

Défis de la recherche méthodologique dans le domaine des statistiques officielles

Kees Zeelenberg¹

Résumé

Il existe plusieurs domaines et sujets de recherche méthodologique en statistiques officielles. Nous expliquons pourquoi ils sont importants, et pourquoi il s'agit des plus importants pour les statistiques officielles. Nous décrivons les principaux sujets dans ces domaines de recherche et nous donnons un aperçu de ce qui semble le plus prometteur pour les aborder. Nous nous penchons ici sur :

- la qualité des comptes nationaux, et plus particulièrement le taux de croissance du revenu national brut;
- les mégadonnées, et plus particulièrement la façon de créer des estimations représentatives et de tirer le meilleur parti possible des mégadonnées, lorsque cela semble difficile ou impossible;

Nous abordons aussi :

- l'amélioration de l'actualité des estimations statistiques provisoires et finales;
- l'analyse statistique, plus particulièrement des phénomènes complexes et cohérents.

Ces sujets font partie de l'actuel programme de recherche méthodologique stratégique qui a été adopté récemment à Statistique Pays-Bas.

Mots-clés : méthodologie statistique; recherche; statistiques officielles.

1. Introduction

Le présent document porte sur quatre questions :

1. Quels sont les principaux défis liés aux statistiques officielles?
2. Comment la méthodologie peut-elle aider à surmonter ces défis?
3. Comment les mégadonnées peuvent-elles aider à surmonter ces défis?
4. Quelle est la meilleure stratégie de recherche pour les méthodologistes qui traitent des mégadonnées?

Ainsi, nous partirons de points assez généraux, puis nous mettrons l'accent sur la méthodologie et sur les mégadonnées. Dans le document, nous tentons de présenter cela sous forme d'aperçu, du point de vue de ce qui semble important en matière de mégadonnées et de statistiques officielles. Les exemples, ainsi que les principaux points découlent ou dépendent du programme de recherche méthodologique et du programme d'innovation de Statistique Pays-Bas². Dans certains cas, il s'agit d'une perspective plus personnelle, dans un domaine où nous n'effectuons pas encore de recherche très active.

2. Défis pour les statistiques officielles

Je mettrai l'accent sur deux des défis liés aux statistiques officielles :

- la qualité;
- les besoins des utilisateurs.

¹Kees Zeelenberg, Statistique Pays-Bas, C.P. 24500, 2490 HA, La Haye, Pays-Bas (k.zeelenberg@cbs.nl).

² De façon plus particulière, les personnes suivantes ont participé : Bart Bakker, Barteld Braaksma, Jan van den Brakel, Bart Buelens, Joep Burger, Jacco Daalmans, Piet Daas, Maaïke Hersevoort, Jeroen Pannekoek, Frank Pijpers, Marko Puts, Sander Scholtus, Ton de Waal; voir aussi la bibliographie.

Évidemment, la qualité est notre « créneau ». C'est ce qui nous distingue de nombreux producteurs de données. Il sera démontré ci-après que nous avons certains problèmes de qualité, par exemple, au sujet du PNB et de son taux de croissance.

En deuxième lieu, de nos jours, les utilisateurs ne sont pas tant à la recherche de données que d'*information*. Ils veulent en savoir davantage au sujet de ce qui sous-tend les données. Ils souhaitent obtenir cette information plus rapidement, et qu'elle soit fondée sur des données de meilleure qualité. Ce besoin est motivé principalement par la complexité de plus en plus grande de la société et le rythme plus rapide auquel elle évolue. Cela fait en sorte que les INS doivent mettre l'accent sur trois sujets :

- l'actualité;
- l'information et l'analyse;
- les relations complexes dans la société.

Les mégadonnées sont de nouvelles sources d'information et fournissent de nouvelles façons de créer l'information, ce qui fait qu'elles peuvent être très utiles. Cela s'appelle l'utilisation analytique des mégadonnées.

Pour l'analyse de relations complexes et pour des contributions plus opportunes aux débats de la société, il faut tout d'abord que nous puissions utiliser des liens et combiner des données de diverses bases. C'est ce que l'on appelle un *lac de données* et ce que nous tentons de créer; pas une base de données unique pouvant servir à toutes les fins, mais plutôt un système de bases de données pouvant être couplées de façon souple et à partir desquelles des données peuvent être rapidement extraites.

L'analyse des relations complexes et l'accent qui est mis sur l'information font en sorte que l'on se concentre davantage sur les méthodes, ce qui fait que, même si on laisse de côté les mégadonnées, les méthodologistes devront jouer un rôle important. Outre des méthodes plus précises pour les mégadonnées (voir la section 4), des méthodes multiniveaux semblent prometteuses dans ce cas.

Nous devons aussi apprendre à créer et à utiliser des classifications souples. Dans le cas des entreprises, la classification type des industries (la CITI et ses contreparties nationales ou régionales) ne doit plus occuper une position centrale, étant donné que pour de nombreux phénomènes complexes, d'autres classifications sont beaucoup plus importantes, par exemple, les petites/grandes entreprises, les entreprises d'exportation/nationales, et les entreprises utilisant des TI/produisant des TI. La classification industrielle type est aussi démodée au niveau conceptuel : les entreprises sont classées selon leur activité principale, mais de nos jours, elles changent plus fréquemment d'activités, et leurs activités auxiliaires sont parfois aussi importantes que leur activité principale. En outre, d'un point de vue sociétal, il est beaucoup plus important de connaître les produits et services fournis et demandés que l'industrie dans laquelle ils sont produits. Encore une fois, cela entraînera une utilisation plus grande des méthodes statistiques, comme l'analyse par grappes et d'autres méthodes de classification, ainsi que méthodes d'analyse.

Même si les méthodes de la section 4 ne sont pas utilisables pour les statistiques officielles (descriptives), elles pourraient être mises à contribution de façon profitable pour les analyses statistiques et, ainsi, pour l'analyse des relations complexes dans la société.

Évidemment, d'autres défis se posent, par exemple :

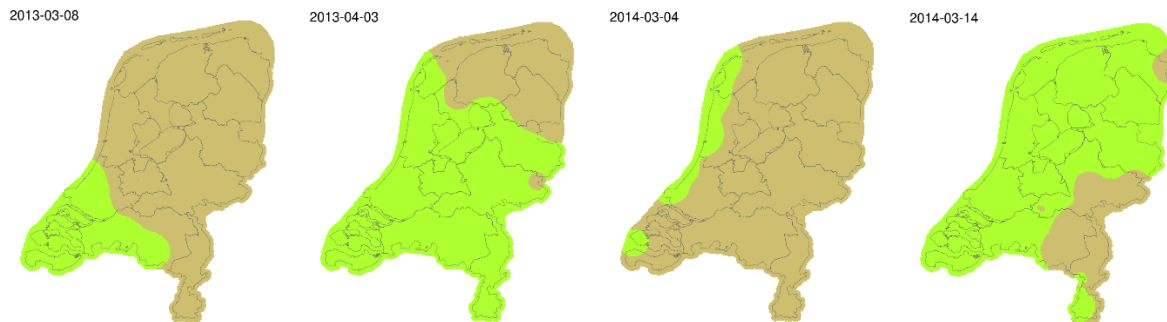
- la non-réponse, spécialement aujourd'hui, étant donné que nous avons une stratégie de recours à *Internet en premier* pour la collecte des données, les enquêtes sur le Web étant destinées à prendre de plus en plus la place des autres enquêtes;
- le couplage, et plus particulièrement pour les mégadonnées, ces dernières ne comprenant pas de variable de couplage claire, comme le numéro de sécurité sociale; l'appariement statistique, le couplage probabiliste et l'imputation sont par conséquent devenus beaucoup plus pertinents;
- la protection de la vie privée, qui continue d'être très importante auprès du grand public, et plus encore aujourd'hui à l'ère des mégadonnées et des données ouvertes;
- et, évidemment, les restrictions budgétaires, qui incitent à une plus grande efficacité.

Les méthodologistes peuvent apporter une contribution importante, et nous nous penchons sur ces sujets. Toutefois, je ne les aborderai pas ici et je mettrai l'accent sur la qualité et les besoins des utilisateurs (information plutôt que données).

3. Exemple : L'arrivée du printemps

Laissez-moi commencer en utilisant un exemple intéressant, qui ne repose pas sur des données d'enquête ou des données administratives, mais utilise plutôt une forme de données brutes trouvées dans Internet, et voyons ce que nous pouvons en tirer. L'exemple concerne l'anémone des bois. Il s'agit de la première fleur à éclore au printemps en Europe de l'Ouest. Les dates des premières floraisons sont enregistrées par des bénévoles, qui se rendent dans les parcs et dans les bois pour observer les fleurs³. Il est donc possible d'utiliser ces dates de premières floraisons pour savoir quand le printemps arrive⁴. L'analyse a été effectuée par Maaïke Hersevoort (Statistique Pays-Bas) et Hamed Mehdipoor (Université Twente).

Figure 3-1.
Arrivée du printemps aux Pays-Bas, 2013 et 2014



La figure 3-1 montre l'arrivée du printemps, déterminée par la floraison de l'anémone des bois. En 2013, le printemps commence dans le sud, et en 2014, dans l'ouest. En outre, en 2013, le printemps commence une semaine et demie plus tard qu'en 2014 et se termine environ trois semaines plus tard; c'est donc dire qu'en 2013, le printemps arrive plus tard et que cette arrivée se fait d'abord plus rapidement, et plus lentement par la suite qu'en 2014.

Que pourrions-nous tirer de cet exemple?

- Les données sont recueillies par des bénévoles. Les observateurs ne sont pas les mêmes chaque année, ils ne sortent pas tous les jours, ni au même moment de la journée. C'est donc dire que ces données ne sont pas réparties également dans le temps et dans l'espace. En termes plus familiers, nous sommes en présence de non-réponse et de sélectivité.
- C'est pourquoi les auteurs de cet exemple ont dû avoir recours à la modélisation pour améliorer la qualité.
- Ils ont toutefois aussi déterminé que la modélisation nécessite beaucoup de main-d'œuvre, c'est-à-dire que l'élaboration d'un processus de production n'est pas toujours simple ou peu coûteux.
- Ils ont donc créé de l'*information* concernant l'arrivée du printemps, ne se limitant pas à recueillir ou épurer un ensemble de données.

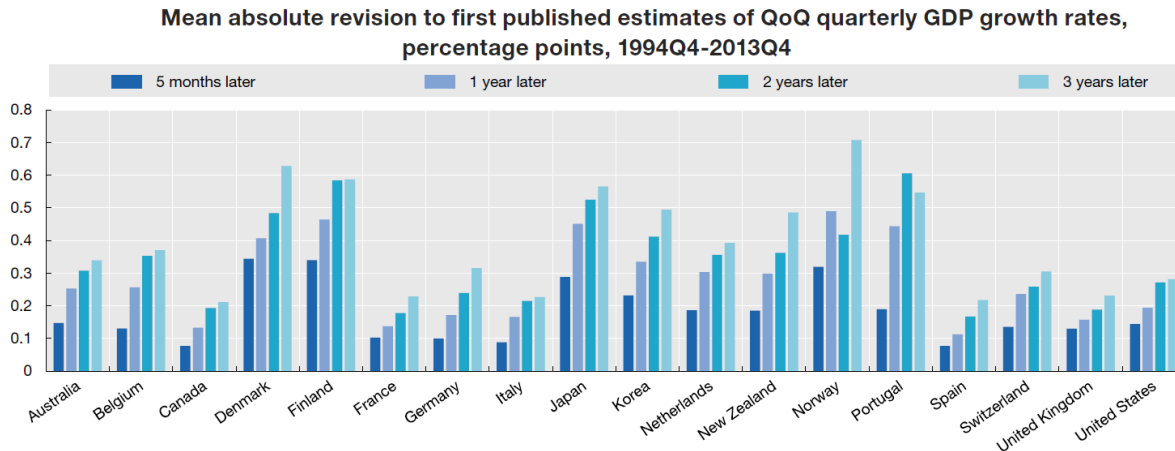
Voici certains des problèmes dont je parlerai.

³ Les données se trouvent à www.natuurkalender.nl.

⁴ En fait, l'arrivée du printemps est le sujet de l'un des poèmes les plus connus en langue hollandaise : Mai, par Herman Gorter; voir Gorter (2015).

4. Qualité du PNB

Figure 4-1.



La figure 4-1 (figure 3 de Zwijnenburg, 2015) montre, pour près de 20 États membres de l'OCDE, la révision absolue moyenne (RAM) de la première estimation de la croissance du PNB, pour plusieurs périodes. Concentrons-nous sur les révisions finales après trois ans, soit le trait en bleu pâle, qui est situé le plus à droite pour chaque pays. Tout d'abord, le Canada, la France, l'Allemagne, l'Italie, l'Espagne, la Suisse, le Royaume-Uni et les États-Unis affichent les révisions les plus faibles, avec une RAM moyenne inférieure à 0,25 %. Un deuxième groupe, constitué de l'Australie, de la Belgique, des Pays-Bas et de la Nouvelle-Zélande, présente des révisions de taille moyenne, avec une RAM se situant entre 0,25 % et 0,35 %. Un troisième groupe, constitué du Danemark, de la Finlande, du Japon, de la Corée, de la Norvège et du Portugal, enregistre les révisions les plus élevées, avec une RAM moyenne supérieure à 0,35 %. Ainsi, pour plusieurs pays, les révisions sont assez importantes, les Pays-Bas se situant quelque part au milieu, tant du point de vue du pourcentage que de la taille de la révision.

Toutefois, pour les Pays-Bas, la situation est beaucoup moins bonne lorsqu'il s'agit de la révision moyenne, ce qui fait que si l'on ne tient pas compte des valeurs absolues (Zwijnenburg, 2015, figure 2), les Pays-Bas figurent parmi les pays ayant le pire rendement, avec une révision moyenne de 0,1 %. En outre, près du tiers de l'incertitude dans les prévisions du PNB par l'agence nationale chargée des prévisions (CPB) est attribuable à des révisions des estimations antérieures du PNB par Statistique Pays-Bas (Elbourne et coll., 2015). Avec une croissance moyenne d'un trimestre à l'autre de 0,2 % du PNB (2006-2015), il est clair qu'il n'y a pas un problème dans ce cas. Les commentaires de nos principaux utilisateurs en font d'ailleurs foi.

La principale cause des révisions du PNB semble être la qualité des données de base :

- Pour le troisième mois du dernier trimestre, nous n'avons à peu près pas de données.
- Pour le deuxième mois, nous avons des données très préliminaires seulement pour certaines composantes du PNB.
- Pour certaines industries, nous n'avons même pas de données mensuelles.

Ainsi, pour améliorer le PNB et réduire les révisions, nous avons besoin de meilleures estimations préliminaires et premières estimations!

Que pouvons-nous faire au point de vue méthodologique en ce qui a trait à la qualité du PNB? Tout d'abord, ce que nous pourrions et nous devrions faire, à mon avis, c'est d'examiner des méthodes bayésiennes pour évaluer la qualité du PNB. Par le passé, diverses études ont porté sur des méthodes d'intégration dans plusieurs pays, par exemple, au Royaume-Uni, en Italie, aux Pays-Bas et peut-être aussi dans d'autres pays. Elles ont mené à une mise en œuvre réussie de méthodes d'étalonnage pour le PNB, par exemple, étalonner les données trimestrielles en fonction des données annuelles, les données des composantes en fonction des totaux, etc. Nous pouvons toutefois élaborer à ce sujet et, en commençant avec un modèle d'intégration à petite échelle, utiliser des méthodes bayésiennes pour simuler le processus d'intégration. Il existe des exemples à très petite échelle de cela, par exemple, dans Van Tongeren,

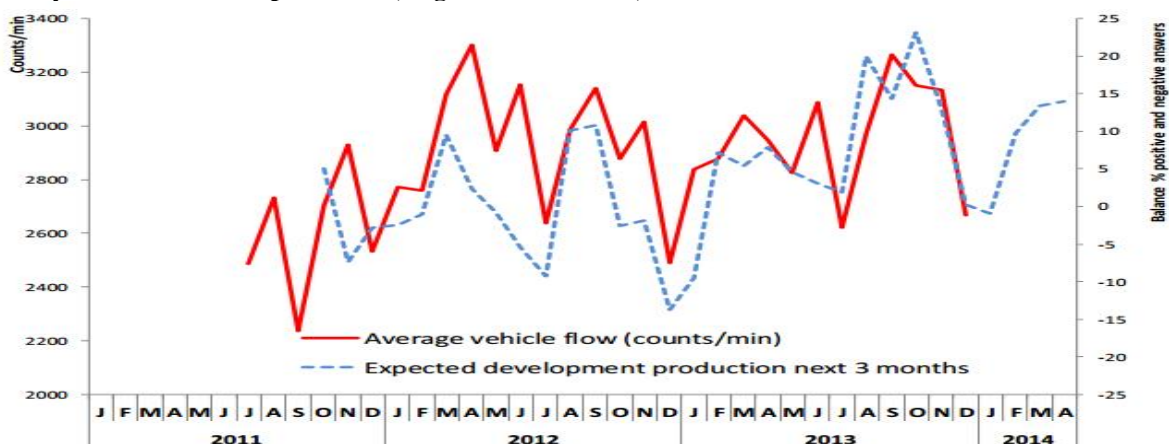
Magnus et De Vos (2001). À une deuxième étape, nous pouvons étendre le modèle à des niveaux inférieurs et à un plus grand nombre de composantes et de détails. Un tel modèle de processus d'intégration nous donnera un meilleur aperçu de la qualité du PNB⁵.

Toutefois aussi, l'utilisation de mégadonnées peut contribuer à l'amélioration de la qualité des estimations du PNB. Il existe quatre sources possibles de mégadonnées :

- *Bases de données fiscales* : Celles-ci servent principalement aux impôts sur les salaires et aux taxes sur les ventes. Elles ne sont toutefois pas suffisamment à jour. Elles sont en retard d'environ deux mois : un mois pour la déclaration des impôts et taxes et un mois pour l'entrée des données dans les bases.
- *Comptes des compagnies* : Il sera possible, au cours des prochaines années, d'accéder directement aux comptes des compagnies. L'accès direct signifie que les modules de déclaration pour l'imposition et les statistiques auront été intégrés dans le logiciel comptable. Toutefois, cela n'est pas encore possible et, même lorsque cela sera mis en œuvre, cela s'appliquera uniquement aux comptes annuels.
- *Transactions bancaires* des entreprises et des ménages : il s'agit clairement de l'avenue la plus prometteuse. Il existe une chambre de compensation seulement pour les banques des Pays-Bas, et elle est très rapide; par ailleurs, les banques doivent déclarer des données quotidiennement, hebdomadairement et mensuellement à la banque centrale. Toutefois, il existe évidemment de très grandes préoccupations en matière de protection de la vie privée dans ce cas. Il sera donc très très difficile d'entreprendre les premières étapes, et il faudra beaucoup de temps, et plus de temps encore avant que la mise en œuvre se fasse.
- *Estimations fondées sur un modèle* à partir de mégadonnées; il ne s'agit pas d'une source de mégadonnées proprement dite, évidemment, mais d'une façon d'utiliser les mégadonnées. C'est donc dire que pour le moment, il s'agit de la meilleure option.

Van Ruth (2015) donne un exemple de ce que les estimations fondées sur un modèle à partir de mégadonnées ont à offrir. La figure 4-2 montre la relation entre l'intensité de la circulation et la production (valeur ajoutée) dans une région importante des Pays-Bas, autour de la ville d'Eindhoven. L'intensité de la circulation est mesurée à partir de capteurs routiers sur la surface des routes, et la production est tirée de l'enquête mensuelle auprès des entreprises. Statistiquement, les séries semblent coïncider, et la désaisonnalisation et une décomposition tendance-cycle peuvent peut-être supprimer certains bruits et améliorer davantage le modèle. À l'heure actuelle, la mesure de l'intensité de la circulation comporte un délai très court, peut-être un ou deux jours. C'est donc dire que cette mesure, ainsi que des relations similaires, peuvent être utiles pour élaborer de meilleures premières estimations et des estimations préliminaires améliorées de la production.

Figure 4-2
Flux moyen de véhicules et production, région d'Eindhoven, 2011-2014



⁵ Ces méthodes bayésiennes sont des exemples d'estimations fondées sur un modèle. Parmi les autres exemples figurent les modèles de séries chronologiques structurelles et les techniques d'intégration, par exemple, pour les ensembles de microdonnées. J'y reviendrai plus tard.

5. Stratégie méthodologique pour l'utilisation des mégadonnées dans les statistiques officielles

5.1. Problème de représentativité des mégadonnées

Comme nous le savons tous maintenant, il est difficile de transformer les mégadonnées en statistiques et en information, leur représentativité posant un problème. Le problème de représentativité des mégadonnées est l'un des obstacles méthodologiques les plus importants à leur utilisation dans les statistiques officielles! Elles sont difficiles à coupler et à apparier avec d'autres ensembles de données qui comprennent de l'information contextuelle. Par exemple, avec les capteurs de circulation, vous pouvez observer une voiture qui passe, mais vous ne savez pas qui s'y trouve, à qui elle appartient et pourquoi cette personne passe à cet endroit. En outre, les mégadonnées sont sélectives. Par exemple, sur Twitter, les hommes de 30 ans environ sont surreprésentés.

Précisément parce que le couplage avec des variables contextuelles est souvent impossible, les méthodes habituelles des techniques d'enquête ne peuvent pas être utilisées directement. Toutefois, nous pouvons utiliser des méthodes pour les échantillons non probabilistes. Vous pouvez penser ici en termes de propension à être observé dans un grand ensemble de données, et comme dans une enquête, nous pouvons analyser la propension à répondre. Cela s'apparente beaucoup à ce que l'on appelle le profilage dans le marketing. Par exemple, nous pouvons estimer la probabilité qu'une personne dans un ensemble de données de Twitter soit un homme ou une femme. Ainsi, vous pouvez être en mesure de corriger une répartition entre les sexes non équilibrée. De façon plus générale, vous tentez de modéliser les rapports entre les variables de l'ensemble de mégadonnées et, si possible, d'autres ensembles de données.

Essentiellement, il existe trois catégories de méthodes pour modéliser les ensembles de mégadonnées dans les statistiques officielles :

1. pseudo-méthodes fondées sur le plan de sondage;
2. techniques d'apprentissage automatique;
3. méthodes d'intégration dans lesquelles, en termes simples, les mégadonnées servent comme variables auxiliaires.

5.2. Pseudo-méthodes fondées sur le plan de sondage

Dans ce cas, nous agissons comme si, dans une certaine mesure, un mécanisme probabiliste avait généré les données. Nous pouvons par la suite procéder de la façon habituelle, par exemple, utiliser les caractéristiques des événements dans les données pour constituer des strates, puis utiliser un estimateur de stratification a posteriori pour tenir compte de la sélectivité, mais cela n'est pas toujours sensé et est difficile à vérifier.

5.3. Techniques d'apprentissage automatique

Celles-ci sont beaucoup plus récentes, mais beaucoup d'ouvrages ont déjà été publiés à ce sujet. En termes simples, avec l'aide de ces techniques, vous laissez les données décider quelles sont les strates. Plusieurs techniques bien connues comprennent les suivantes :

- arbres de régression;
- K plus proches voisins;
- réseaux de neurones artificiels;
- machines à vecteurs de support.

Ces techniques sont aussi très utiles pour l'utilisation des mégadonnées à des fins analytiques, que nous avons brièvement mentionnée précédemment.

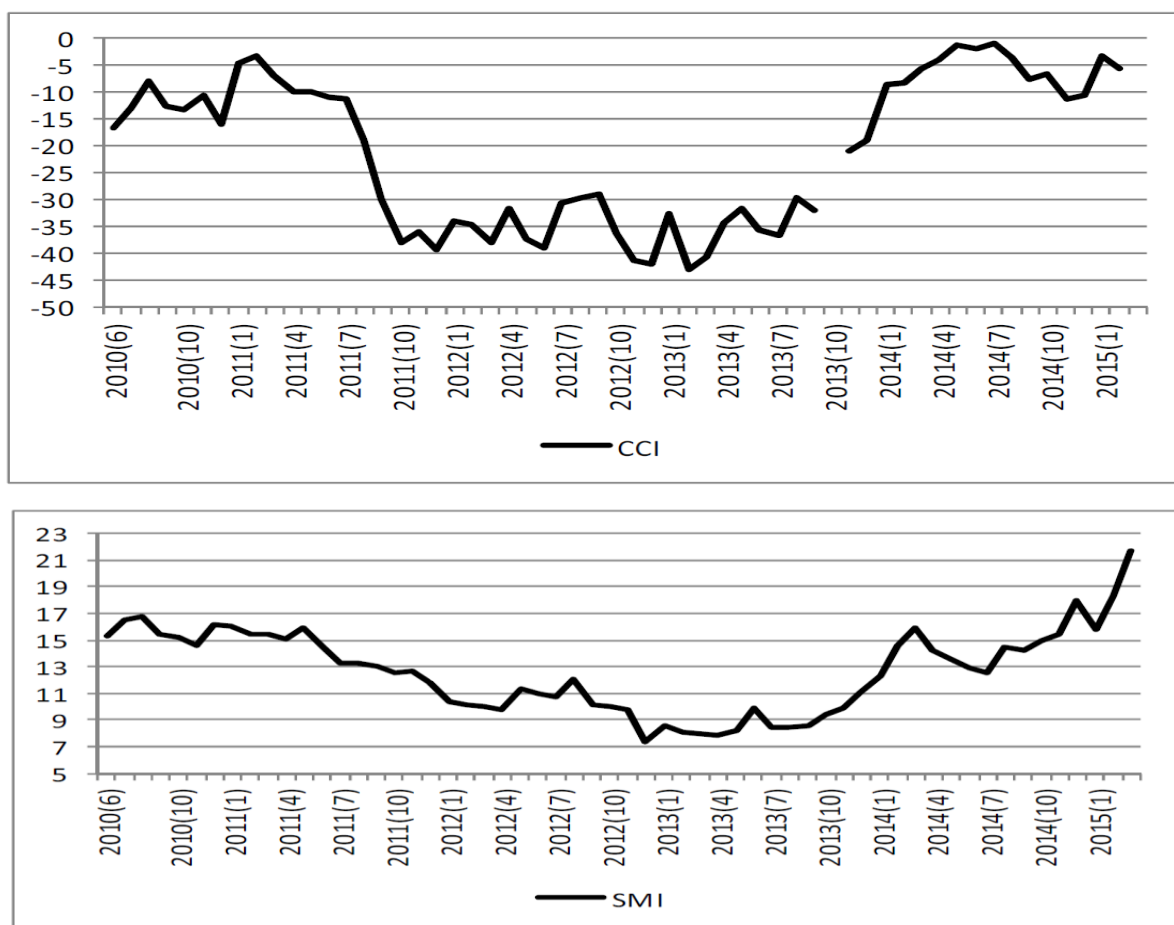
Grâce à ces techniques d'apprentissage automatique, nous avons procédé à une étude de simulation relativement à la base de données des véhicules automobiles (Van den Brakel, Buelens and Burger, 2015). Il s'agit d'une très grande base de données, comportant plus de 7 millions d'enregistrements, soit tous les véhicules automobiles des Pays-Bas. Les résultats de la simulation montrent que les techniques d'apprentissage automatique sont prometteuses. De façon plus particulière, ce sont les arbres de régression qui sont prometteurs, parce qu'ils constituent dans une certaine

mesure un intermédiaire entre les techniques fondées sur le plan de sondage et d'autres techniques d'apprentissage automatique plus fluides. Ils facilitent aussi davantage la compréhension de ce qui se produit « dans les coulisses », et sont par conséquent plus acceptables pour les INS, qui doivent toujours être transparents au sujet de leur méthodologie.

5.4. Techniques d'intégration

Dans ce cas, nous tentons d'améliorer l'estimation habituelle découlant d'enquêtes ou de registres à l'aide des mégadonnées. Par exemple, nous avons examiné le rapport entre l'indice de confiance des consommateurs et un indice du sentiment dans les médias sociaux (Van den Brakel et coll., 2016). L'indice de confiance des consommateurs provient d'une enquête mensuelle comportant environ 1 000 répondants. L'indice du sentiment dans les médias sociaux est fondé sur des mégadonnées de Twitter et de FaceBook.

Figure 5.4-1
Indice de confiance des consommateurs (CCI) et indice du sentiment dans les médias sociaux (SMI), 2010-2015



La partie supérieure de la figure 5.4-1 montre l'indice de confiance des consommateurs et la partie inférieure, l'indice du sentiment dans les médias sociaux. À noter que les échelles verticales sont différentes. La période va de 2009 à 2014 et les données sont des données mensuelles. Nous voyons que les deux séries évoluent plus ou moins en parallèle. Cela est confirmé par une analyse statistique officielle comportant un modèle de séries chronologiques structurales, qui montre que les deux séries sont cointégrées et qu'il est possible d'améliorer l'indice de confiance des consommateurs à l'aide du modèle. Les avantages sont les suivants :

- plus grande précision;
- plus grande fréquence;
- sous-populations plus détaillées.

5.5. Modèles utilisés pour les statistiques officielles

Dans la présente section et dans la précédente, on a démontré que l'analyse fondée sur un modèle prendra beaucoup d'importance, plus particulièrement lorsque les mégadonnées seront utilisées davantage dans les statistiques officielles. En outre, le présent document a aussi permis de démontrer, entre les lignes, que les avantages des mégadonnées peuvent être suffisamment importants pour que les INS ne puissent plus se permettre de les laisser de côté.

Les INS ont toujours été prudents, à juste titre, en ce qui a trait à l'utilisation de modèles dans les statistiques officielles. Lorsqu'il s'agit d'utiliser les mégadonnées, nous devrions toujours essayer de demeurer dans le cadre des statistiques officielles. Par ailleurs, dans plusieurs domaines, nous utilisons déjà des modèles, comme dans la désaisonnalisation et dans la correction pour tenir compte de la non-réponse⁶. Afin de demeurer dans le cadre des statistiques officielles, nous devrions par conséquent non seulement refuser d'utiliser des modèles, mais chercher des méthodes qui sont acceptables. Des lignes directrices et des restrictions quant à l'utilisation des modèles dans les statistiques officielles ont été élaborées par Buelens, De Wolff et Zeelenberg (2015), par exemple :

- aucune prévision;
- prévision immédiate uniquement en présence de données pertinentes (mais pas sur la base de modèles comportementaux);
- vérification exhaustive de modèle;
- transparence au sujet de la méthodologie.

5.6. Processus de production génériques pour les mégadonnées

Nous devons aussi tenter de créer des processus de production plus génériques pour les mégadonnées. Cette question n'a pas été abordée dans le présent document, parce qu'il s'agit davantage d'une tâche liée au développement qu'à la recherche. Toutefois, elle est importante. Si nous n'avons pas de tels processus de production génériques, il sera toujours très coûteux d'utiliser les mégadonnées. Encore une fois, la solution du problème de représentativité constitue une étape importante vers des processus de production plus génériques.

6. Résumé

Une stratégie méthodologique pour les mégadonnées dans les statistiques officielles pourrait prendre la forme suivante :

Accent sur des *estimations fondées sur un modèle*. Cela est utile pour

- améliorer la qualité;
- analyser les relations complexes;
- résoudre le problème de représentativité des mégadonnées.

Bibliographie

van den Brakel, J., et Bethlehem, J. (2008) "Model-Based Estimation for Official Statistics," Document de travail 2008-02, Statistique Pays-Bas. <https://www.cbs.nl/nl-nl/achtergrond/2008/10/model-based-estimation-for-official-statistics>

van den Brakel, J., E. Söhler, P. Daas, et B. Buelens (2016), "Social media as a data source for official statistics; the Dutch Consumer Confidence Index," Document de travail 2016-01, Statistique Pays-Bas. <https://www.cbs.nl/nl-nl/achtergrond/2016/07/social-media-as-a-data-source-for-official-statistics-the-dutch-consumer-confidence-index>

⁶ D'autres exemples figurent dans Bethlehem et Van den Brakel (2008) et Buelens, De Wolff et Zeelenberg (2015).

Braaksma, B., et K. Zeelenberg (2015), “ ‘Re-make/Re-Model’: Should Big Data Change the Modelling Paradigm in Official Statistics?” *Statistical Journal of the IAOS* 31 (2), pp. 193–202. doi: [10.3233/sji-150892](https://doi.org/10.3233/sji-150892).

Buelens, B., P.-P. de Wolff, et K. Zeelenberg (2015), “Model based estimation at Statistics Netherlands,” article présenté au Advisory Council on Methodology and Quality, 2 avril 2015, Statistique Pays-Bas.

Elbourne, A., K. Grabska, H. Kranendonk, et J. Rhuggenaath (2015), “The effects of CBS revisions on CPB forecasts,” CPB Background Document, CPB Netherlands Bureau for Economic Policy Analysis. <http://www.cpb.nl/en/publication/the-effects-of-cbs-revisions-on-cpb-forecasts>

van Ruth, F. (2014), “Traffic intensity as indicator of regional economic activity,” Document de travail 2014-21, Statistique Pays-Bas. <https://www.cbs.nl/nl-nl/achtergrond/2014/34/traffic-intensity-as-indicator-of-regional-economic-activity>

Smekens, M. , et K. Zeelenberg (2015), “Lean Six Sigma at Statistics Netherlands,” *Statistical Journal of the IAOS* 31 (4): 583–86. doi: [10.3233/SJI-150930](https://doi.org/10.3233/SJI-150930).

Vincent, P. (2015), *Herman Gorter - Poems of 1890: A Selection*. UCL Press, Londres 2015. Accès libre à <http://www.ucl.ac.uk/ucl-press/browse-books/poems-1890/index>

Zeelenberg, K. (2015), “Product and service innovation at Statistics Netherlands,” Article présenté à la 63e séance plénière de la Conférence des statisticiens européens, Genève, 16 juin 2014. http://www.unece.org/index.php?id=38920#jfmulticontent_c48917-3

Zwijnenburg, J. (2015), “Revisions of quarterly GDP in selected OECD countries,” *Cahier statistique 22 de l’OCDE*. <https://www.oecd.org/std/thestatisticsbrief.htm>