

CONCEPTS ET DÉFINITIONS

L'Enquête mensuelle sur les industries manufacturières (EMIM) produit des séries statistiques sur l'activité manufacturière — livraisons, stocks, commandes en carnet et nouvelles commandes. Chacune des valeurs associées à ces composantes représentent une projection mensuelle des données de l'Enquête annuelle des manufactures (EAM). Cette dernière recueille les données auprès de l'ensemble de l'industrie manufacturière canadienne.

Les données recueillies dans le cadre de l'Enquête mensuelle sur les industries manufacturières (EMIM) sont maintenant classées d'après le Système de classification des industries de l'Amérique du Nord (SCIAN) de 1997, qui remplace la Classification type des industries (CTI) de 1980. L'année de référence 2000 est la dernière pour laquelle des données sont publiées sur la base de la CTI.

L'Enquête mensuelle sur les industries manufacturières a adopté le SCIAN pour l'année de référence 2001, qui répertorie plus de deux cents industries manufacturières, tandis que les données relatives aux années précédentes ont été recalculées d'après le nouveau système de classification, en remontant jusqu'à 1992. Des séries désaisonnalisées sont disponibles pour les principaux agrégats.

1. Livraisons

Les livraisons 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 livraisons, pour le Canada et les provinces, sont disponibles selon le SCIAN, au niveau de détails des six chiffres.

Dans le cas des industries des aéronefs et des pièces d'aéronef et de la construction navale, on se fonde sur la valeur de la production plutôt que sur celle des livraisons. Pour déterminer cette valeur, on ajuste la valeur mensuelle des livraisons en fonction de la variation mensuelle des stocks détenus pour les produits en cours et les produits finis). 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 la livraison 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, de leurs produits en cours et de leurs produits finis 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'EAM. 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 livraisons 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 livraisons et les commandes en carnet. Toutes les livraisons 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 livraisons 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.

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 — 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 livraisons à l'égard de travaux partiellement exécutés (travaux en cours).

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.

On a également profité de l'occasion pour introduire le renouvellement de l'échantillon pour la première fois dans le plan d'enquête. La plupart des plus petites compagnies à qui on demande de participer à l'enquête ne le font que pour un temps limité.

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'EAM. Des estimations provinciales seront produites pour toutes les variables. Une mesure de la qualité (CV) sera également produite.

Volets de l'enquête remaniée

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 2 % de l'estimation totale des livraisons manufacturières 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 représentatif formé de quelque 11 000 établissements. 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 livraisons ou des ventes. Chaque cellule industrie-province a une strate à tirage complet formée des 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 font partie de la tranche supérieure de 45 % de l'estimation nationale des livraisons manufacturières.

Chaque cellule industrie-province ne peut compter plus de trois strates à tirage partiel. Il n'est pas nécessaire que tous les établissements compris dans ces strates soient échantillonnés avec certitude. Un échantillon aléatoire est tiré à partir des autres strates. Un poids égal à l'inverse de la probabilité de sélection est attribué aux réponses de ces établissements échantillonnés.

L'échantillon initial a été prélevé vers la fin de 1998 et a été rafraîchi chaque mois par l'introduction d'un échantillon de nouvelles unités de population.

Collecte des données

La collecte des données, la saisie des données, la vérification préliminaire et le suivi auprès des non-répondants sont effectués par les sept 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.

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, EAM, 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

Pour établir les estimations, on applique un poids aux réponses des établissements. Ce poids est égal à l'inverse de la probabilité de sélection des établissements échantillonnés. Les unités à tirage complet s'autoreprésentent.

Étalonnage

Les données de l'EMIM font périodiquement l'objet d'un étalonnage en fonction des plus récents résultats de l'Enquête annuelle des manufactures (environ dix-huit mois après la fin de l'année de référence). L'étalonnage est une opération qui consiste à ajuster les estimations mensuelles au niveau des données de la plus complète Enquête annuelle des manufactures. Le résultat présente pour les deux enquêtes des totaux compatible au niveau agrégé. Les estimations de l'EMIM ont été étalonnées en fonction de celles de l'EAM de 1998 à l'occasion de la publication des données du mois de référence janvier 2001.

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

Ce genre d'erreur existe parce que les observations portent uniquement sur un échantillon, et non sur l'ensemble de la population. L'erreur dépend de facteurs tels que la taille de l'échantillon, la variabilité de la population, le plan de sondage et la méthode d'estimation. Pour une taille donnée d'échantillon par exemple, 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.)

Comme dans une enquête par échantillonnage, on tire des conclusions sur l'ensemble d'une population à partir des données concernant une partie seulement. Les résultats seront probablement différents de ceux qu'on aurait obtenus en recensant toute la population dans les mêmes conditions. La principale caractéristique des sondages probabilistes, c'est que l'erreur d'échantillonnage peut être directement mesurée à partir de l'échantillon.

2. Erreurs non liées à l'échantillonnage

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 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 Craig Kuntz, de l'Enquête mensuelle sur les industries manufacturières, au 951-7092 (kuncrai@statcan.ca).

CV nationaux par caractéristique

Mois	Livraisons	Stocks de matières premières	Stocks de produits en cours	Stocks de produits fins	Commandes en carnets
			%		
Octobre 2000	0,61	1,17	0,86	1,32	1,57
Novembre 2000	0,62	1,16	0,95	1,36	1,75
Décembre 2000	0,56	1,17	0,74	1,48	1,81
Janvier 2001	0,61	1,13	0,81	1,47	1,86
Février 2001	0,57	1,13	0,82	1,39	1,57
Mars 2001	0,58	1,16	0,83	1,36	1,47
Avril 2001	0,57	1,19	0,78	1,41	1,53
Mai 2001	0,55	1,18	0,72	1,44	1,40
Juin 2001	0,59	1,18	0,70	1,49	1,49
Juillet 2001	0,61	1,09	0,73	1,47	1,43
Août 2001	0,59	1,12	0,77	1,52	1,46
Septembre 2001	0,60	1,11	0,79	1,50	1,44
Octobre 2001	0,61	1,13	0,75	1,55	1,56

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 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 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 qui suit présente les trois types de taux pondéré pour chaque élément de l'ensemble de l'activité manufacturière à l'échelle nationale. Dans le tableau, les taux (exprimés en pourcentage) sont des moyennes calculées pour les derniers treize mois.

Taux pondérés nationaux moyens par caractéristique

Caractéristique	Réponse		Imputation	Vérification
		%		
Livraisons	93,42		4,39	2,19
Matières premières	80,59		16,57	2,84
Produits en cours	82,48		11,08	6,44
Produits finis	84,36		12,70	2,94
Commandes en carnet	86,51		5,41	8,08

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 X11ARMMI/88 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.

La partie X-11 du programme X11ARMMI/88 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.

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.

Le total désaisonnalisé au niveau du Canada est dérivé de façon indirecte en faisant la somme des genres de commerce désaisonnalisés séparément au préalable.