calcul de l'IPC, dès qu'elles seront disponibles de façon régulière. Le scénario envisagé est de remplacer chaque prix collecté dans les magasins de ce détaillant faisant partie de l'échantillon actuel, par des données de lecteurs optiques. Cette première introduction aura l'avantage d'immédiatement diminuer les coûts de collecte et ne nécessitera pas de changements aux systèmes actuels (ou très peu). Par contre, cette dernière n'augmentera pas la taille de l'échantillon et n'utilisera pas les quantités vendues afin de pondérer les différents produits. Une fois la phase 3 réalisée, le but sera ensuite d'utiliser les données de lecteurs optiques afin d'augmenter la taille de l'échantillon de l'IPC, puis d'introduire une méthode de mise en œuvre tirant avantage de l'information sur les quantités vendues disponible, tout en étant compatible avec le concept de « panier fixe ».

### 5.3.2 Phase 4 – Mise en œuvre simple et systématique étendue

La phase 4 va étendre la méthode développée à la phase 3 à plus de détaillants, et ce, à travers diverses industries (non seulement l'industrie alimentaire). Par contre, les données ne seront encore pas utilisées à leur plein potentiel – ceci sera l'objet de la phase 5.

### 5.3.3 Phase 5 – Mise en œuvre avancée

Afin d'utiliser les données de lecteurs optiques à leur plein potentiel, les formules d'indices pour l'IPC canadien ainsi que les concepts seront réévalués. Puisque l'information sur les quantités vendues est disponible, les données de lecteurs optiques ouvrent la porte à l'adoption de formules utilisant plus d'information, telles que les formules d'indices « superlatifs », qui sont des formules considérées meilleures par rapport à celles normalement utilisées, suivant le concept d'un « panier fixe ». En ce moment, il y a de nombreuses recherches sur les données de lecteurs optiques et le développement des formules supérieures. Le but de la phase 5 sera d'analyser ces nouvelles formules et de les mettre en œuvre.

## 6. Conclusion

Les données de lecteurs optiques présentent de nombreux intérêts pour le programme de l'IPC. Notamment, elles serviront à améliorer la qualité et la représentativité de l'Indice et aideront à la réduction des coûts de collecte. L'utilisation des données de lecteurs optiques résultera aussi en la diminution des erreurs non dues à l'échantillonnage. D'un autre côté, elles apportent leur lot de défis. Une fois les enjeux liés à l'acquisition et l'exploration des données surmontés, il restera à résoudre le défi de la classification des données avec un taux de succès satisfaisant, afin qu'elles soient utilisables dans l'IPC. Afin de compléter cet aspect du travail, des techniques d'apprentissage automatique sont considérées et en cours d'exploration. De plus, il faudra répondre adéquatement à la nouvelle forme des données, qui présentent des prix unitaires hebdomadaires et soulèvent différents scénarios possibles pour définir un produit et agréger les données. Des données de production doivent être reçues de façon régulière avant de débiter la mise en œuvre simple et systématique. En parallèle, un plan de contingence devra être développé afin de réagir rapidement à l'éventualité d'une coupure dans la transmission des données. Finalement, l'acquisition de données d'autres détaillants est envisagée afin d'élargir la couverture du marché et maximiser l'utilisation des données de lecteurs optiques. À long terme, l'idée est d'amener l'IPC vers des formules supérieures, qui utilisent le plein potentiel de cette riche source de données.

## Bibliographie

Richardson, David H. (2000), « Scanner Indexes for the CPI », article présenté au *Sixth Meeting of the International Working Group on Price Indices*, Australie.

Schut, Cecile. (2001), « Using scanner data to compile price indices: practical problems », article présenté au *Joint Statistical Commission and Economic Commission for Europe / International Labour Organisation (ECE/ILO) Meeting on Consumer Prices*, Genève.

Statistique Canada (2015). *L'indice des prix à la consommation canadien document de référence*. Disponible au : <http://www.statcan.gc.ca/pub/62-553-x/62-553-x2015001-fra.htm>.

Statistique Canada (2016). *Guide de la Loi sur la Statistique*. Disponible au: [http://icn-rci.statcan.ca/10/10b/10b\\_001-fra.html](http://icn-rci.statcan.ca/10/10b/10b_001-fra.html).

van der Grient, Heymerick et de Haan, Jan. (2010), « The use of supermarket scanner data in the Dutch CPI », article présenté au *Joint ECE/ILO Workshop on Scanner Data*, Genève.