

CONCEPTS, DÉFINITIONS ET QUALITÉ DES DONNÉES

L'Enquête mensuelle sur les industries manufacturières (EMIM) contient des séries de données statistiques sur les ventes de biens manufacturés, les stocks, les commandes en carnet et les nouvelles commandes des fabricants. Chacune des valeurs associées à ces composantes représente une projection mensuelle des données de l'Enquête annuelle sur les manufactures et l'exploitation forestière (EAMEF).

L'EMIM est une enquête par échantillonnage menée auprès d'environ 10 500 établissements manufacturiers canadiens, lesquels sont catégorisés en plus de 220 industries. Les industries sont classées selon le Système de classification des industries de l'Amérique du Nord (SCIAN) de 2007. Des séries désaisonnalisées sont disponibles pour les principaux agrégats.

Un établissement comporte la plus petite unité de fabrication en mesure d'informer sur les variables à l'étude. Les données recueillies par l'EMIM tracent le « portrait » de la valeur des ventes de biens fabriqués, réalisées par le secteur manufacturier canadien et nous permettent d'analyser la situation de l'économie canadienne, ainsi que la santé d'industries spécifiques à court et à moyen termes. Les données de l'enquête sont employées par des intervenants des secteurs privé et public, notamment par Statistique Canada, les gouvernements fédéraux et provinciaux, des entités commerciales et industrielles, des organismes non gouvernementaux nationaux et internationaux, des experts-conseils, la presse d'affaires et de simples citoyens. Ces données permettent d'analyser la part du marché, les tendances, l'étalonnage des entreprises, l'analyse des politiques, l'élaboration des programmes, la politique fiscale et la politique commerciale.

1. Ventes de biens fabriqués

Les ventes de biens fabriqués (anciennement les livraisons des produits de propre fabrication) sont définies comme étant la valeur des produits fabriqués par les établissements qui ont été livrés à des clients. Elles n'incluent aucune activité de commerce de gros et aucune recette provenant de la location de matériel ou de la vente d'électricité. Dans la pratique, certains répondants déclarent les transactions financières plutôt que les paiements ayant trait au travail fait. Les ventes de biens fabriqués, pour le Canada et les provinces, sont disponibles selon le SCIAN, au niveau de détails des trois chiffres.

Dans le cas des industries des produits aérospatiaux et leurs pièces d'aéronef et de la construction navale, on se fonde sur la valeur de la production plutôt que sur celle des ventes de biens fabriqués. Pour déterminer cette valeur, on ajuste la valeur mensuelle des ventes de biens fabriqués en fonction de la variation mensuelle des stocks détenus pour les stocks de biens/travaux en cours de fabrication et les produits finis fabriqués). Ce calcul ne tient pas compte des matières premières, car les chiffres de production visent à mesurer le « travail fait » durant le mois. On cherche ainsi à réduire la distorsion causée par les ventes de biens fabriqués de produits d'une grande valeur comme lors d'une vente complétée.

2. Stocks

Il est important de mesurer la valeur des composantes des stocks et ce, aussi bien pour les études économiques que pour le calcul de la valeur de la production économique. On demande aux répondants de déclarer la valeur comptable (au prix coûtant) de leurs matières premières et composantes, de leurs biens / travaux en cours de fabrication et de leurs produits finis fabriqués séparément. Dans certains cas, les répondants estiment la valeur totale de leurs stocks, que l'on répartit ensuite en fonction des pourcentages déclarés dans le cadre de l'EAMEF. Les niveaux des stocks sont calculés pour l'ensemble du Canada et non par province.

3. Commandes

a) Commandes en carnet

Les commandes en carnet sont une réserve de commandes qui généreront les ventes de biens fabriqués futures, à la condition qu'elles ne soient pas annulées. À l'instar des stocks, les commandes en carnet et les nouvelles commandes sont estimées pour l'ensemble du Canada et non par province.

L'EMIM produit des estimations relativement aux commandes en carnet pour toutes les industries à l'exception de celles où les commandes sont habituellement livrées à partir des stocks détenus et où l'on ne tient pas, par conséquent, de carnet de commande.

b) Nouvelles commandes

Les nouvelles commandes représentent la demande courante pour les produits manufacturés. Les estimations portant sur les nouvelles commandes sont établies à partir des données sur les ventes de biens fabriqués et les commandes en carnet. Toutes les ventes de biens fabriqués effectuées au cours d'un mois donné résultent d'une commande reçue ce mois-là ou avant. Ainsi, la valeur des nouvelles commandes s'obtient en additionnant les ventes de biens fabriqués au cours d'un mois donné avec la différence entre les commandes en carnet pour ce mois et les commandes en carnet pour le mois précédent.

4. Biens non-durables/biens durables

a) Industries des biens non-durables

Aliments (SCIAN 311),
Boissons et produits du tabac (312),
Usines de textiles (313),
Usines de produits textiles (314),
Vêtements (315),
Produits en cuir et produits analogues (316),
Papier (322),
Impressions & activités connexes de soutien (323),
Produits du pétrole et du charbon (324),
Produits chimiques (325) et
Produits en caoutchouc et en plastique (326).

b) Industries des biens durables

Produits en bois (SCIAN 321),
Produits minéraux non-métalliques (327),
Première transformation des métaux (331),
Fabrication des produits métalliques (332),
Machines (333),
Produits informatiques et électroniques (334),
Matériel, appareils et composants électriques (335),
Matériel de transport (336),
Meubles et produits connexes (337) et
Activités diverses de fabrication (339).

CONCEPTION DE L'ENQUÊTE ET MÉTHODOLOGIE

L'Enquête mensuelle sur les industries manufacturières (EMIM) a fait l'objet d'un vaste remaniement à compter du mois de référence août 1999.

L'analyse conceptuelle

En 1998, il a été décidé qu'avant d'entreprendre tout travail de remaniement, les concepts et les définitions de base du programme seraient vérifiés.

Cela a été fait de deux façons : premièrement, une analyse des exigences des utilisateurs a été amorcée. À cet effet, un rapport interne a été réexaminé pour s'assurer qu'on avait satisfait aux exigences des utilisateurs au cours de cet exercice. De plus, une analyse auprès des principaux utilisateurs des comptes nationaux a été entreprise. Cette analyse avait principalement pour objet de repérer toute faille présente dans les données.

Deuxièmement, avec en mains l'information sur ces failles et exigences, une étude des répondants de l'enquête a été menée pour vérifier leur capacité de fournir des données existantes et nouvelles. L'étude avait également pour objet de vérifier si les définitions utilisées par les analystes de l'enquête correspondaient à celles utilisées par les répondants.

Cet exercice a entraîné une réduction du nombre de questions de l'enquête, soit de seize à sept. La plupart des questions abandonnées portaient sur les ventes de biens fabriqués à l'égard de travaux partiellement exécutés.

En 2007, la terminologie de l'EMIM a été mise à jour afin qu'elle soit conforme au plan comptable (PC). Avec la diffusion du mois de référence d'août 2007, l'EMIM a harmonisé ses concepts avec ceux de l'EAMEF. La variable anciennement intitulée « Livraisons » correspond désormais à « Vente de biens fabriqués ». En outre, des modifications mineures ont été apportées aux noms des composants des stocks. Les définitions et les renseignements issus de l'enquête n'ont toutefois pas été modifiés.

Méthodologie

La nouvelle conception de l'échantillon tient compte du Système de classification des industries de l'Amérique du Nord (SCIAN) de 1997 et accorde une importance beaucoup plus grande aux données provinciales. La stratification est effectuée par province avec des exigences égales en matière de qualité pour chaque province. Les grandes unités sont sélectionnées avec certitude, alors que les petites unités le sont selon une probabilité fondée sur la qualité souhaitée de l'estimation pour chaque cellule.

La technique d'estimation permet à l'enquête de produire des estimations à partir du SCIAN. Les séries continueront également à faire l'objet d'un rapprochement avec l'EAMEF. Des estimations provinciales seront produites pour toutes les variables. Une mesure de la qualité (CV) sera également produite.

Volets de conception de l'enquête

Population cible et base de sondage

Le Registre des entreprises de Statistique Canada a fourni la base de sondage pour l'EMIM. La population cible de l'EMIM comprend tous les établissements statistiques figurant au Registre des entreprises qui sont classés dans le secteur manufacturier (le SCIAN). La base de sondage de l'EMIM est déterminée à partir de la population cible, après avoir enlevé les établissements faisant partie de la tranche inférieure de 5 % de l'estimation totale des ventes de biens fabriqués pour chaque province. Ces établissements ont été exclus de la base afin de réduire la taille de l'échantillon sans influencer significativement sur la qualité.

L'échantillon

L'échantillon de l'EMIM est un échantillon aléatoire formé de quelque 10 500 établissements. Un nouvel échantillon a été choisi au cours de l'automne 2006, puis un essai parallèle a été réalisé sur une durée de six mois (du mois de référence de septembre 2006 au mois de référence de février 2007). L'échantillon renouvelé est officiellement devenu le nouvel échantillon de l'EMIM à partir de janvier 2007.

Cette étape marque le premier processus de renouvellement de l'échantillon de l'EMIM depuis 2002. Le processus consiste à faire en sorte que l'échantillon soit aussi récent et à jour que possible. Tous les établissements de l'échantillon sont renouvelés en tenant compte des changements dans la valeur de leurs ventes de biens fabriqués; les unités qui ne sont plus actives sont enlevées de l'échantillon et certaines petites unités sont remplacées par d'autres dans la partie de l'échantillon basée sur la TPS alors que d'autres s'alternent dans l'échantillon.

Avant la sélection, la base de sondage est subdivisée en cellules industrie-province. On a surtout utilisé les codes du SCIAN. Selon le nombre d'établissements dans chaque cellule, on a regroupé au sein d'autres subdivisions (appelées strates) les établissements de taille similaire. Pour déterminer la taille d'un établissement, on s'est basé sur les chiffres les plus récents disponibles quant à la valeur annuelle des ventes de biens fabriqués ou des ventes.

Chaque cellule industrie-province comporte une strate à tirage complet où on retrouve les établissements échantillonnés chaque mois avec certitude. Cette strate comprend les plus grandes entreprises statistiques, c'est-à-dire celles qui ont le plus d'impact sur les estimations d'une cellule industrie-province particulière. Ces grandes entreprises statistiques représentent 45 % de l'estimation nationale des ventes de biens fabriqués.

Chaque cellule industrie-province ne peut compter plus de trois strates à tirage partiel. Les établissements de ces strates n'ont pas tous à être échantillonnés avec certitude. On prélève un échantillon aléatoire sur les strates restantes. Un poids égal à l'inverse de la probabilité de sélection est attribué aux réponses de ces établissements échantillonnés. Au sein des cellules à tirage partiel, un échantillon doit être composé au minimum de 10 établissements afin d'obtenir une meilleure stabilité.

La partie à tirage nul de l'échantillon est désormais estimée à partir des données administratives, ce qui fait que 100 % de l'environnement de l'échantillon est couvert. L'estimation de la partie à tirage nul a également permis d'améliorer l'efficacité puisqu'une partie à tirage nul plus importante a été délimitée et l'échantillon a pu être utilisé de manière plus efficace sur la plus petite partie échantillonnée de la base.

Collecte des données

Seulement un sous groupe des établissements de l'échantillon est envoyé pour la collecte. Pour le restant des unités des données administratives sont utilisées pour dériver les ventes de biens fabriqués.. Pour les établissements qui sont collectés, la collecte des données, la saisie des données, la vérification préliminaire et le suivi auprès des non-répondants sont effectués par les bureaux régionaux de Statistique Canada. On communique avec les établissements échantillonnés par la poste ou par téléphone, selon ce qu'ils préfèrent. La saisie des données et la vérification préliminaire sont effectuées en même temps afin de garantir la validité des données.

Dans certains cas, on reçoit des rapports globaux des entreprises ou compagnies qui comptent plus d'un établissement au sein de l'échantillon et où les répondants préfèrent ne pas produire un rapport distinct pour chaque établissement. On fait immédiatement le suivi auprès des entreprises qui n'ont pas répondu ou dont les données contiennent des erreurs.

Utilisation de données administratives

La gestion du fardeau de réponse est un défi constant pour Statistique Canada. Afin de tenter de réduire le fardeau de réponse, particulièrement auprès des petites entreprises, Statistique Canada a étudié différentes options de rechange à la réalisation d'enquêtes. Les fichiers de données administratives sont une grande source de données relatives aux entreprises et Statistique Canada travaille actuellement à l'exploitation du plein potentiel de cette riche source de données. De ce fait, depuis le mois de référence d'août 2004, l'EMIM a réduit le nombre d'établissements simples de l'échantillon qui sont interviewés directement et tire plutôt les données sur les ventes de biens fabriqués de ces établissements des dossiers de la taxe sur les produits et services (TPS) au moyen d'un modèle statistique. Le modèle rend compte de la différence entre les ventes de biens fabriqués (déclarées de l'EMIM) et les ventes (déclarées aux fins de la TPS) en plus du délai entre la période de référence de l'enquête et la période de référence du dossier de TPS.

Parallèlement à l'échantillon le plus récent, depuis janvier 2007, environ 2 500 établissements simples ont été sélectionnés pour composer la partie de l'échantillon basée sur la TPS.

Les estimations des stocks et des commandes en carnet des établissements dont les données sur les ventes de biens fabriqués sont basées sur la TPS sont obtenues au moyen du système d'imputation de l'EMIM. Le système d'imputation applique, aux valeurs des mois précédents, les variations mensuelles et annuelles des entreprises répondantes semblables. Avec le plus récent échantillon, on a raffiné les règles d'inclusion des établissements dont les données sont basées sur la TPS de manière à accroître le nombre de ces établissements dans les industries qui tiennent des stocks plus bas. Ainsi, on réduira au minimum les répercussions des établissements dont les données sont basées sur la TPS pour lesquels on doit estimer les stocks.

Des renseignements détaillés sur la méthodologie utilisée afin de modéliser les estimations des ventes de biens fabriqués dérivées de sources de données administratives se trouvent dans le document « **l'Enquête mensuelle sur les industries manufacturières : l'utilisation de données administratives** » (catalogue no 31-533-XIF).

QUALITÉ DES DONNÉES

Vérification statistique et imputation

Les données sont analysées à l'intérieur de chaque cellule industrie-province. Les valeurs extrêmes sont inscrites sur une liste, pour qu'on puisse en faire un contrôle en fonction de l'ampleur de l'écart par rapport à la moyenne. On communique avec les répondants afin de vérifier les valeurs extrêmes. Les enregistrements qui sont rejetés à la vérification statistique sont considérés comme des valeurs aberrantes et ne servent pas au calcul des valeurs d'imputation.

Des valeurs d'imputation sont attribuées aux cas de non-réponse, pour les établissements qui n'ont pas répondu ou qui n'ont répondu que partiellement au questionnaire d'enquête. Diverses méthodes d'imputation (tendances des cellules industrie-province, réponses antérieures, EAMEF, etc.) sont employées selon la variable exigeant le traitement. Après l'imputation, le personnel de l'EMIM effectue une vérification finale des réponses qui ont été imputées.

Révisions

Parallèlement à l'élaboration des estimations préliminaires du mois courant, les estimations des trois mois précédents sont révisées pour tenir compte des réponses tardives. Les données sont révisées lorsqu'on reçoit des réponses tardives ou lorsqu'on a reçu antérieurement des réponses erronées.

Estimation

Les estimations sont fournies en fonction des réponses d'un échantillon d'établissements manufacturiers associées aux données administratives relatives à une partie des plus petits établissements. L'échantillonnage de l'enquête comprend la couverture totale des établissements manufacturiers importants de chaque industrie-province, ainsi que la couverture partielle des petites et moyennes entreprises. On répartit au prorata les rapports combinés des sociétés à unités multiples parmi leurs établissements, et les ajustements faits pour tenir compte de la facturation provisoire traduisent les recettes reçues pour les travaux effectués dans le cadre des contrats importants. Environ 2 500 des petites et moyennes entreprises sondées ne reçoivent pas de questionnaire. Les données relatives à leurs ventes de biens fabriqués sont obtenues à partir des recettes reçues indiquées dans les fichiers de la TPS. La partie qui n'est pas représentée dans l'échantillon (la partie à tirage nul) comporte les établissements en-dessous du seuil spécifié dans chaque province et industrie. Les sous-totaux relatifs à cette partie sont également déduits en fonction de leurs recettes.

On estime les valeurs des ventes de biens fabriqués, des stocks et des commandes en carnet en pondérant les réponses, les valeurs dérivées des fichiers de la TPS et les imputations par le nombre d'établissements que chacune représente. Les estimations pondérées sont ensuite ajoutées à la partie à tirage nul. Les estimations des ventes de biens fabriqués sont produites par province, mais aucun détail géographique n'est établi pour les stocks et les commandes, car nombre d'entreprises ne peuvent pas déclarer la valeur comptable de ces articles mensuellement.

Étalonnage

Jusqu'en 2003 (inclusivement), l'Enquête mensuelle sur les industries manufacturières (EMIM) était étalonnée en fonction de l'Enquête annuelle sur les manufactures et l'exploitation forestière (EAMEF). L'étalonnage consistait en l'examen régulier des estimations de l'EMIM dans le cadre des données annuelles fournies par l'EAMEF. L'étalonnage réaligait le niveau annualisé de l'EMIM en fonction des dernières données annuelles vérifiées fournies par l'EAMEF.

En 2006-2007, Statistique Canada a mené une enquête importante pour déterminer s'il était judicieux de conserver le processus d'étalonnage. Les résultats ont indiqué que l'étalonnage des estimations de l'EMIM en fonction de l'EAMEF devrait cesser. Grâce au renouvellement de l'échantillon de l'EMIM en 2007, on a déterminé que l'étalonnage ne serait plus nécessaire (rétroactif à l'année 2004) puisque l'EMIM représentait depuis précisément 100 % de l'univers de l'échantillon. Le rapprochement entre les données de l'EMIM et l'EAMEF sera maintenu afin de résoudre d'éventuelles anomalies.

Depuis le mois de référence de janvier 2007, un nouvel échantillon a été introduit. La pratique normale veut qu'au bout de quelques années l'échantillon soit renouvelé afin de garantir que la base de sondage est à jour sur le plan des naissances, des décès et des autres changements relatifs à la population. L'échantillon renouvelé est lié au niveau détaillé pour empêcher les ruptures de données et pour s'assurer de la continuité de la série chronologique. Il est conçu de façon à mieux représenter l'industrie manufacturière aux échelles nationale et provinciale.

Comparaisons et rapprochements avec d'autres sources de données

À chaque année, au moment où la Section de l'Enquête annuelle sur les manufactures et l'exploitation forestière élabore ses estimations annuelles, la Section de l'Enquête mensuelle sur les industries manufacturières travaille conjointement avec celle de l'EAMEF à comparer et à rapprocher les écarts de valeurs importants entre les données fondées sur les exercices financiers que fournit l'EAMEF et les données annualisées de l'EMIM aux niveaux des strates et des industries.

L'objectif de cet exercice de rapprochement de données est de faire ressortir et de résoudre les différences importantes entre les deux enquêtes et d'aider à réduire au minimum les différences entre les microdonnées de l'EMIM et de l'EAMEF.

Erreurs d'échantillonnage et autres erreurs

Ce bulletin présente des estimations fondées sur une enquête par échantillonnage qui risquent, par conséquent, d'être entachées d'erreurs. La section qui suit vise à faciliter, pour le lecteur, l'interprétation des estimations qui sont publiées.

Les estimations établies sur la base d'une enquête par échantillonnage sont sujettes à différents types d'erreurs. On les regroupe en deux grandes catégories: les erreurs qui sont dues à l'échantillonnage et celles qui ne le sont pas.

1. Erreurs d'échantillonnage

Les erreurs d'échantillonnage constituent un risque inhérent aux enquêtes par échantillonnage. Elles découlent de la différence notée entre la valeur d'une variable échantillonnée aléatoirement et celle d'une variable obtenue grâce à un recensement (ou la moyenne de toutes les valeurs aléatoires possibles). Ce genre d'erreur existe parce que les observations portent uniquement sur un échantillon, non sur l'ensemble de la population.

L'erreur d'échantillonnage dépend de facteurs tels que la taille de l'échantillon, la variabilité de la population, le plan de sondage et la méthode d'estimation. Par exemple, pour une taille donnée d'échantillon, l'erreur d'échantillonnage sera fonction de la méthode de stratification adoptée, de l'attribution de l'échantillon, du choix des unités sondées et de la méthode de sélection. (On peut même, dans le cadre d'un seul plan de sondage, effectuer plusieurs calculs pour arriver à la méthode d'estimation la plus efficace.) La plus importante caractéristique des sondages probabilistes, c'est que l'erreur d'échantillonnage peut être mesurée à partir de l'échantillon lui-même.

2. Erreurs non liées à l'échantillonnage

Les erreurs qui ne sont pas liées à l'échantillonnage découlent d'une faille systématique dans la structure de la procédure de collecte des données ou dans l'élaboration d'une variable ou de l'ensemble des variables à l'étude. Ces erreurs occasionnent une différence entre la valeur d'une variable obtenue par échantillonnage ou par recensement et la valeur réelle de cette variable. Ces erreurs se retrouvent tant dans les recensements que dans les enquêtes par échantillonnage. Elles sont dues à un ou plusieurs facteurs parmi les suivants:

a) Erreur de couverture. L'erreur peut résulter d'un listage incomplet et d'une couverture insuffisante de la population visée.

b) La réponse. Ici, l'erreur peut être attribuable à la conception du questionnaire et aux caractéristiques de la question, à l'incapacité ou au refus de l'enquêté de fournir des renseignements exacts, à l'interprétation fautive des questions ou à des problèmes d'ordre sémantique.

c) La non-réponse. Certains enquêtés refusent de répondre, tandis que d'autres en sont incapables ou encore répondent trop tard. Les données relatives aux non-répondants peuvent être imputées à partir des chiffres fournis par les répondants ou à l'aide des statistiques antérieures sur les non-répondants, lorsque celles-ci existent.

On ne connaît généralement pas avec précision l'importance de l'erreur d'imputation; celle-ci varie beaucoup selon les caractéristiques qui distinguent les répondants des non-répondants. Comme ce type d'erreur prend habituellement de l'ampleur à mesure que diminue le taux de réponse, on s'efforce d'obtenir le meilleur taux de réponse possible.

d) Le traitement. L'erreur peut se produire lors des diverses étapes du traitement (codage, entrée, vérification, pondération, totalisation, etc.). Il est difficile de mesurer les erreurs non liées à l'échantillonnage. De plus, il faut les cerner à un niveau où elles ne nuisent pas à l'utilisation ou à l'interprétation des chiffres définitifs.

Des mesures ont été prises afin de minimiser les erreurs non liées à l'échantillonnage. Ainsi, les unités ont été définies avec beaucoup de précision au moyen des listes les plus à jour. Les questionnaires ont été conçus avec soin afin de réduire au minimum les différentes interprétations possibles. De plus, les diverses étapes de vérification et de traitement ont fait l'objet de contrôles d'acceptation détaillés et on n'a absolument rien négligé pour que le taux de non-réponse et le fardeau de réponse soient faibles.

Évaluation de l'erreur d'échantillonnage et de l'erreur non liée à l'échantillonnage

1. Évaluation de l'erreur d'échantillonnage

L'échantillon utilisé aux fins de la présente enquête est un des nombreux échantillons de même taille qui auraient pu être choisis selon le même plan et les mêmes conditions. Si chaque échantillon pouvait faire l'objet d'une enquête menée essentiellement dans les mêmes conditions, il faudrait s'attendre à ce que l'estimation calculée varie d'un échantillon à l'autre.

On nomme valeur probable l'estimation moyenne obtenue de tous les échantillons possibles. Autrement dit, la valeur probable est celle qu'on obtiendrait en recensant toute la population dans des conditions identiques de collecte et de traitement. Une estimation calculée à partir d'une enquête par échantillonnage est dite précise lorsqu'elle s'approche de la valeur probable.

Les estimations fondées sur un échantillon peuvent ne pas correspondre à la valeur probable. Cependant, comme les estimations proviennent d'un échantillon probabiliste, il est possible d'en mesurer la variabilité par rapport à leur valeur probable. La variance d'une estimation, qui en mesure la précision, se définit comme la moyenne, parmi tous les échantillons possibles, des carrés de la différence entre l'estimation et la valeur probable.

Une fois qu'on a calculé l'estimation et sa variance, il devient possible de calculer d'autres mesures de précision. Par exemple, l'erreur-type, soit la racine carrée de la variance, mesure l'erreur d'échantillonnage dans la même unité que l'estimation (en dollars, notamment). Autrement dit, l'erreur-type mesure la précision en termes absolus. Par contre, le coefficient de variation, c'est-à-dire l'erreur-type divisée par l'estimation, mesure la précision en termes relatifs. Ainsi, l'emploi du coefficient de variation facilite la comparaison de l'erreur d'échantillonnage de deux estimations.

Dans cette publication, on utilise le coefficient de variation pour évaluer l'erreur d'échantillonnage des estimations. Cependant, puisque le coefficient de variation publié pour cette enquête est calculé à partir des réponses des unités, il mesure aussi une certaine erreur non liée à l'échantillonnage.

Voici la formule utilisée pour calculer les coefficients de variation (CV) du tableau 1 :

$$CV(X) = \frac{S(X)}{X}$$

où X est l'estimation et S(X) est l'écart type de X.

Le coefficient de variation est exprimé en pourcentage dans cette publication.

L'estimation et le coefficient de variation nous permettent de construire des intervalles de confiance autour de l'estimation. Ainsi, pour notre échantillon, on peut affirmer qu'avec un niveau de confiance donné, la valeur probable est comprise dans l'intervalle de confiance construit autour de l'estimation. Par exemple, si le coefficient de variation d'une estimation de 12 000 000 \$ est égal à 10 %, l'écart type sera de 1 200 000 \$, soit l'estimation multipliée par le coefficient de variation. Alors, on peut affirmer avec un niveau de confiance de 68 % que la valeur probable sera contenue dans l'intervalle d'une longueur égale à un écart-type autour de l'estimation, soit entre 10 800 000 \$ et 13 200 000 \$. Ou encore, on peut affirmer avec une confiance de 95 % que la valeur probable sera contenue dans l'intervalle d'une longueur de deux écart-types autour de l'estimation, soit entre 9 600 000 \$ et 14 400 000 \$.

Le tableau explicatif 1 qui suit indique, pour l'ensemble du secteur manufacturier, les coefficients de variation nationaux, exprimé en pourcentage, des éléments de l'EMIM. Pour les CV à d'autres niveaux d'agrégation, veuillez communiquer avec la section de Marketing et diffusion au (613) 951-9497, sans frais au 1-866-873-8789 ou par courriel (manufact@statcan.ca).

Tableau explicatif 1: CV nationaux par caractéristique					
MOIS	Ventes de biens fabriqués	Stocks de matières premières et composantes	Stocks de biens/travaux en cours de fabrication	Stocks de produits finis fabriqués	Commandes en carnet
	%	%	%	%	%
Juin 2009	0,69	1,17	1,89	1,25	1,08
Juillet 2009	0,73	1,14	2,08	1,31	1,09
Août 2009	0,75	1,13	1,78	1,33	1,22
Septembre 2009	0,77	1,11	1,69	1,34	1,19
Octobre 2009	0,78	1,15	1,82	1,36	1,22
Novembre 2009	0,87	1,11	1,83	1,38	1,24
Décembre 2009	0,83	1,14	1,98	1,39	1,22
Janvier 2010	0,83	1,17	2,12	1,34	1,24
Février 2010	0,88	1,16	2,43	1,35	1,23
Mars 2010	0,88	1,18	2,50	1,37	1,23
Avril 2010	0,79	1,18	2,37	1,43	1,23
Mai 2010	0,85	1,21	2,51	1,46	1,30
Juin 2010	0,87	1,18	2,64	1,47	1,29

2. Évaluation de l'erreur non liée à l'échantillonnage

L'enquête par échantillonnage et le recensement cherchent tous deux à déterminer la valeur exacte de l'ensemble. L'estimation est dite précise si elle se rapproche de cette valeur. Bien qu'il s'agisse d'une valeur souhaitable, il n'est pas réaliste de supposer que la valeur exacte de chaque unité de l'ensemble ou de l'échantillon peut être obtenue et traitée sans erreur. La différence entre la valeur probable et la valeur exacte de l'ensemble s'appelle le biais. On ne peut calculer les biais systématiques des données en recourant aux mesures de probabilité de l'erreur d'échantillonnage décrites auparavant. La précision d'une estimation est déterminée par l'effet conjugué des erreurs d'échantillonnage et des erreurs non liées à l'échantillonnage.

On distingue trois types d'erreur non liée à l'échantillonnage dans l'EMIM : l'erreur due à la non-réponse, l'erreur d'imputation et l'erreur due à la vérification. Afin d'aider l'utilisateur à évaluer ces trois types d'erreur, le tableau explicatif 2 présente les taux pondérés correspondants. Voici un exemple qui illustre ce qu'est un taux pondéré. Supposons que le taux de réponse d'une cellule comportant un échantillon de 20 unités dont cinq répondent lors d'un mois donné atteint 25 %. Si les cinq unités déclarantes représentent 8 millions de dollars sur l'estimation globale de 10 millions de dollars, le taux de réponse pondéré s'élève à 80 %.

Les trois taux pondérés mentionnés dans le tableau explicatif 2 se définissent comme suit : le taux de réponse pondéré est le pourcentage de l'estimation globale d'un élément qui est basé sur des données déclarées (données vérifiées exclues) ; le taux d'imputation pondéré est le pourcentage de l'estimation globale d'un élément qui est fondé sur des données imputées ; le taux de vérification pondéré est le pourcentage de l'estimation globale d'un élément qui est basé sur des données vérifiées (lesquelles peuvent avoir été déclarées ou imputées initialement).

Le tableau explicatif 2 qui suit présente les trois types de taux pondérés pour chaque élément de l'ensemble de l'activité manufacturière à l'échelle nationale. Dans le tableau, les taux sont exprimés en pourcentage.

Tableau explicatif 2: Taux pondérés nationaux par source et par caractéristique						
Caractéristique	Source d'enquête			Source des données administratives		
	Réponse	Imputation	Vérification	Modélisation	Imputation	Vérification
	%	%	%	%	%	%
Ventes de biens fabriqués	85,03	4,25	3,72	6,31	0,53	0,16
Matières premières et composantes	72,69	12,66	4,53	0	10,13	0,00
Stocks de biens/travaux en cours de fabrication	58,62	9,36	24,15	0	7,68	0,19
Stocks de produits finis fabriqués	74,28	10,92	5,07	0	8,17	1,55
Commandes en carnet	54,24	6,35	34,63	0	3,84	0,95

Interprétation simultanée des mesures d'erreur

Il faut tenir compte simultanément de la mesure d'erreurs non liées à l'échantillonnage ainsi que du coefficient de variation pour avoir un aperçu de la qualité des estimations. Plus le coefficient de variation sera bas et que le taux de réponse pondéré sera élevé, meilleure sera l'estimation publiée.

Désaisonnalisation

Les séries chronologiques économiques comportent les éléments essentiels à la description, l'explication et la prévision du comportement d'un phénomène économique. Ce sont des enregistrements statistiques de l'évolution des processus économiques dans le temps. L'observation par les économistes et les statisticiens de l'activité économique à l'aide des séries chronologiques a donc permis de distinguer quatre composantes principales du comportement de ces séries : le mouvement à long terme ou tendance, le mouvement cyclique, les variations saisonnières et les fluctuations irrégulières. Ces mouvements sont causés par différents facteurs, soit économiques, climatiques ou institutionnels. Les variations saisonnières sont les fluctuations périodiques plus ou moins régulières qui se produisent au cours d'une année en raison du cycle météorologique normal, des congés fixes et d'autres événements qui se répètent à intervalles avec une certaine régularité pour influencer de façon significative le taux d'activité économique.

Afin de favoriser l'interprétation exacte de l'évolution fondamentale d'un phénomène économique et de produire une meilleure prévision, Statistique Canada rajuste les séries chronologiques au moyen de la méthode de désaisonnalisation X12-ARMMI de façon à minimiser l'impact des variations saisonnières sur les séries. Cette technique consiste essentiellement à ajouter les estimations d'une année de données brutes à la fin de la série initiale avant de procéder à la désaisonnalisation proprement dite. Les données estimées proviennent de prévisions réalisées par des modèles ARMMI (modèles autorégressifs à moyennes mobiles intégrées) du type Box-Jenkins.

Le programme X12 fait surtout appel à la méthode de rapport aux moyennes mobiles pour effectuer le lissage de la série modifiée et obtenir une estimation provisoire de la tendance-cycle, calculer les rapports de la série initiale (ajustée) aux estimations de la tendance-cycle, et estimer les facteurs saisonniers à partir de ces dits rapports. Les facteurs saisonniers définitifs ne sont produits que lorsque ces opérations ont été exécutées à plusieurs reprises.

La technique utilisée consiste essentiellement, dans un premier temps, à corriger la série initiale de toute sorte d'effets indésirables, tels l'effet des jours ouvrables et l'effet de Pâques, par un module appelé regARIMA.

L'estimation de ces effets se fait grâce à l'utilisation de modèles de régression à erreurs ARMMI. On peut également extrapoler la série d'au moins une année à l'aide du modèle. Dans un deuxième temps, la série brute, pré-ajustée et extrapolée s'il y a lieu, est désaisonnalisée par la méthode X-12.

Les étapes déterminant les facteurs saisonniers nécessaires au calcul des données désaisonnalisées finales sont exécutées à chaque mois. Cette approche garantit que la série non désaisonnalisée, à partir de laquelle sont calculées les estimations des facteurs saisonniers, inclut toutes les données les plus récentes relativement à ladite série, c.-à-d., les données non désaisonnalisées qui portent sur le mois courant et les données non désaisonnalisées révisées du mois précédent.

Bien que la désaisonnalisation permette de mieux comprendre la tendance-cycle fondamentale d'une série, la série désaisonnalisée n'en contient pas moins une composante irrégulière. De légères variations d'un mois à l'autre dans la série désaisonnalisée peuvent n'être que de simples mouvements irréguliers. Pour avoir une meilleure idée de la tendance fondamentale, les utilisateurs doivent donc examiner les séries désaisonnalisées d'un certain nombre de mois.

Les séries de données agrégées au niveau du Canada sont maintenant désaisonnalisées de façon directe, ce qui signifie que les totaux désaisonnalisés sont obtenus au moyen de X12 ARMMI. Ces totaux sont ensuite utilisés pour réconcilier les séries de totaux provinciaux qui ont été désaisonnalisés séparément au préalable.

Pour ce qui est des autres séries agrégées, on a utilisé la désaisonnalisation indirecte. En d'autres mots, leurs totaux désaisonnalisés sont dérivés de façon indirecte en faisant la somme des genres de commerce désaisonnalisés séparément au préalable.

Tendance

Une série désaisonnalisée peut montrer encore les effets d'irrégularités et de circonstances spéciales ; et cela peut masquer la tendance. La tendance à court terme montre la direction prise dans une série désaisonnalisée en affectant une moyenne à travers les mois de sorte que ces mouvements irréguliers soient aplanis. Le résultat est une série plus stable. La tendance pour le dernier mois apparaît sous réserve de rectification, car les valeurs des mois à venir sont incluses dans le calcul de la moyenne.

Déflation des ventes de biens fabriqués, des commandes et des stocks manufacturiers

La variation des valeurs des données publiées dans le cadre de l'Enquête mensuelle sur les industries manufacturières (EMIM) peut être attribuable à l'évolution des prix ou des quantités mesurées, ou des deux. Pour analyser l'activité du secteur manufacturier, il est souvent souhaitable de séparer les variations causées par des changements de prix de celles causées par des changements des quantités produites. Cet ajustement est connu sous le nom de déflation.

La déflation consiste à diviser les valeurs aux prix courants obtenues grâce à l'enquête par des indices de prix pertinents afin d'obtenir des estimations évaluées aux prix d'une période antérieure, actuellement l'année 2002. On dit des valeurs finales qu'elles sont « aux prix de 2002 ». Il faut noter que l'expression « aux prix courants » désigne le moment où s'est déroulée l'activité et non le moment présent, ni le moment de la compilation.

Les estimations déflatées de l'EMIM reflètent les prix de 2002, qui constitue l'année de base. On a choisi l'année 1997 parce qu'elle correspond à l'année de base des indices de prix utilisés pour déflater les estimations de l'EMIM. L'utilisation des prix d'une année de base pour mesurer l'activité courante produit une mesure représentative du volume d'activité actuel par rapport à cette année de base. Les variations actuelles du volume ne se reflètent convenablement dans les mesures à prix constants que si l'importance relative actuelle des industries n'est pas très différente de celle observée pendant l'année de base.

La déflation des estimations de l'EMIM est effectuée à un niveau industriel très détaillé, équivalant aux classes à 6 chiffres du Système de classification des industries de l'Amérique du Nord (SCIAN). Pour chaque industrie à ce niveau, on utilise des indices de prix composites qui décrivent la fluctuation des prix des différents groupes de produits fabriqués par cette industrie.

Sauf quelques rares exceptions, les indices de prix sont des moyennes pondérées des indices des prix des produits industriels (IPPI). Les poids utilisés sont tirés des tableaux annuels des entrées-sorties du Canada et varient d'année

en année. Comme les tableaux d'entrées-sorties sont publiés avec un retard d'environ deux ans et demi, les poids utilisés pour les plus récentes années sont fondés sur les tableaux d'entrées-sorties les plus récents.

On utilise le même indice de prix pour déflater les ventes de biens fabriqués, les nouvelles commandes et les commandes en carnet d'une industrie. Les poids servant à la compilation de cet indice de prix sont tirés des tableaux de sorties, évalués aux prix à la production. Les prix à la production sont les prix des produits à leur sortie de l'établissement de fabrication et ne comprennent pas des éléments tels que les frais de transport, les taxes, etc. Ainsi, l'indice de prix de chaque industrie reflète la production des établissements de l'industrie.

Les indices de prix qu'on utilise pour déflater les stocks de biens/travaux en cours de fabrication et les stocks de produits finis fabriqués d'une industrie sont des moyennes mobiles de l'indice de prix utilisé pour les ventes de biens fabriqués. En ce qui concerne les stocks de biens/travaux en cours de fabrication, le nombre de termes de la moyenne mobile correspond à la durée du processus de fabrication. On obtient la durée en calculant la moyenne, au cours des 48 mois précédents, du ratio entre les stocks de biens/travaux en cours de fabrication à la fin du mois et la production de l'industrie, cette dernière étant égale aux ventes de biens fabriqués additionnées des variations des stocks de biens/travaux en cours de fabrication et de produits finis fabriqués.

En ce qui a trait aux stocks de produits finis fabriqués, le nombre de termes dans la moyenne mobile reflète la durée pendant laquelle un produit fini demeure en stock. On obtient ce chiffre, connu sous le nom de période de rotation des stocks, en calculant la moyenne, au cours des 48 mois précédents, du ratio entre les stocks de produits finis fabriqués à la fin du mois et les ventes de biens fabriqués.

Pour déflater les stocks de matières premières et composantes, les indices de prix correspondant à la consommation des matières premières sont obtenus en tant que moyennes pondérées des IPPI. Les poids sont tirés des tableaux d'entrées évalués aux prix à la consommation, c'est-à-dire que ces prix comprennent les marges de gros, les frais de transport et les taxes, etc. L'indice de prix qui en découle reflète donc la structure des coûts des matières premières de chaque industrie.

Les stocks de matières premières et composantes sont ensuite déflatés à l'aide d'une moyenne mobile de l'indice de prix correspondant à la consommation des matières premières. Le nombre de termes de la moyenne mobile correspond au taux de consommation des matières premières. Le taux est calculé comme étant la moyenne, au cours des quatre années précédentes, du ratio entre les stocks de matières premières et composantes à la fin de l'année et les entrées intermédiaires de l'industrie.