

Article

Symposium 2008 :
Collecte des données : défis, réalisations et nouvelles orientations

Analyse de cohérence – Amélioration de la qualité des comptes nationaux

par Anna-Greta Erikson

2009



Analyse de cohérence – Amélioration de la qualité des comptes nationaux

Anna-Greta Erikson¹

Résumé

Dans de nombreux pays, l'amélioration des statistiques économiques est au nombre des grandes priorités du 21^e siècle. L'accent est mis, d'abord et avant tout, sur la qualité des comptes nationaux, tant annuels que trimestriels. À cet égard, la qualité des données sur les entreprises les plus grandes joue un rôle essentiel. L'analyse de cohérence est un outil fort utile pour s'assurer que les données fournies par ces entreprises sont de bonne qualité. Par cohérence, nous entendons que les données provenant de diverses sources concordent et brossent un tableau logique du développement de ces entreprises. Une analyse de cohérence efficace est généralement une tâche ardue qui consiste principalement à recueillir des données de différentes sources afin de les comparer de façon structurée. Au cours des deux dernières années, de grands progrès ont été accomplis à Statistics Sweden en ce qui concerne l'amélioration des routines servant à l'analyse de cohérence. Nous avons construit un outil TI qui recueille les données sur les plus grandes entreprises auprès d'un grand nombre de sources et les présente de manière structurée et logique, et nous avons élaboré une approche systématique d'analyse trimestrielle des données destinée aux comptes nationaux. Le présent article décrit les travaux effectués dans ces deux domaines et donne un aperçu de l'outil TI et des routines retenues.

Mots clés : Cohérence, qualité supérieure, assurance de la qualité.

1. Introduction

Dans plusieurs pays, des ressources particulières ont été affectées au traitement des plus grandes entreprises ou fournisseurs clés de données statistiques. En 2004, dans le cadre d'une approche systématique en vue d'améliorer la qualité des comptes nationaux, Statistics Sweden a créé la section des gestionnaires des grandes entreprises. La tâche principale de cette section est de gérer les fournisseurs clés de statistiques économiques, c'est-à-dire les entreprises les plus grandes en Suède, afin d'accroître la cohérence et la qualité des données qu'elles fournissent. Le présent article, axé sur l'approche suivie pour améliorer la qualité des données recueillies auprès des plus grandes entreprises, expose les travaux concernant l'analyse de cohérence effectuée dans le cadre du programme de gestion des grandes entreprises à Statistics Sweden.

2. Programme de gestion des grandes entreprises

L'objectif le plus important de Statistics Sweden, lorsqu'il a créé un groupe de gestionnaires des grandes entreprises, était d'améliorer la qualité des statistiques du Système de comptabilité nationale. Ces gestionnaires ont été chargés d'améliorer la gestion des plus grandes entreprises de Suède, car la qualité des données du Système de comptabilité nationale dépend de la qualité et de la cohérence des données recueillies auprès de ces fournisseurs clés.

2.1 Les fournisseurs clés de données

En Suède, les gestionnaires des grandes entreprises se concentrent sur les 50 plus grandes entreprises. Celles-ci sont repérées en examinant les principaux indicateurs économiques, c'est-à-dire le chiffre d'affaire/chiffre net de ventes, la valeur ajoutée, l'effectif, la valeur de la production et le total du bilan. En 2005, les 50 plus grandes entreprises représentaient près de 30 % de la contribution des entreprises non financières au produit intérieur brut. L'objectif est

¹Anna-Greta Erikson, Statistics Sweden, DFO/SF, SE-701 89 ÖREBRO, Sweden (annagreta.erikson@scb.se)

de pouvoir concentrer les efforts sur les fournisseurs de données les plus importants du système de statistique économique. Les gestionnaires surveillent les grandes entreprises en permanence, même quand il n'existe aucun problème dans le processus de collecte des données, parce que, si un problème survient dans la collecte ou la fourniture des données sur ces entreprises, l'effet sur les comptes nationaux risque d'être important.

2.2 Travail des gestionnaires des grandes entreprises

Nous avons choisi de donner une grande portée au travail des gestionnaires des grandes entreprises, afin qu'ils arrivent à bien connaître les entreprises dont ils ont la responsabilité et qu'ils acquièrent une perspective globale des entreprises. En ce qui concerne l'analyse de cohérence, les gestionnaires sont déjà familiarisés avec les données recueillies et avec les problèmes que pose le processus de collecte des données. Ils apprennent aussi à reconnaître les tendances et les phénomènes similaires parmi les entreprises qu'ils gèrent.

2.2.1 Établir des liens

L'une des tâches les plus importantes des gestionnaires est de nouer des liens avec les entreprises les plus grandes et, plus précisément, d'établir un dialogue avec des employés travaillant dans divers domaines, comme les ressources humaines et la comptabilité. Dans ce processus, il est important que les gestionnaires des grandes entreprises rendent visite aux entreprises afin de les informer du travail qu'ils font et de l'importance qu'ont les données de l'entreprise pour la production de statistiques. Cette démarche a l'avantage supplémentaire de réduire le fardeau perçu de la fourniture des données, en permettant aux entreprises de mieux comprendre le processus et de mettre un « visage » sur les demandes de données.

2.2.2 Point de contact unique

Le gestionnaire des grandes entreprises est un point de contact unique entre le bureau de la statistique et l'entreprise. À ce titre, il est à la fois un intermédiaire et la personne qui fait part aux entreprises des besoins et des demandes de statistiques, et qui leur transmet l'information concernant les changements effectués au bureau de la statistique. Il joue aussi un rôle essentiel dans le dépistage des problèmes importants qui pourraient survenir.

2.2.3 Apprendre à connaître les entreprises

Il est essentiel que le gestionnaire apprenne vraiment à connaître les entreprises et le groupe auquel elles appartiennent. Cela signifie qu'il doit se familiariser avec le contexte juridique et opérationnel de l'organisation et être mis continuellement au courant des changements apportés à cette dernière. Il est également important qu'il connaisse la capacité de fourniture de données de l'entreprise et d'autres caractéristiques uniques, telles que celles des systèmes de comptabilité et autres systèmes administratifs.

2.2.4 Mettre à jour le registre des entreprises et dresser le profil des unités statistiques

Une autre tâche des gestionnaires des grandes entreprises consiste à veiller en permanence à la mise à jour du registre des entreprises. En outre, grâce à la surveillance des unités légales et des unités statistiques, ils peuvent décider quelles sont les unités pertinentes et éventuellement déclarantes (entreprise, unité d'activité économique, etc.). Comme les gestionnaires ne travaillent qu'avec quelques groupes d'entreprises, ils peuvent utiliser toutes les sources disponibles d'information pour mettre à jour le registre des entreprises et pour établir le profil des unités statistiques; ils n'ont pas besoin de s'appuyer uniquement sur les sources générales du registre.

2.2.5 Recueillir et vérifier les données

Les gestionnaires des grandes entreprises manipulent quotidiennement les données concernant les entreprises dont ils ont la charge. Dans le cas de certaines enquêtes annuelles, cela signifie entretenir des contacts réguliers avec les fournisseurs de données. Pour les statistiques à court terme et d'autres enquêtes annuelles, les gestionnaires jouent le rôle de conseillers et de ressources en vue de faciliter la résolution des problèmes causés par la production de

statistiques. Ce travail pratique durant le processus de collecte des données leur permet de mieux comprendre le système d'enquêtes. La tâche comporte la surveillance continue du processus permanent de collecte des données et l'utilisation de cette surveillance comme outil de dépistage des problèmes éventuels en ce qui concerne la fourniture de données et le processus de collecte des données. En cas de problème, les gestionnaires peuvent décider d'adopter des solutions personnalisées de fourniture des données pour des entreprises particulières.

2.2.6 Cohérence des données

Le principal objectif de la création d'un groupe de gestionnaires des grandes entreprises à Statistics Sweden est l'amélioration de la qualité des statistiques du Système de comptabilité nationale. Les gestionnaires ont pour mandat de veiller à la qualité et à la cohérence des données fournies par les plus grandes entreprises.

Grâce à l'analyse de cohérence, nous supposons que le besoin de vérification des données diminuera avec le temps. Le fait de travailler en étroite collaboration avec les entreprises et d'établir des solutions propres à chacune lorsque cela s'avère nécessaire signifiera que les entreprises produiront des données de haute qualité durant le processus normal de fourniture des données et que le besoin de vérification après la collecte diminuera.

3. Analyse de cohérence

En Suède, les gestionnaires des grandes entreprises ont exécuté progressivement le travail pratique. L'analyse de cohérence, l'une de leurs tâches les plus importantes, figure au programme depuis le début de 2004. Pour commencer, les travaux se sont concentrés sur la cohérence des données pour la production des statistiques annuelles, l'analyse étant intégrée dans le processus de vérification ordinaire des données de trois enquêtes annuelles. Ces enquêtes sont les principales sources de données pour la production des comptes nationaux et constituent le fondement du système d'analyse de cohérence.

3.1 Approche systématique – Extension de l'analyse de cohérence

Dès le début, il a été prévu d'élargir la portée de l'analyse de cohérence et d'y inclure des enquêtes et des champs de données supplémentaires. Il était également important d'intégrer les statistiques à court terme dans cette analyse. En 2005, nous avons adopté une approche systématique s'appuyant sur la notion que les enquêtes et les registres font tous partie d'un même système : le Système de statistiques économiques. Très complexe, ce dernier est constitué des enquêtes et des registres utilisés directement dans les comptes nationaux ou reliés à des enquêtes ou à des registres utilisés dans les comptes nationaux. Il permet de cerner divers types de liens entre les enquêtes et les registres.

3.2 Identification des liens

En ce qui a trait aux variables, les travaux ont débuté par l'identification des liens entre les enquêtes et les registres dans le système. Afin d'intégrer les statistiques annuelles ainsi qu'à court terme, de même qu'un champ de données plus étendu dans l'analyse de cohérence, il a également fallu structurer les grands ensembles de données d'une certaine façon. Nous avons commencé par regrouper les variables en divers champs de données, en nous appuyant sur la notion que les indicateurs sont liés les uns aux autres. Nous avons créé sept domaines de cohérence appropriés pour les besoins de notre analyse.

3.3 Domaines de cohérence et liens

Les domaines de cohérence cernés sont la production, le commerce extérieur, la consommation intermédiaire, les salaires et traitements, l'emploi, l'investissement et les stocks, etc. Nous avons également défini plusieurs liens dans certains domaines de cohérence. Le but de la décomposition des domaines de cohérence en diverses combinaisons de variables est de pouvoir donner un aperçu plus clair et une analyse plus efficace d'un grand ensemble de variables. Les liens entre les enquêtes peuvent être directs, ce qui signifie que ces enquêtes contiennent des

indicateurs identiques ou quasi identiques, ou indirects, ce qui signifie que les indicateurs contenus dans les deux enquêtes ne sont pas identiques, mais sont néanmoins liés, en quelque sorte, les uns aux autres.

4. L’outil d’analyse de cohérence

L’approche structurée a abouti à la création d’un outil personnalisé pour l’analyse de cohérence, principalement parce que l’entreposage des données est décentralisé à Statistics Sweden. Autrement dit, les données concernant les statistiques des entreprises étaient, et sont encore, situées sur divers serveurs. Le stockage des données n’est pas normalisé, c’est-à-dire que les structures des différentes bases de données ne sont pas harmonisées. De toute façon, même si l’entreposage des données avait été centralisé, il aurait fallu structurer les données et produire des agrégats et des périodes de référence comparables.

4.1 Fonctionnalité

L’outil personnalisé a été construit en vue d’obtenir les données requises à partir de divers bases de données et serveurs, et de recueillir des données à jour en provenance de divers types d’enquêtes et de registres.

4.1.1 Périodicité

L’outil personnalisé permet de résoudre le problème des diverses périodicités des données, c’est-à-dire qu’il agrège les données à court terme en données annuelles, ainsi que les données mensuelles en données trimestrielles. Ce traitement des diverses périodicités des données facilite la comparaison de ces dernières.

4.1.2 Agrégation

L’outil agrège aussi les données recueillies pour diverses unités statistiques, en vue de les rendre comparables. Les données recueillies au niveau de l’unité locale sont agrégées au niveau de l’unité d’activité économique, et les données recueillies au niveau de l’unité d’activité économique sont agrégées au niveau de l’entreprise. Au niveau de l’entreprise, les données sont agrégées au niveau de l’unité de cohérence, qui est le type unique d’unité utilisé pour l’analyse de cohérence.

4.1.3 Transformation

L’outil personnalisé transforme aussi les données recueillies en variables mieux adaptées à l’analyse de cohérence. Autrement dit, il agrège les variables pour lesquelles les données sont recueillies à un niveau plus pointu de détail, de manière à ce qu’elles soient davantage comparables à d’autres variables. Les moyennes trimestrielles sont calculées d’après les données mensuelles, et les moyennes annuelles, d’après les données trimestrielles. Le but est de pouvoir comparer des données qui ne peuvent pas être simplement agrégées.

4.1.4 Présentation

L’outil présente aussi les données sous-jacentes sur demande. En étendant la présentation, il est possible de voir les données d’origine tirées des bases de données originales pour chaque enquête et registre figurant dans la présentation. Il est donc facile d’examiner les données avant la transformation et l’agrégation. Il est également facile de copier les données dans un fichier Excel, où leur traitement peut être poursuivi.

4.2 Procédures

Pour lancer une analyse de cohérence dans l’outil personnalisé, il faut d’abord établir la valeur de certains paramètres fondamentaux.

4.2.1 Choix de l'unité statistique

L'analyse doit débiter par le choix de l'unité statistique la plus appropriée. Le nombre d'enquêtes ou de registres à partir desquels il est possible d'obtenir les données varie selon le type d'unité pour laquelle il est choisi de recueillir des données. Il existe cinq unités distinctes pour lesquelles des données sont disponibles dans le système :

- unité de cohérence;
- entreprise;
- unité légale;
- unité d'activité économique;
- unité locale.

L'unité utilisée le plus fréquemment pour recueillir des données est l'entreprise. L'unité de cohérence est une unité créée uniquement pour les besoins de l'analyse des séries chronologiques dans l'analyse de cohérence. Elle permet de tenir compte des effets des changements organisationnels au cours du temps, en créant une unité qui est davantage comparable dans le temps. Une unité de cohérence est composée d'une ou de plusieurs entreprises.

Si l'on désire recueillir des données à un niveau plus agrégé (par exemple l'unité de cohérence), il est possible d'accéder aux données concernant les unités sous-jacentes, c'est-à-dire les entreprises, les unités légales, les unités d'activités économiques ou les unités locales des entreprises. Si l'on désire recueillir des données à partir de l'unité d'entreprise, il est également possible d'accéder et d'analyser les données sur les unités légales, les unités d'activités économiques ou les unités locales des entreprises.

4.2.2 Choix de la période

L'étape suivante consiste à préciser la période sur laquelle doit porter l'analyse. Il faut choisir l'année pour laquelle les données doivent être recueillies. Il faut aussi indiquer si l'analyse des données doit être effectuée sur une base annuelle, trimestrielle ou mensuelle. S'il est choisi d'analyser des données trimestrielles ou mensuelles, les données peuvent être recueillies pour un seul trimestre ou un seul mois, ou pour tous les trimestres ou tous les mois de l'année.

L'outil permet également de recueillir des données sous forme de séries chronologiques. Des données sont disponibles pour 2003 et les années suivantes.

4.2.3 Choix du domaine des données

La troisième étape consiste à décider du domaine de données à analyser. Huit domaines de cohérence sont définis dans l'outil d'analyse de cohérence, à savoir :

- production;
- production et commerce extérieur;
- consommation intermédiaire;
- consommation intermédiaire et commerce extérieur;
- emploi;
- traitements et salaires;
- investissements;
- stocks, etc.

Nous avons choisi de combiner dans l'outil un domaine de cohérence original, le commerce extérieur, avec les domaines de la production et de la consommation intermédiaire. La raison de la collecte de données concernant à la fois la production et les exportations est que l'on obtient ainsi un tableau plus complet des flux au sein de l'unité. Grâce à la combinaison aux données sur la production, il est possible de voir les flux de réexportation dans les données sur les exportations. Il en est de même des données sur la consommation intermédiaire recueillies en combinaison avec les données sur le commerce extérieur.

Pour un domaine de cohérence particulier, nous avons défini un ou plusieurs liens. Par exemple, le domaine de cohérence concernant les données sur la production comporte quatre liens définis. Le premier lien permet de regrouper des données concernant les ventes nettes et les livraisons de biens déclarées qui figurent dans diverses sources. Le deuxième permet de réunir des données concernant la valeur de la production de biens et les ventes nettes générées par les activités de fabrication qui sont déclarées dans diverses sources. Le troisième lien permet de réunir des données sur les ventes nettes générées par les activités d'échange recueillies dans diverses sources. Enfin, le quatrième lien permet de réunir des données sur les ventes nettes générées par d'autres activités qui sont déclarées dans diverses sources.

Dans le domaine de cohérence concernant les données sur la production et les données sur le commerce extérieur, il existe à la fois des liens directs et indirects. Le premier lien, les exportations totales, rassemble des données concernant les exportations totales de biens et de services provenant de diverses sources. Le deuxième lien, les exportations totales et la production, réunit des données concernant la production, ainsi que les exportations et les importations. Le troisième lien, la production de biens et les exportations de biens, réunit des données concernant la production de biens et les exportations ainsi que les exportations de biens, groupées selon le niveau souhaité de code de bien. Le quatrième lien, la production de services et les échanges extérieurs de services, permet de recueillir des données concernant la production de services ainsi que les exportations et les importations de services, selon des groupements prédéterminés de services.

4.3 Mise en œuvre

Quand l'outil a commencé à être utilisé pour le travail courant des gestionnaires des grandes entreprises, son inutilité est apparue évidente, non seulement pour déceler tôt le manque de cohérence dans le système, mais aussi afin de superviser le processus de collecte des données. Les gestionnaires ont la possibilité de superviser le processus de collecte des statistiques à court terme ainsi que des statistiques annuelles, c'est-à-dire de voir si les changements dus aux restructurations et aux réorganisations se manifestent comme prévu.

4.3.1 Analyse de cohérence annuelle

L'outil personnalisé permet d'intégrer des enquêtes supplémentaires dans l'analyse de cohérence ordinaire effectuée annuellement; l'accès aux données concernant les statistiques à court terme et les statistiques annuelles est désormais facile. Comme l'outil fournit aussi aux gestionnaires un champ de données plus général, la mise en oeuvre des routines permettant d'intégrer ces aspects dans l'analyse régulière a débuté, mais n'est pas entièrement achevée.

4.3.2 Analyse de cohérence trimestrielle

Parallèlement à la mise en oeuvre de nouvelles routines pour l'analyse de cohérence annuelle, des routines pour l'analyse de cohérence sur une base trimestrielle sont également mises en oeuvre graduellement dans le système de gestion des grandes entreprises. Ces travaux de développement sont effectués en collaboration avec le personnel des comptes nationaux. Grâce à cette collaboration, des rapports de cohérence trimestriels sont produits pour un sous-ensemble d'entreprises auxquelles est accordée une haute priorité. Les entreprises prioritaires sont principalement celles qui dominent entièrement un certain produit ou un groupe de la NACE dans les comptes nationaux, ou des entreprises exploitées dans des domaines où il existe souvent des problèmes d'incohérence.

Les rapports de cohérence trimestriels ont deux objectifs : le premier est de prendre conscience des incohérences à un stade précoce et de pouvoir y remédier temporairement dans les comptes nationaux; et le deuxième consiste à pouvoir entreprendre une enquête plus approfondie. Sur une base trimestrielle, on ne dispose habituellement pas de suffisamment de temps pour résoudre les problèmes d'incohérence concernant plusieurs enquêtes statistiques, mais si le personnel des comptes nationaux est au courant du manque de cohérence dans certains domaines, il peut en tenir compte au niveau macroéconomique. Entre-temps, les gestionnaires des grandes entreprises lancent d'autres analyses et communiquent avec les entreprises concernées et les divers spécialistes du domaine. Idéalement, quand

l'analyse de cohérence du trimestre suivant se termine, les données concernant l'entreprise en question sont cohérentes.

4.3.3 Rapports de cohérence trimestriels

Dans les rapports d'analyse de cohérence trimestriels, les données sont présentées de façon normalisée. Le rapport est constitué de quatre grandes fiches de données présentant des données chronologiques pour divers domaines de cohérence : la production, le commerce extérieur, les traitements et salaires, ainsi que l'emploi et les heures de travail. Les séries chronologiques contiennent des données pour l'analyse trimestrielle ainsi qu'annuelle. En plus des séries chronologiques, les rapports comprennent des séries indices, où les données sont comparées à celles de la même période l'année précédente. Les séries indices sont présentées pour un sous-ensemble de variables. De plus, certains ratios sont fournis, tels les salaires et traitements par nombre d'heures travaillées, les salaires et traitements par employé et une approximation de l'évolution de la productivité.

5. L'avenir

Bien que nous n'ayons pas encore mis en œuvre complètement le nouvel outil personnalisé pour le travail courant des gestionnaires des grandes entreprises, nous avons cerné certains domaines dans lesquels les travaux de développement devraient se poursuivre. Nous aimerions intégrer la signalisation automatique des incohérences soupçonnées qui ont un effet important sur les estimations statistiques en général et sur les agrégats des comptes nationaux en particulier. Le but serait de rendre le processus plus efficace et de raccourcir la durée de l'analyse. Idéalement, les gestionnaires aimeraient pouvoir « prédire » les questions qui seront posées par le personnel des comptes nationaux et procéder immédiatement à des analyses plus approfondies et à des contacts supplémentaires avec les entreprises.

Un autre développement pourrait être l'ajout d'une fonction qui permet de simuler des valeurs vérifiées et corrigées dans certains domaines, et de voir les effets qu'elles ont sur la cohérence dans un domaine connexe. Le but ne serait pas de fabriquer la cohérence, mais, de nouveau, de simplifier le processus d'analyse de cohérence, et de pouvoir intervenir plus efficacement et contacter les entreprises « une fois pour toutes ».

L'analyse de cohérence est un domaine de travail compliqué, surtout dans le cas des grandes entreprises complexes. L'automatisation du travail dans ce domaine ne peut pas remplacer la connaissance et la compréhension des unités particulières analysées. Cependant, elle permet d'assurer un niveau commun d'analyse et réduire la dépendance à l'égard d'une personne particulière.

Bibliographie

Erikson, A. (2007). Large Enterprise Management - Higher Quality and Lower Burden through Better Relations, *Proceedings of the International Conference on Establishment Surveys (ICES-III)*, American Statistical Association, 825-829.

Erikson, J. (2002). Coherence Analysis as a Tool for Questionnaire Evaluation in Enterprise Statistics, *International Conference on Questionnaire Design Evaluation and Testing (QDET)*