

### 3.3 Sources

#### 3.3.1 Surveyed Data

The majority of industries covered under the expenditures series are surveyed. All establishments selected for the sample during the three survey periods (see section 3.3.4) will receive either the regular survey questionnaire (short or long form), a specialized survey questionnaire (long or short form) or the new project questionnaire. The type of questionnaire an establishment receives depends on the industry, the expected level of expenditure, the survey being conducted and whether or not the establishment is classified as a new project (i.e. out of frame or outlier).

The regular short questionnaire is most often used during each of the three survey periods. This questionnaire collects basic information on capital construction, capital machinery and equipment, repair construction and repair machinery and equipment, gross book value, capacity utilization in the manufacturing and mining sectors, reasons for change in expenditures, work in progress and leasing. Note that establishments are asked to report repair expenditures on the actual survey only. An establishment will receive one of the other questionnaire types if it is expected to spend a large amount on capital, has been operating in a specialized industry or has been categorized as a new project.

The regular long questionnaire is used only during the actual survey period and is distributed to establishments that have previously reported large capital expenditures. This questionnaire goes beyond the basic data assembled by the short form to collect information related to asset detail, asset value, reason for disposals, interest payments capitalized, number of robots and leases by type of asset (see appendix).

Specialized questionnaires are used for the mining industry and the petroleum and natural gas industry. New project questionnaires are sent to new establishments that are considered to be either not yet on the frame because they are not in production or outliers on the frame.

Apart from surveying establishments, the capital expenditures series also uses reporting arrangements in the data collection process. Some respondents operating within Canada are unable to provide the required provincial breakdown of expenditures during the reporting periods. Consolidated reports are used to collect data from such respondents. These reports are subsequently allocated to the provinces based on related characteristics. It might also be the case that the number of establishments are too numerous for conventional sampling. To facilitate the reporting of capital expenditures by these establishments, data are collected through a reporting entity known as provincial establishments. However, the locations covered under the provincial establishment's report must all be within the same industry.

All respondents are asked to report expenditures for their 12 months fiscal period for which the final day occurs between April 1 of the reference year and March 31 of the following year.

### 3.3 Sources

#### 3.3.1 Données enquêtées

La majorité des secteurs couverts par les séries sur les investissements sont enquêtés. Tous les établissements sélectionnés durant les trois périodes d'enquête (voir section 3.3.4) reçoivent soit le questionnaire ordinaire (long ou abrégé), soit un questionnaire spécialisé (long ou abrégé) soit le questionnaire de nouveau projet. Le genre de questionnaire reçu par un établissement dépend de son secteur d'activité, du niveau de dépenses anticipé, de l'enquête en cause et du fait que l'établissement fait partie ou non des nouveaux projets (hors de la base de sondage ou cas aberrant).

Le questionnaire ordinaire abrégé est le plus utilisé pour les trois périodes d'enquête. Ce questionnaire vise à recueillir des renseignements de base portant sur les constructions, le matériel et l'outillage immobilisés, les réparations apportées aux constructions ainsi qu'au matériel et à l'outillage, la valeur comptable brute, l'utilisation de la capacité dans les secteurs manufacturier et minier, les motifs des changements de dépenses, les travaux en cours et les contrats de location à bail. Soulignons que l'établissement est tenu de revoir ses dépenses en réparations durant l'enquête sur les données réelles seulement. L'établissement reçoit l'un des autres types de questionnaire si on l'on croit qu'il consacre des sommes importantes aux immobilisations, s'il est engagé dans un secteur spécialisé ou s'il fait partie des nouveaux projets.

Le questionnaire ordinaire long ne sert que pour l'enquête sur les données réelles et vise les établissements ayant déjà déclaré des dépenses importantes. Il vise à rassembler des renseignements plus approfondis que le questionnaire abrégé et touche le détail des dépenses par type d'actif, la valeur brute des actifs, les motifs d'aliénation, les frais d'intérêt capitalisés, le nombre de robots et les contrats de location selon le genre de bien (voir annexe).

Les questionnaires spécialisés concernent le secteur minier et celui du pétrole et du gaz naturel. Les questionnaires de projets nouveaux sont destinés aux nouveaux établissements qui ne font pas encore partie de la base de sondage parce qu'ils ne produisent pas ou sont classés comme des cas aberrants.

Outre le sondage des établissements, la série des dépenses en immobilisations repose aussi sur des modes de cueillette particuliers. En effet, certains répondants opérant au Canada sont dans l'impossibilité de fournir une ventilation provinciale de leurs dépenses pour les périodes en cause. On utilise alors un rapport consolidé. La répartition entre les provinces est calculée à partir de ce rapport selon les caractéristiques pertinentes. Il peut aussi arriver que le nombre de sites administrés par l'établissement soit trop élevé pour un sondage classique. Dans ce cas, pour simplifier la déclaration, on fait appel, pour la collecte des données, à une entité appelée établissement provincial. Cependant, les sites représentés par l'établissement provincial doivent tous faire partie de la même industrie.

On demande aux répondants de déclarer les dépenses pour l'exercice financier de douze mois dont le dernier jour se situe entre le 1<sup>er</sup> avril de l'année demandée et le 31 mars de l'année suivante.

### 3.3.2 Non-Surveyed Data

Although the capital expenditures series provides estimates of the expenditures attributable to each SIC division, they are not all surveyed. In these cases, estimates of capital expenditures are produced based on indicators of production, consumption and costs associated with operation in that industry.

Non-surveyed estimates are produced for the forecast and revised forecast capital expenditures in the *agriculture* industry (Division A)<sup>1</sup>. These estimates are based on three different sets of indicators. First, Agriculture Canada's net farm income forecast is used in a simulation model of capital expenditures. Second, forecast and revised forecast estimates of agriculture related machinery and equipment sales, from the Canadian Farm and Industrial Equipment Institute, are used in the estimation of machinery and equipment expenditures. Finally, data from Statistics Canada's Industry Division and International Trade Division are assembled according to the apparent domestic availability method, which has been designed to estimate farm equipment consumption in Canada. These three indicators are then combined to produce a final trend estimate of capital expenditure for SIC major group 01, Agriculture.

The value of capital expenditures in the *fishing* industry (Division B), for all survey periods, is based on the statistical modelling of data obtained from the Department of Fisheries and Oceans Canada and from Industrial Organization and Finance Division of Statistics Canada. Nevertheless, industry group 032, Services to Fishing and industry group 033, Trapping, are not covered by these other sources and are not estimated for in the capital expenditures series.

Estimated changes in capital expenditures in the *construction* industry (Division F) for all survey periods are based on the trend observed in construction activity in the whole economy. The underlying assumption is that the value of new construction work put in place, both in residential and non-residential sectors, is providing a reliable indicator of the demand placed on the construction industry, and therefore of the industries' own investment in capital. However, major group 44, Services to Construction, has not been covered by the capital expenditures survey and is not estimated for in the capital expenditures series.

1. Actual and Preliminary Actual data are collected through Agriculture Division surveys (see section 3.3.3 for more details).

### 3.3.2 Données non enquêtées

Bien que la série des dépenses en immobilisations fournisse une estimation des dépenses attribuables à chaque division de la CTI, toutes ne sont pas enquêtées. Lorsqu'elles ne le sont pas, l'estimation des dépenses est fondée sur les indices de production, de consommation et de coûts associés au secteur en cause.

Des estimations sans recensement sont produites pour les prévisions et les prévisions révisées de dépenses du secteur de l'*agriculture* (division A)<sup>1</sup>. Ces estimations sont fondées sur trois groupes d'indicateurs. Premièrement, les prévisions d'Agriculture Canada sur le revenu agricole net sont utilisées dans un modèle de simulation des dépenses en immobilisations. Deuxièmement, les prévisions et prévisions révisées de l'Institut canadien de l'équipement agricole et industriel au sujet du matériel et de l'outillage agricoles servent à l'estimation des dépenses en matériel et outillage. Enfin, les données de la Division de l'industrie et de la Division du commerce international de Statistique Canada sont colligées selon la méthode de la disponibilité intérieure apparente, conçue pour estimer la consommation d'outillage agricole au Canada. Ces trois indicateurs sont alors combinés pour donner une estimation finale de la tendance des dépenses en immobilisations du grand groupe 01 de la CTI, soit l'agriculture.

La valeur des dépenses en immobilisations du secteur de la *pêche* (division B), pour toutes les périodes d'enquête, est fondée sur la modélisation statistique des données obtenues du ministère des Pêches et Océans du Canada et de la Division de l'organisation et des finances de l'industrie de Statistique Canada. Toutefois, les groupes industriels 032 et 033, soit les services relatifs à la pêche et le piégeage respectivement, ne sont pas couverts par ces sources et ne font pas l'objet d'estimations dans la série des dépenses en immobilisations.

L'estimation des changements dans les dépenses en immobilisations dans le secteur de la *construction* (division F), pour toutes les périodes d'enquête, se base sur la tendance observée dans le domaine du bâtiment pour l'ensemble de l'économie. Cette méthode repose sur l'hypothèse selon laquelle la valeur des mises en chantier, aussi bien dans le secteur résidentiel que dans les secteurs non résidentiels, constitue un indice fiable de la demande dans le domaine du bâtiment, donc des investissements des entreprises qui en font partie. Cependant, le grand groupe 44, soit les services relatifs à la construction, n'est pas visé par l'enquête sur les dépenses en immobilisations et ne fait pas partie des estimations de la série.

1. Les données réelles et réelles préliminaires sont recueillies par des enquêtes de la Division de l'agriculture (voir la section 3.3.3).

In addition, *housing* investment is produced by the Current Investment Indicators Section and is based on projected housing starts, building costs and the value of alterations and improvements in each province. *Residential infrastructure* put in place by developers has been estimated for and the value of that infrastructure which will be turned over to municipalities upon completion has been included in the capital expenditures series under local government investments in capital.

### 3.3.3 Data Collection Arrangements

Within Statistics Canada several divisions participate in the collection of data which are incorporated into the final production of capital expenditure estimates by the Investment and Capital Stock Division. The Agriculture Division collects information on actual and preliminary actual capital expenditures from the Farm Financial Survey and Crop Surveys. The Public Institutions Division expedites the collection process by providing information from its Local Government Capital Expenditure Survey, while Industry Division contributes small establishment data from the Net Cash Expenditures Survey of the oil and gas industry. Housing estimates are produced by the Current Investment Indicators Section (Investment and Capital Stock Division).

Furthermore, the capital expenditures series consolidates data collected by agencies or departments external to Statistics Canada. Data collected by each provincial/territorial statistical focal point related to education (provincial/ territorial schools), health and provincial governments are incorporated into the capital expenditures series. Natural Resources Canada, Energy Policy Sector organizes collection of data from large establishments within the oil and gas industry; this data complements Statistics Canada's Manufacturing, Construction, Energy Division's survey of small establishments operating in the oil and gas industry. Mining industry data are collected at the provincial level by provincial energy, mines and resources departments in Newfoundland, Nova Scotia, Quebec, Ontario, Manitoba and British Columbia.

Finalement, les chiffres sur les investissements dans le domaine du *logement* sont produits par la Section des indicateurs courants de l'investissement et sont fondés sur les mises en chantier projetées, les coûts de construction et la valeur des rénovations dans chaque province. L'*infrastructure résidentielle* mise en place par les promoteurs est estimée et la valeur des infrastructures qui seront remises aux municipalités une fois achevées est comprise dans la série des dépenses en immobilisations sous la rubrique des administrations locales.

### 3.3.3 Mode de collecte des données

Plusieurs divisions de Statistique Canada participent à la collecte des données comprises dans les estimations finales des dépenses en immobilisations produites par la Division de l'investissement et du stock de capital. La Division de l'agriculture recueille des données sur les dépenses réelles et réelles préliminaires en immobilisations à l'aide de l'Enquête financière sur les fermes et de l'Enquête sur les cultures. La Division des institutions publiques accélère le processus en fournissant les données de son Enquête sur les dépenses en immobilisations des administrations locales, tandis que la Division de l'industrie apporte des données sur les petites entreprises en se fondant sur son Enquête sur les dépenses nettes de l'industrie pétrolière et gazière. Les estimations portant sur le logement sont produites par la Section des indicateurs courants de l'investissement (Division de l'investissement et du stock de capital).

De plus, la série des dépenses en immobilisations comprend des données recueillies par des organismes et ministères extérieurs à Statistique Canada. Ainsi, les données rassemblées par chaque coordonnateur statistique provincial ou territorial dans le domaine de l'enseignement (écoles provinciales et territoriales), de la santé et de l'administration publique sont incorporées dans les séries des dépenses en immobilisations. Le Secteur de la politique énergétique de Ressources naturelles Canada organise la collecte des données provenant des grands établissements des secteurs pétroliers et gaziers; ces données viennent compléter l'enquête effectuée par la Division de la fabrication, de la construction et de l'énergie de Statistique Canada auprès des petits établissements actifs dans ces secteurs. Les chiffres portant sur l'industrie minière sont recueillis à l'échelle provinciale par les ministères provinciaux de l'énergie, des mines et des richesses naturelles de Terre-Neuve, de la Nouvelle-Écosse, du Québec, de l'Ontario, du Manitoba et de la Colombie-Britannique.

### 3.3.4 Survey Periods

The three survey periods are organized and timed to collect four sets of annual data related to intentions, revised intentions, preliminary actual and actual capital and repair expenditures for all sectors of the economy (see Table 1).

**Table 1** Capital Expenditures Series Data Collection

Release Date	Data	Collection Period
February (Y)	Intentions (Y)	Nov. (Y-1) to Feb. (Y)
	Preliminary Actual (Y-1)	Nov. (Y-1) to Feb. (Y)
	Actual (Y-2)	March (Y-1) to Oct. (Y-1)
July (Y)	Revised Intentions (Y)	May (Y) to July (Y)

Y = current calendar year

### 3.3.4 Périodes d'enquête

Les trois périodes d'enquête sont organisées de telle sorte qu'on recueille quatre ensembles de données annuelles ayant trait aux perspectives, aux perspectives révisées, aux données réelles préliminaires et aux données réelles relativement aux dépenses en immobilisations et en réparations pour toutes les branches de l'économie (voir tableau 1).

**Tableau 1** Collecte des données pour la série des dépenses en immobilisations

Date de publication	Données	Période de collecte
Février (A)	Perspectives (A)	Nov. (A-1) à févr. (A)
	Réelles préliminaires (A-1)	Nov. (A-1) à févr. (A)
	Réelles (A-2)	Mars (A-1) à oct. (A-1)
Juillet (A)	Perspectives révisées(A)	Mai (A) à juillet (A)

A = année civile courante

## 3.4 Quality Assurance

### 3.4.1 Non-Response Follow-Up

Low response rate to the survey within a specific industry and province/territory represents the primary reason for follow-up. Initially, a general reminder is sent in the form of a mailout to the entire delinquent portion of the sample. If non-response continues, establishments in areas of lowest coverage are solicited by telephone for the return of the completed questionnaire. Actively canvassing sampled non-response establishments increases the response rate and, as a result, estimation for the non-sampled portion of the frame are made more accurate (see 3.5 methodology).

### 3.4.2 Editing

After the questionnaires have been completed and returned, the process of quality assurance continues through data editing. Data are screened at the micro level for internal, survey over survey and year over year inconsistencies.

Add-check edits identify expenditure data that are incorrectly reported in dollars rather than thousands, percentage data failing to add to 100 percent and/or inconsistencies related to the reported totals. Large-difference edits evaluate the consistency of reported expenditures by comparing the current data with reports from a previous survey within the same year and from a different year. On the actual survey for respondents receiving long forms, asset detail edits identify all establishments reporting expenditures on assets or asset details which are inconsistent with previous questionnaire returns or inconsistent with assets commonly used in the respondent's industry. Edit tests will flag reported data for confirmation based on thresholds which are set after evaluating industry coverage and geographic location. In addition, new and large project data are collected from newspapers, trade journals and industry reports. This information is compared to reported data and any inconsistencies are flagged for confirmation.

Once an establishment's reported expenditures data have been flagged by the edit process, additional questionnaire data are consulted for an explanation. For example, the questionnaire section entitled, "Reasons for changes in capital expenditures", contains respondent supplied explanations for changes in capital expenditure. However, if the reason for the inconsistency cannot be ascertained from the questionnaire or other industry information, the reporting establishment is contacted directly for confirmation. Based on this inquiry the data reported are updated to include either new data or an explanation of expenditures.

Other micro data editing may occur for reported Gross Book Value or Capacity Utilization. Gross book value edits occur when the reported gross book value of an

## 3.4 Assurance de la qualité

### 3.4.1 Suivi des questionnaires non retournés

Les mesures de suivi sont le plus souvent motivées par un faible taux de réponse dans un secteur ou dans une province ou un territoire en particulier. Pour commencer, on envoie un rappel général par la poste à toutes les unités n'ayant pas encore répondu. Si la situation ne se règle pas, on téléphone aux établissements faisant partie des secteurs les plus lacunaires pour les inciter à retourner le questionnaire rempli. Le fait d'intervenir activement auprès des établissements enquêtés qui ne répondent pas accroît le taux de réponse et permet ainsi d'obtenir des résultats plus précis pour l'estimation de la partie non enquêtée de la base de sondage (voir 3.5, Méthode).

### 3.4.2 Vérification

Lorsque les questionnaires remplis sont reçus, le processus d'assurance de la qualité se poursuit par la vérification statistique des données, qui consiste à scruter les données pour détecter les incohérences internes et les anomalies par rapport aux autres enquêtes ou aux autres années.

La vérification des additions sert à repérer les chiffres indiqués en unités plutôt qu'en milliers, les pourcentages dont la somme n'arrive pas à 100% et les totaux inexacts. La vérification des différences consiste à évaluer la vraisemblance des dépenses déclarées en les comparant avec les dépenses indiquées lors des enquêtes antérieures de la même année et d'une autre année. Pour l'enquête sur les données réelles auprès des répondants qui reçoivent le questionnaire long, la vérification du détail des biens permet de découvrir les établissements qui présentent des renseignements sur les biens d'une manière qui ne concorde pas avec les déclarations des questionnaires antérieurs ou avec les réalités communes dans l'industrie. Ces vérifications permettent d'identifier les données pour confirmation selon des seuils fixés après évaluation de la portion couverte par l'enquête par rapport à l'industrie et selon l'emplacement géographique. De plus, des données sur les grands projets et sur les projets nouveaux sont recueillies dans les journaux, dans les revues spécialisées et dans les rapports d'industrie. Ces renseignements sont confrontés aux données déclarées et toute incohérence est identifiée pour confirmation.

Lorsque les données d'un établissement ont été identifiées à l'étape de la vérification, on consulte les autres données du questionnaire pour éclaircir le cas. Par exemple, une section du questionnaire demande au répondant d'indiquer les motifs des changements dans ses dépenses en immobilisations. Toutefois, si l'incohérence ne peut être expliquée à partir du questionnaire ni par d'autres renseignements portant sur l'industrie, on appelle directement l'établissement pour demander une confirmation. Les données sont alors mises à jour, soit par une correction, soit par une explication.

D'autres vérifications des microdonnées peuvent être effectuées pour les valeurs comptables brutes ou l'utilisation de la capacité. On fait une vérification des valeurs comptables brutes lorsque la

establishments assets does not coincide with the previously reported gross book value plus current investment in new capital net of disposals. In this case, the establishment is contacted for confirmation of (or an update to) the reported data. Capacity utilization edits identify all those manufacturing and mining establishments operating at less than expected manufacturing or mining capacity. If previous reports are significantly different from the current questionnaire response, the establishment is contacted to confirm or update the reported data.

### 3.4.3 Macro Data Evaluation

After the estimation process (see 4.5.3 estimation), a comprehensive data set exists for the surveyed and non-surveyed portions of the universe (frame) and therefore trend analysis for the various industries can begin. Commencing with an evaluation of the year over year (or percentage) change in each industry, provinces/territories that have industries or sub-industries experiencing unusual activity are highlighted. In addition, this type of analysis also identifies industries which have the largest impact on Canadian aggregates.

Macro analysis continues with the assessment of information which may be effecting the expenditures in a specific province or industry. This additional information might come in the form of economic indicators such as GDP, productivity, capacity utilization, profits or technological innovation. Factors influencing the expenditures might also include government policies (fiscal policy, monetary policy, grants and/or subsidies) or industry specific information such as meters drilled, import/export data or building permits. Although causality is not drawn, the analysis attempts to link information directly and indirectly related to the industry with recent trends in capital expenditures. As a by product of this analysis, those industries experiencing exceptional activity will undergo further micro data evaluation to determine the reason for the large year over year change.

valeur déclarée des immobilisations d'un établissement n'égal pas la valeur comptable brute déclarée antérieurement plus les investissements courants en immobilisations nouvelles déduction faite des aliénations. Dans un cas de ce genre, on appelle l'établissement pour qu'il confirme ou corrige les données. La vérification de l'utilisation de la capacité consiste à repérer tous les établissements manufacturiers et miniers qui fonctionnent à une capacité inférieure aux prévisions. Si les déclarations antérieures présentent des différences considérables par rapport aux données du questionnaire examiné, on appelle l'établissement pour qu'il confirme ou corrige les chiffres.

### 3.4.3 Évaluation des macrodonnées

L'exercice d'estimation (voir 4.5.3, Estimation) permet d'obtenir un ensemble complet de données pour les parties enquêtées et non enquêtées de l'univers (base de sondage), donc il est possible de commencer l'analyse des tendances des divers secteurs. On commence par une évaluation des changements annuels (ou en pourcentage) dans chaque industrie pour repérer les provinces et territoires comportant des industries ou des sous-industries qui présentent une activité inhabituelle. Ce genre d'analyse met aussi en relief les industries qui ont le plus de poids dans les totaux canadiens.

La macroanalyse se poursuit par l'ajout des facteurs pouvant avoir un effet sur les dépenses dans une province ou dans une industrie en particulier. Ces renseignements supplémentaires peuvent prendre la forme d'indices économiques comme le PIB, la productivité, l'utilisation de la capacité, les bénéfices ou l'innovation technologique. D'autres facteurs peuvent influencer sur les dépenses, par exemple les politiques gouvernementales (politique fiscale, politique monétaire, subventions) ou les données propres à une industrie comme le nombre de mètres forés, l'import-export et les permis de construction. Sans établir un lien de cause à effet, cette analyse vise à mettre en rapport les informations ayant directement et indirectement trait à l'industrie avec les tendances récentes en matière de dépenses d'immobilisations. À la suite de cet exercice, les industries qui présentent une activité exceptionnelle sont soumises à des évaluations de microdonnées plus poussées en vue d'expliquer les différences importantes entre les années.

## 3.5 Methodology

### Backcasting of 91 to 99 estimations

This part of the methodology section refers to the data presented in the special article. Hence, section 3.5.0 is special and is only concerned with backcasted NAICS data for reference years 1991 to 1999. The following sections (3.5.1 to 3.5.11) only discuss the methodology used to process the 1999 Revised Intentions data.

With the new industrial classification called NAICS (North American Industrial Classification System), estimates under the old industrial classification called SIC (Standard Industrial Classification) are out-dated. Starting from the 1999 Survey on Revised Intentions, only NAICS based estimates will be published.

Since estimates under NAICS starts in 1999, no previous data are available to help build the usual forecast models or to help calculate growth rates to determine the trend of the data. To fill that need for data and recreate a series of NAICS estimates, two statistical techniques were used. The first one is used to reproduce estimates under NAICS for years 1997 and 1998, as well as for the intentions of 1999. The second technique is used to recreate NAICS estimates for years 1991 to 1996.

What is left of section 3.5.0 gives further detail on each of the techniques. Once the backcasted estimates are available, a calendarization technique is applied to correct the estimates so that they better represent the calendar year. This last technique is explained at the end of the section.

### Backcasting estimates from 1997 to 1999

Since the new classification started in 1997, the universe and the sample used to produce SIC-based estimates were also coded to NAICS. Coded with both classifications, these data were used to estimate for the 1997 Actual Survey, the 1998 Revised Intentions and Preliminary Actual Surveys and the 1999 Intentions Survey. For reference year 1998, only the Preliminary Actual data are re-published.

To produce estimates as reliable as possible under NAICS, four important survey steps were redone for the processing of the data: the outlier detection, imputation, calculation of g-weights and estimation.

For the outlier detection, the usual methodology has been applied (see section 3.5.6). However, the detection groups were built using NAICS groupings (instead of SIC groupings).

Concerning imputation, the only difference compared to the usual methodology is also linked to the creation of the groupings. These follow the NAICS aggregation

## 3.5 Méthodologie

### Reproduction des estimations de 91 à 99

Cette partie de la section de méthodologie fait référence aux données présentées dans l'article spécial. La section 3.5.0 est donc spéciale et traite des données reproduites selon le SCIAN pour les années de références de 1991 à 1999. Les sections suivantes (3.5.1 à 3.5.11) discutent uniquement de la méthodologie employée pour traiter les données de l'enquête sur les perspectives révisées de 1999.

Avec la venue de la nouvelle classification industrielle appelée SCIAN (Système de classification industrielle de l'Amérique du Nord), les estimations selon l'ancienne classification industrielle appelée CTI (Classification type des industries) deviennent désuètes. À partir de l'enquête sur les perspectives révisées de 1999, seulement les estimations selon le SCIAN seront diffusées.

Puisque les estimations selon le SCIAN commencent en 1999, on ne possède pas de données antérieures pour nous aider à bâtir les modèles de prévision habituels ou encore nous permettre de calculer les taux de croissance pour déterminer la tendance inhérente aux données. Pour pallier à ce manque de données et recréer une série d'estimations selon le SCIAN, deux techniques statistiques ont été utilisées. La première nous permet de reproduire des estimations selon le SCIAN pour les années 1997, 1998 ainsi que pour les perspectives de 1999. La seconde technique elle, nous permet de recréer des estimations selon le SCIAN pour les années 1991 à 1996.

Le reste de la section 3.5.0 explique de façon détaillée chacune des techniques. Une fois les estimations produites, une technique de calendrialisation est appliquée pour nous permettre de corriger les estimations de façon à représenter adéquatement l'année civile. Cette dernière méthode est expliquée à la fin de la section.

### Reproduction des estimations de 1997 à 1999

Étant donné que la nouvelle classification est entrée en vigueur en 1997, l'univers ainsi que l'échantillon qui avaient servi à produire des estimations selon la CTI étaient également codés selon le SCIAN. Ces données codées selon les deux classifications ont donc pu servir à estimer les enquêtes des données réelles de 1997, les perspectives révisées ainsi que les données réelles provisoires de 1998 en plus des perspectives de 1999. Pour l'année 1998, seulement les données réelles provisoires sont republiées.

Pour produire les estimations les plus fiables possibles selon le SCIAN, on a refait quatre étapes importantes du traitement des données : la détection des données aberrantes, l'imputation, le calcul des poids de calage et le calcul des estimations.

Pour la détection des données aberrantes, la méthodologie habituelle a été appliquée (voir section 3.5.6). Cependant, les groupes de détection ont été construits en regroupant les industries selon le SCIAN (au lieu de la CTI).

En ce qui a trait à l'imputation, le seul changement par rapport à la méthodologie habituelle est également relié aux regroupements. Ceux-ci suivent des niveaux d'agrégats selon le SCIAN et sont par

levels which make them more appropriate. For more details on imputation, consult section 3.5.7.

The calculation of the g-weights was done using NAICS groups instead of SIC groups. This adjustment to the sampling weight compensates for the stratification developed under SIC. This example reflects the benefit of re-calculating the g-weights.

Suppose there is a take-some SIC stratum in a given province. There are 10 units selected from a population of 100. When we look closely at the distribution of these units in terms of NAICS, we find that only 2 cover one sector while the 8 others cover a second sector. However, in the population these are split equally between both NAICS sectors.

SIC: 10 sampled out of 100 in the population  
NAICS A: 2 out of 50      NAICS B: 8 out of 50

The original sampling weight for each of the selected units is 10. Instead of using directly the original weight, we apply a correction using the g-weight in such a way that the final weight is greater for the 2 units in NAICS A and smaller for each of the 8 units in NAICS B.

This example shows that even though the stratification was not done using NAICS, we can still make sure that the NAICS universe is well represented by the final weights. In the case of the specified surveys, we re-calculated the g-weights using groups at the Canada level with the variable revenue known for both the sample and the population.

Since a new up to date and more complete universe was available when recalculating the weights, its revenue were used. In the preceding example, the new distribution and an increase in the universe's size could change the totals of 50 and 50 to values like 40 and 75.

Finally, after the weights have been recalculated, the domain estimates can be calculated for the NAICS publication. Domain estimation is used so that only NAICS with units in that NAICS domain are represented. If a particular NAICS domain does not have any units selected in the sample, no estimate is produced. Once the estimates are backcasted, they can be analyzed for publication.

### Backcasting estimates from 1991 to 1996

To backcast estimates from 1991 to 1996, only the control totals of these years were used (no micro-data). For years 1997, 1998 and 1999, estimates under NAICS were backcasted. Since these micro-data are still coded under SIC we can produce estimates with both classifications and find the relationship between the SIC and NAICS classifications. Also, we have micro-data from four survey occasions (one in 1997, two in 1998 and one in 1999) from which to calculate the relationship.

To make use of these micro-data from the four surveys, we combined them as if they were independent and

le fait même plus appropriés. Pour plus de détails au sujet de l'imputation, il faut consulter la section 3.5.7.

Le calcul des poids de calage s'est effectué à l'aide de groupes selon le SCIAN au lieu de groupes selon la CTI. Cet ajustement au poids d'échantillonnage permet de compenser pour la stratification effectuée en termes de CTI. Un exemple exprime mieux le bienfait du recalcul du poids de calage.

Supposons une strate à tirage partiel en termes de CTI pour une province donnée. Il y a 10 unités qui sont sélectionnées parmi 100 dans la population. Lorsque l'on regarde attentivement la distribution de ces unités en termes de SCIAN, on se rend compte que seulement 2 couvrent un secteur et que les 8 autres couvrent un deuxième secteur SCIAN même si la population se divise également entre les deux SCIAN.

CTI : 10 sélectionnés parmi 100 dans la population  
SCIAN A : 2 parmi 50      SCIAN B : 8 parmi 50

Le poids original d'échantillonnage pour chacune des 10 unités est de 10. Au lieu d'utiliser directement ce poids, celui-ci est corrigé par le poids de calage de façon à obtenir un poids final plus grand pour les deux unités du SCIAN A et plus petit pour chacune des 8 unités du SCIAN B.

Cet exemple montre que même si la stratification n'a pas tenu compte du SCIAN, on peut quand même s'assurer que l'univers SCIAN soit bien représenté dans le calcul des poids finaux. Dans le cas des enquêtes spécifiées, on a procédé au calage selon des groupes au niveau Canada à l'aide de la variable revenu connue pour l'échantillon et la population.

Puisqu'au moment de recalculer les poids un nouvel univers à jour et plus complet était disponible, les comptes de ceux-ci ont été utilisés. Dans l'exemple précédent, la nouvelle distribution et l'augmentation de l'univers pourraient changer les totaux de 50 et 50 pour des valeurs comme, par exemple, 40 et 75.

Finalement, après avoir recalculé les poids, on calcule les estimations pour les domaines de publication selon le SCIAN. L'estimation par domaine est utilisée de telle sorte qu'un total est calculé seulement sur les domaines présents selon le SCIAN. Si un domaine SCIAN particulier n'a malheureusement aucune unité sélectionnée dans l'échantillon, aucune estimation n'est produite. Une fois les estimations reproduites, elles peuvent être analysées pour fins de publication.

### Reproduction des estimations de 1991 à 1996

Pour reproduire les estimations des années 1991 à 1996, seulement les totaux de contrôle de ces années ont été utilisés (aucune micro donnée). Pour les années 1997, 1998 et 1999, on a réussi à reproduire les estimations selon le SCIAN. Or, ces micro données sont encore codées selon la CTI. On peut donc produire les estimations selon les deux classifications et trouver le lien entre la classification CTI et celle du SCIAN. De plus, on a quatre jeux de micro données (une enquête en 1997, deux en 1998 et une en 1999) avec lesquels on peut calculer ce lien.

Pour tirer profit des micro données des quatre enquêtes, on a combiné les quatre échantillons comme s'ils étaient indépendants



calculated a concordance factor on the four surveys simultaneously. To give more weight to the surveys with actual data (Actual 1997 and Preliminary Actual 1998) than the one with intentions data (Revised Intentions 1998 and Intentions 1999), a 33% weight was given to the two first and a 17% weight given to the last two when calculating a concordance factor.

Once the data were ready to be processed, the correspondence factors were calculated. First, let's define the working cell to be the cross-product of province/territory and the most disaggregated publication level. The working cell is defined under both classifications, SIC and NAICS. From the micro-data we can now calculate for an SIC's working cell the proportion that goes into each of the related NAICS' working cells.

Once the proportions are calculated, they can be applied to the known SIC estimates from 1991 to 1996. Then, we can sum over the NAICS publication cells to produce the backcasted estimates.

### Calendarization of the backcasted data

Since the results must follow closely the calendar year they refer to (and not the period covered by all fiscal periods of the respondents), the NAICS backcasted data need to be calendarized. In the case where the micro-data are available i.e. for the 1997, 1998 and 1999 surveys, the methodology is described in section 3.5.5.

Since this methodology is based on micro-data, it cannot be applied to the 1991 to 1996 results (these being based on previously published data). However, using micro-data from the 1997, 1998 and 1999 where we have micro-data fiscal and calendarized under NAICS from four survey occasions, we can estimate what should be the calendarized values for the 1991 to 1999 backcasted NAICS estimates.

For each of the 1997, 1998 and 1999 set of calendarized micro-data, we first calculate the portion of investment that is part of the reference year, the portion transferred to the previous reference year and the portion transferred to the following year. For example, if for the 1998 reference year a respondent has a fiscal period from June 1<sup>st</sup> 1997 to May 31<sup>st</sup> 1998, there is a portion of its investments that is part of the reference year (January to May 1998) and a portion that is part of the previous reference year (June to December 1997).

Once all the portions have been calculated, they can be summed at the NAICS' working cell level for the three different years. Once again, this is done for the four surveys where micro-data is available and the totals follow the distribution 33%, 33%, 17% and 17% depending if the estimates come from actual or intentions data.

We are now ready to recalculate from the 1991 to 1996 backcasted NAICS totals what proportion moves to the

et calculé un facteur de concordance sur l'ensemble des quatre enquêtes. Pour donner plus de poids aux enquêtes avec données réelles (données réelles de 97 et données réelles provisoires de 98) par rapport aux enquêtes sur les perspectives (perspectives révisées de 98 et perspectives de 99), un poids de 33% a été accordé aux deux premières alors qu'un poids de 17% a été accordé aux deux dernières dans le calcul du paramètre de concordance.

Une fois les données prêtes à être traitées, les facteurs de concordance ont été calculés. On définit tout d'abord la cellule de travail comme le croisement entre la province/territoire et le niveau de publication industriel le plus désagrégé. Cette cellule est définie selon les deux classifications, CTI et SCIAN. À partir des micro données on peut maintenant calculer pour une cellule de travail en CTI quelle proportion va dans chaque cellule de travail selon le SCIAN.

Une fois les proportions calculées, on peut les appliquer sur les estimations connues de 1991 à 1996 selon la CTI. Puis on n'a qu'à faire la sommation sur les cellules de publication SCIAN pour reproduire les estimations selon le SCIAN.

### Calendrialisation des données reproduites

Puisque les résultats doivent représenter fidèlement l'année civile (et non pas l'ensemble des périodes fiscales des répondants), les données reproduites selon le SCIAN doivent être calendrialisées. Dans le cas où les micro données sont disponibles, i.e. les enquêtes de 1997, 1998 et 1999, on utilise la méthodologie décrite dans la section 3.5.5.

Puisque cette méthodologie est basée sur les micro données, on ne peut l'appliquer aux résultats de 1991 à 1996 (ceux-ci étant uniquement basés sur les données précédemment publiées). Toutefois, en utilisant les micro données de 1997, 1998 et 1999 où, pour quatre enquêtes on possède les micro données fiscales et calendrialisées selon le SCIAN, on peut estimer quelles seraient les valeurs calendrialisées pour les données de 1991 à 1996, reproduites selon le SCIAN.

Pour chacun des ensembles de micro données calendrialisées de 1997, 1998 et 1999, on calcule tout d'abord les investissements qui font partie de l'année de référence, ceux qui sont transférés dans l'année de référence précédente et ceux qui sont transférés dans l'année de référence suivante. Par exemple, si pour l'année de référence 1998 un répondant a une période fiscale du 1<sup>er</sup> juin 1997 au 31 mai 1998, il y a une portion de ses investissements qui se trouve dans l'année de référence (janvier à mai 1998) et une portion qui se trouve dans l'année de référence précédente (juin à décembre 1997).

Une fois toutes les portions calculées, on somme au niveau des cellules de travail selon le SCIAN sur les trois différentes années. Encore une fois, ceci est fait pour les quatre enquêtes pour lesquelles des micro données sont disponibles et les totaux suivent la distribution 33%, 33%, 17% et 17% selon que les estimations proviennent de données réelles ou de perspectives.

On est maintenant prêt à recalculer à partir des totaux de 1991 à 1996 reproduits selon le SCIAN, quelle proportion passe dans

next year, the previous year or stays in the current year. Once distributed, data are summed over the calendar years to produce the NAICS backcasted calendarized estimates.

Special attention was given to the boundary years namely 1991 and 1996. In fact, between 1996 and 1997, the methodology for the calculation of the calendarized estimates changes. The calendarized total from 1996 comes from three sources: what is going from 1995 to 1996 (following the calculated proportion), what stays in 1996 (following the calculated proportion) and what is coming from 1997 back to 1996 (real number based on true micro-data from 1997). In the case of 1991, to estimate what proportion goes from 1990 to 1991, we simply repeated the value going from 1991 to 1992.

The rest of this text only concerns the methodology of the 1999 Revised Intentions Survey.

### 3.5.1 Introduction

The Capital Expenditures Survey (CES) produces data on investment made in Canada, in all types of Canadian industries. These data are gathered three times a year, at three very specific times. This permits follow-up on intentions and achievements in terms of investment, on an annual basis. For fiscal year Y, for example, we have four sources of data: the Survey on Intentions (SI) for which the questionnaire is mailed out in October of year Y-1, the Survey on Revised Intentions (SRI) for which the questionnaire is mailed out in March of year Y, the Survey on Preliminary Actual Data (SPA) for which the questionnaire is mailed out in October of year Y and finally, the Survey on Actual Data (SA) for which the questionnaire is mailed out in March of year Y+1. The data for the four surveys are gathered at three specific moments simply because the SI and SPA questionnaires are combined in a single questionnaire and thus gather these data at the same time.

The Investment and Capital Stock Division (ICSD) produces two annual publications that assemble the results of the aforementioned surveys. In February of year Y, the results of the SI (year Y), SPA (Y-1) and SA (Y-2) surveys are published. In July of year Y, the results of the SRI (Y) survey are published.

The sample for reference year Y is taken in November of year Y-1 and covers the SI (Y+1), SRI (Y), SPA (Y) and SA (Y-1) surveys. It should be noted here that the same sample covers three fiscal years, and conversely, that a single fiscal year is covered by three distinct samples. The following diagram illustrates the various relationships between the sample selection, the surveys covered by the sample as well as the time line from mail-out to publication of the estimates.

l'année suivante, l'année précédente et l'année en cours. Une fois les données distribuées, on somme sur les années civiles pour trouver les nouvelles estimations calendrialisées et reproduites selon le SCIAN.

Une attention spéciale a été portée aux années frontières 1991 et 1996. En effet, entre 1996 et 1997, la méthodologie du calcul des estimations calendrialisées change. Le total calendrialisé de 1996 provient de trois sources : ce qui sort de 1995 vers 1996 (selon la proportion calculée), ce qui reste en 1996 (selon la proportion calculée), et ce qui sort de 1997 vers 1996 (vrai chiffre basé sur les micro données de 1997). Dans le cas de 1991, pour connaître ce qui provient de 1990 en allant vers 1991, on a tout simplement supposé que c'était équivalent à ce qui sortait de 1991 vers 1992.

Le reste de ce texte ne concerne que la méthodologie reliée à l'enquête sur les perspectives révisées de 1999.

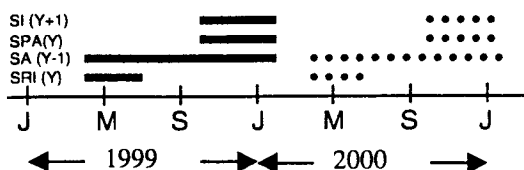
### 3.5.1 Introduction

L'enquête sur les dépenses en immobilisations (EDI) produit des données sur les investissements qui se font au Canada et dans tous les types d'industries canadiennes. Ces données sont recueillies trois fois par année, à trois moments bien précis. Ceci permet d'avoir un suivi sur les intentions et les réalisations en termes d'investissements et ce, sur une base annuelle. On a, par exemple, pour l'année financière A, quatre sources de données: l'enquête sur les perspectives (EP) dont le questionnaire est envoyé au mois d'octobre de l'année A-1, l'enquête sur les perspectives révisées (EPR) dont le questionnaire est envoyé au mois de mars de l'année A, l'enquête sur les données réelles provisoires (ERP) dont le questionnaire est envoyé au mois d'octobre de l'année A et finalement, l'enquête sur les données réelles (ER) dont le questionnaire est envoyé au mois de mars de l'année A+1. Si les données des quatre enquêtes sont recueillies à trois moments précis, c'est tout simplement parce que les questionnaires de l'EP et de l'ERP sont combinés dans un même questionnaire et du fait même, les données récoltées en même temps.

La Division des investissements et stocks de capitaux (DISC) produit annuellement deux publications qui regroupent les résultats des enquêtes susmentionnées. En février de l'année A, on publie les résultats des enquêtes EP(année A), ERP(A-1) et ER(A-2). Au mois de juillet de l'année A, on publie les résultats de l'enquête EPR(A).

L'échantillon pour l'année de référence A est tiré au mois de novembre de l'année A-1 et couvre les enquêtes: EP(A+1), EPR(A), ERP(A) et ER(A-1). On note ici qu'un même échantillon couvre trois années financières et, vice versa, qu'une même année financière est couverte par trois échantillons distincts. Voici un schéma mettant en lumière les différentes relations entre le tirage de l'échantillon, les enquêtes couvertes par celui-ci ainsi que la période entre l'envoi postal et la publication des estimations.

Diagram 1: Description of Survey Activities



In diagram 1, the solid lines indicate the data related to the 1999 sample and the dash lines those from the future 2000 year sample. The months indicated are January (J), May (M) and September (S). The mail-out is done at the end of March or in mid-October while the publications are in July and February.

In the SI, SRI and SPA surveys, the variables of interest are capital expenditures on new construction (CC) and capital expenditures on new machinery and new equipment (CM). In the SA survey, we add repair expenditures on construction (RC) as well as repair expenditures on machinery and equipment (RM). In addition, the SA survey produces more detailed estimates for new capital. In fact, capital expenditures by type of assets are also available in the publication "Capital Expenditures by Type of Asset".

## Methodology by Industrial Sector

As in any survey covering several industrial sectors, the methodology for the CES survey differs from one sector to another and thus requires very detailed explanations that are impossible to cover in one section. The following is how the methodology for the various industrial sectors is divided under the North American Industrial Classification System (NAICS):

### 1. Sector 11 Sub-sector 111 and 112 (Crop and Animal Production Industries):

The survey is conducted by Agriculture Division (AD) which adds investment questions to some of their surveys of farmers. The data are processed by AD and the estimates are re-integrated into the bi-annual publication. Refer to section 3.3.2 for more details.

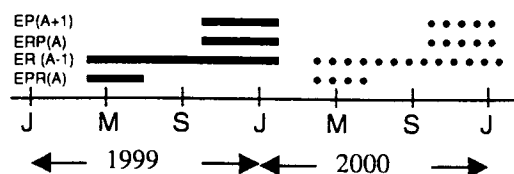
### 2. Sectors 11 Sub-sector 114 (Fishing, Hunting and Trapping Industry) and 23 (Construction Industry):

There is no survey. The data published are based on economic indicators. For more details, refer to section 3.3.2.

### 3. Sector 91 Sub-sector 913 (Local Governments):

The survey is conducted by Public Institutions Division (PID) which uses this opportunity to request the distribution of investment expenditures by function for their own publication "Public Sector Finance". The data,

Schéma 1 : Description des activités de l'enquête



Dans le schéma 1, les lignes pleines correspondent aux données reliées à l'échantillon 1999 et les lignes pointillées à celles de l'échantillon de l'année 2000. Les mois représentés sont janvier (J), mai (M) et septembre (S). Les envois postaux se font à la fin de mars ou à la mi-octobre alors que les publications sortent en juillet et en février.

Lors des enquêtes EP, EPR et ERP, les variables d'intérêt sont les capitaux immobilisés pour de la nouvelle construction (CC) ainsi que les capitaux immobilisés pour de la nouvelle machinerie et du nouvel équipement (CM). Lors de l'enquête ER, on ajoute les capitaux immobilisés pour la réparation reliée à la construction (RC) et les capitaux immobilisés pour la réparation reliée à la machinerie et à l'équipement (RM). De plus, lors de l'enquête ER, on produit des estimations plus détaillées pour les nouveaux capitaux. En effet, des estimations par type d'actifs sont également disponibles dans la publication "Dépenses en immobilisations par type d'actif".

## Méthodologie par secteur industriel

Comme dans toute enquête qui couvre plusieurs secteurs industriels, la méthodologie de l'EDI diffère d'un secteur à un autre et devient par conséquent très fastidieuse à expliquer à l'intérieur d'une même section. Voici comment se scinde la méthodologie pour les différents secteurs industriels selon le Système de classification industrielle de l'Amérique du Nord (SCIAN) :

### 1. Secteur 11 sous-secteurs 111 et 112 (Industrie de la production animale et agricole):

L'enquête est menée par la Division de l'agriculture (DA) qui ajoute les questions d'investissements à certaines de leurs enquêtes auprès des fermiers. Les données sont traitées par la DA et les estimations sont rapatriées dans la publication bi-annuelle. Consultez la section 3.3.2 pour plus de détails.

### 2. Secteurs 11 sous-secteur 114 (Industrie de la pêche, de la chasse et du piégeage) et 23 (Industrie de la construction):

Il n'y a pas d'enquête. Les données publiées sont basées sur des indicateurs économiques. Consultez la section 3.3.2 pour plus de détails.

### 3. Secteur 91 sous-secteur 913 (Administrations locales):

L'enquête est menée par la Division des institutions publiques (DIP) qui en profite pour demander la distribution des dépenses en investissements par fonction qui servira pour leur propre publication "Finance du secteur public". Les données sont

however, are processed by ICSD and usually are in the same format as most of the data gathered by ICSD. For more details on the sampling methodology, see Pandher (1995). It should be noted that in the case of Quebec, a special arrangement provides investment values for the province.

**4. Sectors 21 Sub-sectors 211 (Crude Petroleum and Natural Gas) and 212 (Mining) and 91 Sub-sectors 911, 912 and 914 (Federal Government, Provincial and Territorial Governments and Aboriginal Government):**

A sample using a model based methodology has been preserved. The treatment is the same for the remainder of the samples with only a few exceptions. For more details, see Lacroix (1991).

**5. Sector 21 Canadian Industry 213119 (Other support activities for mining), Sector 55 Canadian Industry 551114 (Head-office), and Sector 81 Sub-sector 814 (Private households):**

There are no surveys and no estimates for this sample.

#### **6. Other Industrial Sectors:**

The methodology used will be described in this section, in particular a model-assisted estimation method.

In fact, the next sections discuss primarily the methodology used for sampling in point 6 (other sectors), and for calendarization, imputation and estimation in points 4 (sectors 21 and 91) and 6. The information on the methodology of the industrial sectors other than that described in point 6 is available in the reference documents cited.

### **3.5.2 Survey Frame**

The frame consists primarily of the Business Register (BR) developed by Statistics Canada. Business Register Division (BRD) is responsible for maintenance and updating of the register. The register is used by a large number of surveys that in turn provide it with feedback to ensure that the latest changes in the business world are incorporated into the BR as quickly as possible.

The BR contains the units required to establish our final survey frame. They are arranged hierarchically as follows: Enterprise - Company - Establishment - Location. An enterprise may comprise several companies, each of which may have several establishments that in turn may operate in several locations. This so-called "statistical" structure is in fact a model of the operational structure described by the enterprise itself. Based on the information available for each level of the operational structure, we define the corresponding statistical structure. For example, to be considered an establishment, a respondent must be able to supply the BR with the wages and rates of pay, income and major inputs in the operational process.

cependant traitées par la DISC et sont habituellement dans le même format que la plupart des données recueillies par la DISC. Pour plus de détails quant à la méthodologie d'échantillonnage, consultez Pandher (1995). Notons que dans le cas du Québec, un arrangement spécial permet d'obtenir les valeurs d'investissements pour la province.

**4. Secteurs 21 sous-secteurs 211 (Pétrole brut et gaz naturel) et 212 (Mines) et 91 sous-secteurs 911, 912 et 914 (Gouvernements fédéraux, provinciaux et territoriaux et autochtones):**

Un échantillon suivant une méthodologie complètement basée sur un modèle a été préservé. Le traitement est le même que pour le reste de l'échantillon à quelques exceptions près. Pour plus de détails, consultez Lacroix (1991).

**5. Secteur 21 industrie canadienne 213119 (Services reliés aux mines), Secteur 55 industrie canadienne 551114 (Bureaux-chefs) et Secteur 81 sous-secteur 814 (Ménages privés):**

Il n'y a pas d'enquête ni aucune estimation pour cet échantillon.

#### **6. Autres secteurs industriels:**

On utilise la méthodologie décrite dans cette section, en particulier une méthode d'estimation assistée d'un modèle.

Les sections suivantes discutent donc principalement de la méthodologie utilisée pour l'échantillonnage du point 6 (autres secteurs), ainsi que pour la calendrialisation, l'imputation et l'estimation des points 4 (secteurs 21 et 91) et 6. L'information concernant la méthodologie des secteurs industriels autres que ceux décrits au point 6, est disponible dans les documents de référence cités.

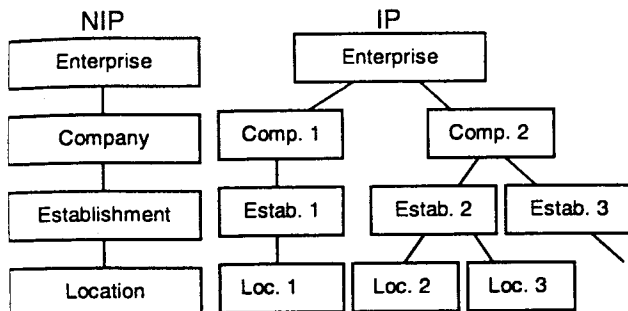
### **3.5.2 Base de sondage**

La base de sondage est principalement formée du Registre des entreprises (RE) développé à Statistique Canada. La Division du registre des entreprises (DRE) est chargée d'en faire l'entretien et la mise à jour. Le registre est utilisé par un grand nombre d'enquêtes qui ne manquent pas de lui retourner de la rétroaction pour s'assurer que les plus récents changements dans le monde des entreprises soient incorporés au RE dans les plus brefs délais.

On retrouve sur le RE les unités nécessaires à l'établissement de notre base de sondage finale. La hiérarchie s'y lit comme suit : Entreprise - Compagnie - Établissement - Emplacement. Une entreprise peut avoir plusieurs compagnies qui peuvent chacune avoir plusieurs établissements qui peuvent à leur tour avoir plusieurs emplacements. Cette structure dite «statistique» est en fait une modélisation de la structure opérationnelle décrite par l'entreprise elle-même. Selon l'information disponible pour chaque niveau de la structure opérationnelle, on définit le niveau statistique correspondant. Par exemple, pour être considéré comme un établissement, on doit pouvoir fournir au RE les salaires et taux de rémunération, le revenu et les intrants principaux dans le processus d'exploitation.

For these units that are part of the non-integrated portion (NIP) of the BR, the statistical structure is linear: an enterprise is related to a single company, a single establishment and a single location. In the integrated portion (IP), the structure may be linear but usually is more complex. Diagram 2 illustrates both structures.

Diagram 2: Statistical Structures



The sampling unit selected for the Capital Expenditure Survey is the establishment, which best corresponds to the gathering and disclosure of investment data. For more details on the BR, refer to Cuthill (1996).

When the sample is drawn in November, a new "image" is taken from the BR. With the new Unified Enterprise Survey, the BR has improved its coverage therefore the "image" is now more complete and up to date. Since the Capital Expenditures Survey is part of the unified survey, it uses this new image for the purpose of sampling.

Since the questionnaires are mailed out in the following March and October, and given the dynamic nature of businesses, we can be certain that new projects will start up after the sample is selected. To be sure that major investments are not "overlooked", units are added to the sample even after the first mailing when the project is deemed important enough. These "new projects", as they are called, are found from newspapers, company reports or lists of building permits. These are sampled with certainty and allow us to avoid gross under-estimation of the value of investment in their industries.

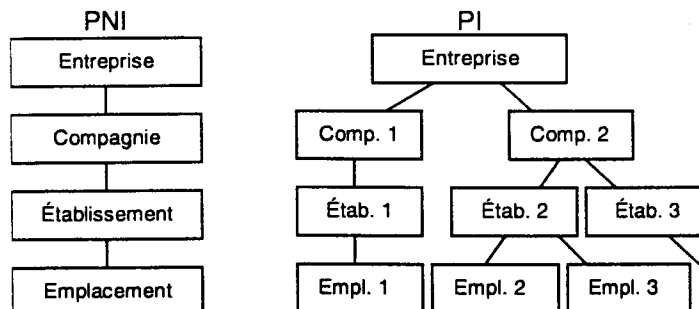
It should be noted that certain units, such as new projects, which we want to have in the sample have incomplete information. Income, which is known for all units on the frame, may be unknown for these units. Since income is used in a range of processes (imputation, estimation, etc.), these units are grouped together to be dealt with separately during processing.

## Grouping

Before sampling begins, all units from the private sector not in the mining and manufacturing industries are grouped together using the following method. All establishments operating in the same province, in the same six-digit-code industrial sector and under the same enterprise have been grouped together in a

Dans le cas des unités formant la portion non-intégrée (PNI) du RE, la structure statistique est linéaire: une entreprise est reliée à une seule compagnie, à un seul établissement et à un seul emplacement. Dans la portion intégrée (PI), la structure peut être linéaire mais est généralement plus complexe. Le schéma 2 illustre les deux structures.

Schéma 2: Structures statistiques



Dans le cadre de l'EDI, l'unité d'échantillonnage choisie est l'établissement, celle-ci correspondant le mieux au besoin de collecte et de divulgation des données d'investissements. Pour plus de détails concernant le RE, consultez Cuthill (1996).

Lorsque l'on tire l'échantillon au mois de novembre, on prend une nouvelle "image" du RE. Avec l'avènement de l'enquête unifiée sur les entreprises, le RE a amélioré sa couverture et cette "image" s'en trouve d'autant plus complète et à jour. L'enquête sur les dépenses en immobilisations fait partie de l'enquête unifiée et se sert de cette nouvelle image pour procéder à l'échantillonnage.

Puisque les questionnaires sont envoyés au mois de mars et au mois d'octobre suivants, et étant donné la nature dynamique des entreprises, il est certain que de nouveaux projets se mettront en branle après la sélection originale de l'échantillon. Pour être certain que d'importants investissements ne soient pas "oubliés", on ajoute des unités à l'échantillon, même après le premier envoi postal, lorsque le projet est jugé assez important. Ces «nouveaux projets» comme nous les appelons, sont trouvés grâce à la lecture de journaux, de rapports de compagnie ou encore grâce aux listes de permis de construction. Ils sont échantillonnés avec certitude et nous permettent d'éviter de sous-estimer grossièrement la valeur des investissements des industries où on les retrouve.

Il est à noter que certaines unités qu'on tient à mettre dans l'échantillon, tels que les nouveaux projets, ont une information incomplète. Le revenu, qui est connu pour toutes les unités de la base, peut être inconnu pour ces unités. Comme le revenu est utilisé au cours de différents processus (imputation, estimation, etc.), ces unités sont regroupées afin de les traiter séparément lors de l'exécution.

## Regroupement

Avant de procéder à l'échantillonnage, toutes les unités dans le secteur privé qui ne faisaient pas partie des industries minières et manufacturières ont été regroupées selon la méthode suivante. Tous les établissements opérant dans la même province, dans le même secteur industriel codé à six chiffres et sous la même entreprise ont été regroupés en un seul super-établissement. Le

single super-establishment. The income of the super-establishment is the sum of all income for the establishments that comprise it, while the remaining information is taken from the head of the group, either the head office where possible, or the establishment with the highest income, where applicable. For the public sector, all the units are in the sample

Once the new universe is constructed with the new super-establishments, all units with income of less than a certain limit are eliminated from the frame unless they constitute head offices or laboratories, in which case the units are chosen with certainty. This procedure is instituted to avoid "losing" these units, which generate practically no income, but might account for substantial investment.

The limit that delineates the out-of-scope units is determined as a function of province and industry. It varies from \$100,000 to \$250,000 depending on the size of the units in the grouping. The limit is calculated in such a way that a maximum of 5% of the total revenue in the group becomes out-of-scope. When all groups have been assembled and the out of scope units have been eliminated, the universe is ready for stratification.

### 3.5.3 Sampling

The sampling is divided into the three traditional parts: stratification, allocation and selection. These are described in the following text.

#### Stratification

The sample must first be stratified by geographic location and industrial classification. The geographic division is based on the 13 provinces and territories, with no other refinement (no infra-provincial stratification). For the industrial stratification, the 1997 NAICS is used at the level required for estimation purposes. If, for example, for a certain industry, the most disaggregated level published corresponds to the 3-digit NAICS, this will be the stratification level. It should be noted that for the remainder of the section, the 6-digit NAICS will be abbreviated as NAICS-6, the 5-digit NAICS as NAICS-5, and so forth.

Table 1 shows, by industry, the most disaggregated publication levels for provincial and Canadian estimates.

**Table 1**  
**Most Disaggregated Publication Levels**

Industry Sector	NAICS Publication Level
11 - Agriculture, Forestry, Fishing and Hunting	3
21 - Mining and Oil and Gas Extraction (NAICS-3 213)	3
21 - Mining and Oil and Gas Extraction (NAICS-5 21231 and 21232)	5
21 - Mining and Oil and Gas Extraction (Other NAICS)	6
22 - Utilities	4

revenu du super-établissement est la somme de tous les revenus des établissements qui le composent et le reste de l'information est tiré de la tête du regroupement, soit le bureau-chef si c'est possible ou, sinon, l'établissement avec le plus grand revenu. Pour le secteur public, toutes les unités font partie de l'échantillon.

Une fois le nouvel univers construit avec les nouveaux super-établissements, toutes les unités qui ont un revenu inférieur à une certaine limite sont éliminées de la base à moins qu'elles ne correspondent à des bureaux-chefs ou à des laboratoires, dans quels cas, les unités sont choisies avec certitude. Cette procédure est mise en place pour éviter de "perdre" ces unités qui ne génèrent pratiquement aucun revenu, mais qui pourraient être l'objet d'investissements substantiels.

La limite inférieure déterminant les unités dans le champs de l'enquête est construite en fonction de la province et du secteur industriel. Celle-ci varie de 100 000\$ à 250 000\$ dépendamment de la taille des unités qui composent l'industrie. En gros, la limite est calculée de telle sorte qu'un maximum de 5% du revenu total du secteur industriel devient hors champs. Lorsque tous les regroupements ont été effectués et que les unités hors champs ont été éliminées, l'univers est prêt à être stratifié.

### 3.5.3 Échantillonnage

L'échantillonnage se divise selon les trois parties traditionnelles: la stratification, la répartition et la sélection. Celles-ci sont décrites dans le texte qui suit.

#### Stratification

On doit tout d'abord stratifier selon le lieu géographique et la classification industrielle. La division géographique se fait selon les 13 provinces et territoires, sans autre raffinement (aucune stratification infra-provinciale). Pour ce qui est de la stratification industrielle, le SCIAN de 1997 est utilisé selon le niveau requis pour les estimations. Si par exemple, pour une certaine industrie, le niveau le plus désagrégé publié correspond au SCIAN à 3 chiffres, ce sera le niveau de stratification. Notons que pour le reste de la section, le SCIAN à 6 chiffres sera abrégé par SCIAN-6, le SCIAN à 5 chiffres par SCIAN-5, etc...

Le tableau 1 indique, par industrie, quels sont les niveaux de publication les plus désagrégés pour les estimations provinciales et canadiennes.

**Tableau 1**  
**Niveaux de publication les plus désagrégés**

Secteur industriel	Niveau de publication SCIAN
11 - Agriculture, foresterie, pêche et chasse	3
21 - Extraction minière, de pétrole et gaz (SCIAN-3 213)	3
21 - Extraction minière, de pétrole et gaz (SCIAN -5 21231 et 21232)	5
21 - Extraction minière, de pétrole et gaz (autres SCIAN)	6
22 - Services publics	4

31-33 Manufacturing (NAICS-3 316 and 323)	3
31-33 Manufacturing (NAICS-4 3121 and NAICS-3 322, 324 and 326)	5
31-33 Manufacturing (Other NAICS)	4
41 - Wholesale Trade	3
44-45 - Retail Trade	3
48-49 - Transportation and Warehousing	3
51 - Information and Cultural Industries	3
52 - Finance and Insurance	3
53 - Real Estate and Rental and Leasing	4
54 - Professional, Scientific and Technical Services	4
55 - Management of Companies and Enterprises	2
56 - Administration and Support, Waste Management and Remediation Services	3
61 - Education Services	4
62 - Health Care and Social Assistance	3
71 - Arts, Entertainment and Recreation	3
72 - Accommodations and Food Services	3
81 - Other Services	3
91 - Public Administration	3

31-33 Fabrication (SCIAN -3 316 et 323)	3
31-33 Fabrication (SCIAN -4 3121 et SCIAN -3 322, 324 et 326)	5
31-33 Fabrication (autres SCIAN)	4
41 - Commerce de gros	3
44-45 - Commerce de détail	3
48-49 - Transport et entreposage	3
51 - Information et culture	3
52 - Finance et assurances	3
53 - Services immobiliers, de location et de location à bail	4
54 - Services professionnels, scientifiques et techniques	4
55 - Gestion de sociétés et d'entreprises	2
56 - Services administratifs, de soutien, de gestion des déchets et d'assainissement	3
61 - Services d'enseignement	4
62 - Soins de santé et assistance sociale	3
71 - Arts, spectacles et loisirs	3
72 - Hébergement et services de restauration	3
81 - Autres services	3
91 - Administrations publiques	3

All provincial publication levels are at the sector level except for the Manufacturing industry where it is at the NAICS-3 level for four provinces: Québec, Ontario, Alberta and British Columbia.

Les niveaux de publication provinciaux correspondent aux secteurs industriels à l'exception de l'industrie de la fabrication où on publie au niveau SCIAN-3 pour quatre provinces: Québec, Ontario, Alberta et Colombie-Britannique

**Allocation**

Once the initial stratification has been introduced, we compute the coefficient of variation (CV) (see section 3.5.8 for more information on CV) to be targeted using the revenue variable to reach the CV set for the most disaggregated publication level. An example helps to better define the situation.

**Répartition**

Une fois la stratification initiale mise en place, on calcule le coefficient de variation (CV) (voir la section 3.5.8 pour plus d'information sur les CV) à viser en utilisant la variable revenu de façon à atteindre le CV fixé pour le niveau de publication le plus désagrégé. Un exemple aide à mieux comprendre la situation.

Assume that we want to publish estimates for sector 72 (Accommodations and Food Services), which corresponds to NAICS-3 at the Canada level and the whole industry at the Province / Territory level. We then construct Table 2, in which the number of provinces has been reduced to 3 and the number of NAICS-3 for the industry as a whole is 2, specifically the sub-sectors (SS) 721 and 722.

Supposons qu'on veuille publier des estimations pour le secteur industriel 72 (Hébergement et services de restauration) pour lequel on publie au niveau SCIAN-3 pour le Canada et au niveau de l'industrie complète par province / territoire. On construit alors le tableau 2, où le nombre de provinces a été simplifié à 3 et le nombre de sous-secteurs (SS) SCIAN-3 pour l'industrie au complet est 2 (721 et 722).

**Table 2  
Cross Publication for Sector 72**

	Prov1	Prov2	Prov3	CV
SS721				15%
SS722				15%
CV	15%	15%	15%	

**Tableau 2  
Croisements de publication pour le secteur 72**

	Prov1	Prov2	Prov3	CV
SS721				15%
SS722				15%
CV	15%	15%	15%	

The initial stratification corresponds to each cell in table 2 and the marginals correspond to the estimates we wish to publish. If, for example, we wish to publish estimates with a target CV of 15%, we must first compute the CV to be targeted for each cell, so that the marginal

La stratification initiale correspond à chacune des cellules du tableau 2 et les marginales correspondent aux estimations qu'on désire publier. Si on désire, par exemple, publier des estimations avec un CV cible de 15%, on doit d'abord calculer le CV à viser pour chacune des cellules de telle sorte que les CV des

CVs are met.

Before we can compute the CV required at the cell level to reach the CV set for the marginals, we must adjust the marginal CVs. In fact, we cannot obtain 15% CVs in both directions, because when we set the variance in one direction to obtain the targeted CV, we automatically set the variance (thus the CV) for the other direction and we are "subject to" the resulting CV. With the knowledge that the CVs in both directions cannot be simultaneously equal to the targeted CV (unless by chance), we have chosen to minimize the distance from the marginal CVs to the target CV. In one direction, we then obtain a resulting CV greater than the target CV and in the other, a CV less than this same CV. This is done by minimizing the distance between the resulting CVs and the target CV under the constraint that the variances must be the same in both directions. In mathematical terms:

$$\text{Minimize } (CV^C - CVA)^2 + (CV^C - CVB)^2$$

under the constraint  $VA=VB$

where  $CVA$  and  $CVB$  represent the CVs attainable in both directions,  $CV^C$  represents the target CV and  $VA$  and  $VB$  represents the variances in both directions.

Let us call the resulting CV the new target CV. In the preceding example, we could end up with new target CVs as in Table 3.

**Table 3**  
**New target CVs(closest to the targeted CV)**

	Prov1	Prov2	Prov3	CV
SS721				11%
SS722				11%
CV	18%	18%	18%	

To reach the new target CV, we must compute what the targeted CVs should be for each of the initial strata by using a raking ratio algorithm as described in Latouche (1988).

Using the letters A and B again to designate the two directions (A the geographic direction and B the industrial direction, for example), we recompute the cell CVs until the combination of the CVs on the same line or in the same column is close enough to the target CV for the corresponding marginal.

marginales soient respectés.

Avant de pouvoir calculer le CV nécessaire au niveau des cellules pour atteindre le CV fixé au niveau des marginales, on doit d'abord ajuster ces CV marginaux. En effet, on ne peut obtenir des CV de 15% dans les deux directions, car lorsque l'on fixe la variance dans une direction pour obtenir le CV visé, on fixe automatiquement la variance (donc le CV) pour l'autre direction et on «subit» le CV résultant. Sachant que les CV des deux directions ne peuvent être simultanément égaux au CV visé (à moins d'un hasard), nous avons choisi de minimiser la distance des CV des marginales au CV cible. On obtient donc, dans une direction, un CV résultant supérieur au CV cible et dans l'autre, un CV inférieur à ce même CV. Ceci est fait en minimisant la distance entre les CV résultants et le CV cible sous la contrainte d'avoir des variances égales dans les deux directions. D'une façon mathématique:

$$\text{Minimiser } (CV^C - CVA)^2 + (CV^C - CVB)^2$$

sous la contrainte  $VA=VB$

où  $CVA$  et  $CVB$  représentent les CV atteignables dans les deux directions,  $CV^C$  représente le CV cible et  $VA$  et  $VB$  représentent les variances dans les deux directions.

Appelons le CV résultant le nouveau CV cible. Dans l'exemple précédent, on pourrait se retrouver avec de nouveaux CV cibles comme dans le tableau 3.

**Tableau 3**  
**Nouveaux CV cibles (les plus près du CV visé)**

	Prov1	Prov2	Prov3	CV
SS721				11%
SS722				11%
CV	18%	18%	18%	

Pour atteindre le nouveau CV cible, on doit calculer ce que devraient être les CV visés pour chacune des strates initiales en utilisant l'algorithme itératif du quotient tel que décrit dans Latouche (1988).

En utilisant à nouveau les lettres A et B pour désigner les deux directions (A la direction géographique et B la direction industrielle par exemple), on recalcule les CV des cellules jusqu'à ce que la combinaison des CV sur une même ligne ou une même colonne soit assez près du CV cible de la marginale correspondante.



$$CV_r^B(\hat{Y}_{ij}) = CV_{(r-1)}^A(\hat{Y}_{ij}) * \frac{CV(\hat{Y}_{.j})\hat{Y}_{.j}}{\sqrt{\sum_j (CV_{(r-1)}^A(\hat{Y}_{ij}))^2 \hat{Y}_{ij}^2}}$$

$$CV_r^A(\hat{Y}_{ij}) = CV_{(r-1)}^B(\hat{Y}_{ij}) * \frac{CV(\hat{Y}_{.i})\hat{Y}_{.i}}{\sqrt{\sum_j (CV_{(r-1)}^B(\hat{Y}_{ij}))^2 \hat{Y}_{ij}^2}}$$

where r denotes the current iteration,  
 r-1 denotes the preceding iteration,  
 i. denotes the marginal in direction A,  
 .j denotes the marginal in direction B,  
 ij denotes a crossover of directions A and B and  
 Y corresponds to the total for the income variable for a given group.

$$CV_r^B(\hat{Y}_{ij}) = CV_{(r-1)}^A(\hat{Y}_{ij}) * \frac{CV(\hat{Y}_{.j})\hat{Y}_{.j}}{\sqrt{\sum_j (CV_{(r-1)}^A(\hat{Y}_{ij}))^2 \hat{Y}_{ij}^2}}$$

$$CV_r^A(\hat{Y}_{ij}) = CV_{(r-1)}^B(\hat{Y}_{ij}) * \frac{CV(\hat{Y}_{.i})\hat{Y}_{.i}}{\sqrt{\sum_j (CV_{(r-1)}^B(\hat{Y}_{ij}))^2 \hat{Y}_{ij}^2}}$$

où r désigne l'itération courante,  
 r-1 désigne l'itération précédente,  
 i. désigne la marginale dans la direction A,  
 .j désigne la marginale dans la direction B,  
 ij désigne un croisement des directions A et B et  
 Y correspond au total de la variable revenu pour un groupement donné.

The algorithm stops when the convergence criterion (0.1%) is met or after a maximum of 10 iterations. It should be noted here that the algorithm converges very quickly and is almost certain to reach the targeted CV for the marginals. Table 4 illustrates the result of the iterative procedure.

Table 4

Cell CVs after Iterations

	Prov1	Prov2	Prov3	CV
SS721	20%	23%	24%	11%
SS722	17%	20%	21%	11%
CV	18%	18%	18%	

Now that the CV is set for each of the initial strata (these correspond to the cells in the preceding table), we can stratify them into two major strata: large, in which the sample is conducted with certainty, and small, in which the sampling is conducted under a probability scheme so the new target CV can be attained. The preferred method for splitting cells in two is that advanced by Hidiroglou (1986) which has the merit of minimizing the sampling size while attaining the target CV. The technique is simple: start with the equation that gives the CV for the initial stratum

$$CV(\hat{Y})^2 = \frac{(N-t)*(N-n(t))}{(n(t)-t)} S^2_{(N-t)} \hat{Y}^2$$

where N denotes the population size,  
 n(t) denotes the total number of units to be sampled,  
 t denotes the total number of units in the take-all stratum,  
 S<sup>2</sup>(n-t) denotes the variance in the take-some stratum and  
 Y corresponds to the total of the income variable for the stratum.

L'algorithme s'arrête lorsque le critère de convergence (0,1%) est rencontré ou après un maximum de 10 itérations. Notons ici que l'algorithme converge très rapidement et on atteint presque à coup sûr le CV visé au niveau des marginales. Le tableau 4 illustre le résultat du procédé itératif.

Tableau 4

CV des cellules après itérations

	Prov1	Prov2	Prov3	CV
SS721	20%	23%	24%	11%
SS722	17%	20%	21%	11%
CV	18%	18%	18%	

Maintenant que le CV est fixé pour chacune des strates initiales (celles-ci correspondent aux cellules du tableau précédent), on peut les stratifier en deux strates de taille: grande taille où l'échantillonnage se fait avec certitude et petite taille où l'échantillonnage se fait selon une probabilité de sélection permettant d'atteindre le nouveau CV cible. La méthode préconisée pour séparer les cellules en deux est celle d'Hidiroglou (1986) qui a le mérite de minimiser la taille échantillonnale tout en atteignant le CV cible. La technique est simple: on part de l'équation qui donne le CV de la strate initiale

$$CV(\hat{Y})^2 = \frac{(N-t)*(N-n(t))}{(n(t)-t)} S^2_{(N-t)} \hat{Y}^2$$

où N désigne la taille de la population,  
 n(t) désigne le nombre total d'unités à échantillonner,  
 t désigne le nombre total d'unités dans la strate à tirage complet,  
 S<sup>2</sup>(n-t) désigne la variance dans la strate à tirage partiel et  
 Y correspond au total de la variable revenu pour la strate.

It can be rewritten to isolate  $n(t)$ , the total number of units to be sampled based on  $t$ , the number of units sampled with certainty:

$$n(t) = t + \frac{(N-t)^2 S_{(N-t)}^2}{CV^2 \hat{Y}^2 + (N-t) S_{(N-t)}^2}$$

We then must clearly understand the function to find its minimum point. This can be attained through an iterative process that computes the following two parameters after converging: the dividing value separating the initial stratum into two final strata as well as the sample size for each of the strata. There will be  $t$  units in the take-all stratum and  $n(t) - t$  units to be taken in the take-some stratum. This process will have taken the minimum number of units to attain the target CV set.

It is highly likely that we will not obtain the precise target CV for the cells. The CV reached is usually close, but for some cells may be as much as 2% below the target CV. The effect of this is a slight change in the CVs targeted for the marginals. Table 5 reproduces the results from Table 4 following application of Hidiroglou's algorithm.

**Table 5**  
**Final Cell CVs after Iterations**

	Prov1	Prov2	Prov3	CV
SS721	20.1%	22.8%	24%	10.8%
SS722	17.2%	21.5%	20.4%	11.7%
CV	18.1%	18.9%	17.8%	

Once this step is complete, we can then proceed with the actual selection of the sample.

**Selection**

For the take-some strata, selection is based on a simple random process. A minimal fraction of 1% is imposed and a minimum of 3 units is required where possible. In the take-all strata, all units are sampled with certainty. This selection method forces no unit into the sample and takes no account of the preceding sample.

On peut la réécrire de façon à isoler  $n(t)$ , le nombre total d'unités à échantillonner en fonction de  $t$ , le nombre d'unités échantillonnées avec certitude:

$$n(t) = t + \frac{(N-t)^2 S_{(N-t)}^2}{CV^2 \hat{Y}^2 + (N-t) S_{(N-t)}^2}$$

Il s'agit alors de bien comprendre la fonction pour trouver son point minimum. Celui-ci peut être atteint selon un processus itératif qui calcule, après avoir convergé, les deux paramètres suivants: la borne qui sépare la strate initiale en deux strates finales ainsi que la taille échantillonnale de chacune des strates. On aura  $t$  unités dans la strate à tirage complet et  $n(t) - t$  unités à tirer dans la strate à tirage partiel. On aura ainsi tiré le nombre minimal d'unités pour atteindre le CV cible fixé.

Il est fort probable qu'on n'obtienne pas exactement le CV cible au niveau des cellules. Le CV atteint est habituellement près, mais peut être pour certaines cellules jusqu'à 2% au-dessus du CV cible. Ceci a pour effet de changer légèrement les CV visés au niveau des marginales. Le tableau 5 reprend les résultats du tableau 4 après l'application de l'algorithme d'Hidiroglou.

**Tableau 5**  
**CV final des cellules après itérations**

	Prov1	Prov2	Prov3	CV
SS721	20.1%	22.8%	24%	10.8%
SS722	17.2%	21.5%	20.4%	11.7%
CV	18.1%	18.9%	17.8%	

Lorsque cette étape est complétée, on peut alors procéder à la sélection proprement dite de l'échantillon.

**Sélection**

Pour les strates à tirage partiel, la sélection se fait selon un processus aléatoire simple. Une fraction minimale de 1% est imposée et un minimum de 3 unités est exigé là où c'est possible. Dans les strates à tirage complet, toutes les unités sont échantillonnées avec certitude. Cette méthode de sélection ne force aucune unité dans l'échantillon et ne tient nullement compte de l'échantillon précédent.

### 3.5.4 Data Editing

Once the sample has been selected, a questionnaire is mailed out and respondents are urged to complete and return it. Units that have not responded are subject to mail and telephone follow-up to ensure the data is obtained. A special effort is made for units in the take-all strata.

Once the data have been captured, some edits are conducted for each establishment. For example, several rules of consistency are in place to ensure that if some fields are coded, all related fields are also coded. For example, we can ensure that the sum of the parts equals the whole, that certain cells are properly filled out, etc.

Some edits focus directly on investment data. For example, if historical data are available, some tolerance rules are applied.

When no historical data are available, all respondents reporting investment of \$10,000,000 or more are the subject of thorough checks. It should be noted that these rules are subject to change.

Finally, a large number of qualitative (rather than quantitative) editing rules are also in place. For more details on editing rules, see Corneau (1995).

### 3.5.5 Calendarization

Once data has been collected and edited, we can proceed with the calendarization of the data. This process will generate data for the January to December period for the reference year when the respondent has given data on another period. In fact, to reduce the response burden, we accept that the respondent provides data on a fiscal basis. For a given year, its fiscal period must end between January 1<sup>st</sup> of the target year and March 31<sup>st</sup> of the following target year.

To prevent the production of estimations linked to many different fiscal periods, calendarization is done. The main idea is relatively simple: first "break" the annual data into monthly data, extrapolate if needed and then sum the monthly values forming the year of interest to get the calendarized data of the respondent.

The method developed by Cholette (1984) is used to "break" the data into monthly portions and extrapolate. The method is similar to a benchmarking technique. We can summarize the algorithm in the following manner:

We are trying to minimize the function

$$O(x) = \sum_{m=2}^T ((x_m - x_{m-1}) - (z_m - z_{m-1}))^2$$

in such a way that the sum of the monthly values ( $x_m$ ) over the fiscal period is equal to the respondent's reported data.

### 3.5.4 Vérification des données

Une fois l'échantillon tiré, un questionnaire est envoyé par la poste invitant le répondant à le retourner dûment rempli. Les unités n'ayant pas répondu font l'objet d'un suivi postal et téléphonique pour s'assurer d'obtenir leurs investissements. Un effort spécial est fait pour les unités faisant partie de strates à tirage complet.

Une fois les données saisies, certaines vérifications sont faites pour chaque établissement. Entre autres, plusieurs règles de cohérence sont en place permettant de s'assurer que si certains champs sont codés, tous ceux qui y sont reliés sont également codés. On peut par exemple, vérifier que la somme des parties égale le tout, que certaines cellules sont bien remplies, etc...

Certaines vérifications ont trait directement aux données d'investissement. Par exemple, si des données historiques sont disponibles, certaines règles de tolérance sont appliquées.

Lorsqu'aucune donnée historique n'est disponible, tous les répondants présentant des investissements de 10 000 000\$ et plus font l'objet de vérifications approfondies. Notons que ces règles sont sujettes à changements.

Finalement, un grand nombre de règles de vérification de nature qualitative (et non quantitative) sont aussi en place. Pour plus de détails sur les règles de vérification, consulter Corneau (1995).

### 3.5.5 Calendrialisation

Une fois la collecte effectuée et les données vérifiées, on procède à la calendrialisation des données. Ce processus permet d'obtenir des données pour la période de janvier à décembre de l'année de référence dans les cas où le répondant fournit ses données sur une autre base. En effet, pour diminuer le fardeau de réponse du répondant, on accepte qu'il fournisse ses données sur la base de sa période fiscale. Pour une année donnée, sa période fiscale doit se terminer entre le 1<sup>er</sup> janvier de l'année visée et le 31 mars suivant l'année visée.

Pour éviter de produire des estimations se rapportant à plusieurs périodes fiscales différentes, on a recours à la calendrialisation. Le principe directeur est assez simple: tout d'abord "briser" les données annuelles en données mensuelles, extrapoler si nécessaire puis sommer les mois qui composent l'année d'intérêt pour ainsi obtenir les données calendrialisées du répondant.

La méthode développée par Cholette (1984) est utilisée pour "briser" les données de façon mensuelle et faire l'interpolation. La méthode s'apparente à la technique d'annualisation de données ("Benchmarking"). En gros, on peut résumer l'algorithme comme suit:

On cherche à minimiser la fonction

$$O(x) = \sum_{m=2}^T ((x_m - x_{m-1}) - (z_m - z_{m-1}))^2$$

de telle sorte que la somme des valeurs mensuelles ( $x_m$ ) sur la période fiscale soit égale à la valeur rapportée par le répondant.

The series of  $z_m$  correspond to known auxiliary information about the respondent such as its cycle or trend. For the survey, this option is not used and the series is simply a constant value which corresponds to minimizing the month to month change (while the fiscal total is still respected).

The available number of months (T) on which the minimization function is calculated depends on the historical information of the respondent. However, since usually a respondent gets at least two questionnaires covering two distinct calendar years, T should at least be equal to 24. Periods that are not covered by the fiscal data (at the beginning and at the end of the series) are extrapolated using the last (or the first) calculated monthly value. The rest of the process can be applied on both calendar and fiscal data of the respondents.

### 3.5.6 Outlier Detection

Once the reported data are on a calendar basis, we proceed with the detection of outliers. Detection may be conducted at four levels, beginning at the most disaggregated. If there are not at least 25 units at this level, we proceed to the next level. As many as three variables may be involved in defining these levels: industrial level, size and geographic area.

There are three size categories: take-all stratum with known income, take-all stratum with unknown income, and take-some stratum.

With respect to geographic areas, units are located in large provinces (QUE, ON, ALTA and BC), mid-sized provinces (NS, NB, MAN and SASK), or small provinces (PEI, YUK, NWT, NUT and NFLD).

The four detection levels are:

- Level 1: NAICS-3 \* Size \* QUE, ON, ALTA, BC, small and mid-sized provinces (separated)
- Level 2: NAICS-3 \* Size \* large provinces and small and mid-sized provinces (together)
- Level 3: NAICS-3 \* Size \* Canada
- Level 4: Sector \* Canada

When publication is at the Sector level for an industry, detection begins at the most aggregate level, ie. level 4.

In addition, the outlier detection module is run before and after imputation. After imputation, this is done with the imputed data and permits detection of outliers among the imputed data.

The Hidiroglou-Berthelot (1986) method is used to detect them. Establishment "i" is considered an outlier if one of the two relations is checked:

$$\begin{aligned} y_i &< M - C \cdot DQ_1 \\ y_i &> M + C \cdot DQ_3 \end{aligned}$$

where  $DQ_1 = \text{Max}(M - Q_1, IA \cdot MI)$ ,  
 $DQ_3 = \text{Max}(Q_3 - M, IA \cdot MI)$ ,

La série des  $z_m$  correspond à l'information auxiliaire connue pour le répondant comme possiblement son cycle ou sa tendance. Pour l'enquête, on n'utilise pas cette option et la série est tout simplement mise égale à une constante ce qui correspond tout simplement à minimiser le changement d'un mois à l'autre (tout en s'assurant de respecter le total fiscal).

Le nombre de mois disponibles (T) sur lequel on calcule la fonction à minimiser dépend de l'information historique que l'on détient sur le répondant. Cependant, étant donné qu'habituellement un répondant reçoit au moins deux questionnaires correspondant à deux années civiles distinctes, T devrait être au minimum égal à 24. Les périodes qui ne sont pas couvertes par les données fiscales (au début ou à la fin de la série) sont extrapolées par la dernière (ou la première) valeur mensuelle calculée. Le reste du traitement décrit peut donc être appliqué sur les données calendrialisées ou sur les données civiles des répondants.

### 3.5.6 Détection de données aberrantes

Une fois les données transformées sur une base calendrier, on procède à la détection de données aberrantes. La détection peut se faire à quatre niveaux. On commence la détection au niveau le plus désagrégé. S'il n'y a pas au moins 25 unités à ce niveau, on passe au niveau suivant. Dans la définition de ces niveaux, jusqu'à trois variables peuvent être impliquées: le niveau industriel, la taille et la région géographique.

En ce qui a trait à la taille, il y a trois catégories: la strate à tirage complet avec revenu connu, la strate à tirage complet avec revenu inconnu et la strate à tirage partiel.

Quant à la région géographique, l'unité se retrouve soit dans les grosses provinces (QC, ON, AB et CB), dans les provinces moyennes (NÉ, NB, MB et SK) ou dans les petites provinces (IPE, TY, TNO, TNU et TN).

Les quatre niveaux de détection sont:

- Niv. 1 : SCIAN-3 \* Taille \* QC, ON, AB, CB, petites et moyennes provinces (séparées)
- Niv. 2 : SCIAN-3 \* Taille \* grosses provinces et petites et moyennes provinces (ensemble)
- Niv. 3 : SCIAN-3 \* Taille \* Canada
- Niv. 4 : Secteur \* Canada

Lorsque dans une industrie, on publie au niveau du secteur, on commence la détection au niveau le plus agrégé, soit le niveau 4.

De plus, la détection des valeurs aberrantes se fait avant et après imputation. La détection suite à l'imputation se fait avec les données imputées et permet ainsi de détecter les données imputées aberrantes.

La méthode d'Hidiroglou-Berthelot (1986) est utilisée pour les détecter. L'établissement "i" est considéré aberrant si une des deux relations est vérifiée:

$$\begin{aligned} y_i &< M - C \cdot DQ_1 \\ y_i &> M + C \cdot DQ_3 \end{aligned}$$

où  $DQ_1 = \text{Max}(M - Q_1, IA \cdot MI)$ ,  
 $DQ_3 = \text{Max}(Q_3 - M, IA \cdot MI)$ ,

M is the median (the point at which exactly 50% of establishments lie on either side),  
 Q<sub>1</sub> is the first quartile (25% of establishments are smaller and 75% are larger),  
 Q<sub>3</sub> is the third quartile (75% of establishments are smaller and 25% are larger),  
 A and C take the values of 0.5 and 20 respectively.

M est la médiane (c'est à dire le point où exactement 50% des établissements se trouvent de part et d'autres),  
 Q<sub>1</sub> est le premier quartile (25% des établissements sont plus petits et 75% sont plus grands),  
 Q<sub>3</sub> est le troisième quartile (75% des établissements sont plus petits et 25% sont plus grands),  
 A et C prennent respectivement les valeurs: 0,5 et 20.

The outliers are detected based on four values of  $y_i$  (eight in the case of SA). The first corresponds to the ratio of CC (calendarized) to revenue (investment takes the value of CC in the y equation) while the second corresponds to the ratio of CM (calendarized) to revenue. In the third and fourth cases, we are using the same ratios, replacing the calendarized values by their fiscal values. If an establishment is found to be an outlier under one of these rules, it is automatically deemed an outlier (for the two investment variables, calendarized and fiscal).

Les données aberrantes sont détectées selon quatre valeurs de  $y_i$  (huit dans le cas de l'ER). La première correspond au ratio de CC calendarisé sur le revenu (invest. prend la valeur de CC dans l'équation de y) alors que la deuxième correspond au ratio de CM calendarisé sur le revenu. Les troisième et quatrième valeurs utilisées sont simplement les mêmes ratios, en utilisant les valeurs fiscales (et non pas calendarisées). Si un établissement est trouvé aberrant selon une ou l'autre de ces règles, il est automatiquement considéré aberrant (pour les deux variables d'investissement, calendarisées et fiscales).

### 3.5.7 Imputation

Records found to be outliers are not imputed since the consistency rules have been applied and the investment reported by the respondent is deemed valid. These records are simply excluded from calculation of the average during imputation of non-respondents. Moreover, if some of the establishments found to be outliers form part of the take-some strata, they are moved up to the take-all strata with known revenues and the selection probability for residual units is recomputed.

### 3.5.7 Imputation

Les enregistrements trouvés aberrants ne sont pas imputés puisque les règles de cohérence ont été appliquées et que l'on considère valide l'investissement déclaré par le répondant. On ne fait qu'exclure ces enregistrements du calcul de moyenne lors de l'imputation des non-répondants. De plus, si certains des établissements trouvés aberrants font partie de strates à tirage partiel, alors ils sont promus dans la strate à tirage complet avec revenu connu et la probabilité de sélection des unités résiduelles est recalculée.

For records to be imputed, three imputation methods are used to proceed with evaluation of the missing data. There is no partial imputation: the two variables of interest, CC and CM (RC and RM are added in the case of the SA) are available or missing for each establishment. The three methods therefore allow us to impute all of the variables in parallel. The first method, used for three of the four survey cycles, is simply a substitution with the historical value:

Pour les enregistrements à imputer, trois méthodes d'imputation sont utilisées pour procéder à l'évaluation des données manquantes. Il n'existe pas d'imputation partielle: les deux variables d'intérêts, CC et CM (on ajoute RC et RM dans le cas de l'ER) sont disponibles ou manquantes pour chacun des établissements. Les trois méthodes permettent donc d'imputer parallèlement toutes les variables. La première méthode, utilisée pour trois des quatre cycles de l'enquête, est tout bonnement la substitution par valeur historique:

$$y_{it} = y_{i(t-1)}$$

$$y_{it} = y_{i(t-1)}$$

where t is the current survey, t-1 the preceding survey and y one of the investment variables (CC or CM). This imputation is done (whenever possible) before the calendarization process. Hence data imputed from a period that could be different from the calendar year are calendarized as well.

où t correspond à l'enquête courante, t-1 à l'enquête précédente et y à une des variables d'investissements (CC ou CM). Cette imputation est faite (là où elle est possible) avant de procéder à la calendarisation. Ainsi les données imputées à partir de données pouvant provenir d'une période autre que l'année calendrier sont également calendarisées.

We should note here that time t-1 refers to any other survey during which the data have been reported and which cover the same reference year. The only exception arises from the fact that the RC and RM variables are not required in the questionnaires covering the same reference year. The historical data are then those from the preceding reference year. In the case of the SI, the only imputation method used is the current ratio:

Notons ici que le temps t-1 fait référence à toute autre enquête lors de laquelle les données ont été rapportées et qui couvriraient la même année de référence. La seule exception provient du fait que les variables RC et RM ne sont pas requises lors de questionnaires couvrant la même année de référence. Les données historiques sont alors celles de l'année de référence précédente. Dans le cas de l'EP, la seule méthode d'imputation utilisée correspond à celle du quotient des valeurs courantes:

$$y_{it} = \frac{\bar{y}_t}{\bar{x}_t} x_{it}$$

where x is revenue.

This method is also used for the three other surveys in cases where there is no historical information reported for the year surveyed.

Note that units which are part of take-all strata with unknown income are imputed with the average of current values rather than the ratio:

$$y_{it} = \bar{y}_t$$

An important factor when computing the imputed value is the level at which imputation is conducted. In fact, the imputation is conducted if the imputation group includes at least 10 establishments for which the questionnaire is complete and if these represent at least 25% of units in the group.

### Imputation Groups

The initial imputation group corresponds to the stratum used for sampling once it is updated with the new data gathered. If one of the preceding constraints (10 units, 25% of units) is not met, we move to a more aggregated imputation group within the same industrial group and in the same size group, but in which all provinces are combined. As in outlier detection, the possible sizes are take-all stratum with known income, take-all stratum with unknown income and take-some stratum.

If the constraints still are not met, the industries are grouped. For example, all NAICS-6s from a given NAICS-5 are combined. We remain at the Canada level and within the same size group. The most aggregated level we can reach corresponds to the groups for all NAICS-3s in a given sector, at the Canada level, for one size group where the last level of the take-all stratum with known and unknown revenues are regrouped. Two examples will provide a better understanding.

If an establishment in the Canadian mining industry 212114 in Ontario that is part of the take-some group is to be imputed, we obtain the following sequence:

212114 - Ontario - take-some stratum  
 212114 - Canada - take-some stratum  
 21211 - Canada - take-some stratum  
 2121 - Canada - take-some stratum  
 212 - Canada - take-some stratum  
 Mining and Oil and Gas Extraction  
 sector - Canada - take-some stratum

If an establishment in sector 55 (Management of Companies and Enterprises) in Quebec that is part of the take-all group with unknown revenues is to be imputed, we obtain the following sequence:

$$y_{it} = \frac{\bar{y}_t}{\bar{x}_t} x_{it}$$

où x correspond au revenu.

Cette méthode est également utilisée pour les trois autres enquêtes dans le cas où on n'a aucune information historique rapportée pour l'année enquêtée.

Notons que les unités qui font partie de strates à tirage complet avec revenu inconnu sont imputées avec la moyenne des valeurs courantes au lieu du quotient:

$$y_{it} = \bar{y}_t$$

Un facteur important lors du calcul de la valeur imputée est le niveau auquel se fait l'imputation. En effet, l'imputation a lieu si le groupe d'imputation comprend au moins 10 établissements dont le questionnaire est complet et que ceux-ci représentent au moins 25% des unités du groupe.

### Groupes d'imputation

Le groupe d'imputation initial correspond à la strate utilisée pour l'échantillonnage une fois mise à jour avec les nouvelles données recueillies. Si l'une ou l'autre des contraintes précédentes (10 unités, 25% des unités) n'est pas satisfaite, on passe à un groupe d'imputation plus agrégé se situant dans le même groupe industriel et dans le même groupe de taille mais où toutes les provinces sont combinées. Comme dans la détection des valeurs aberrantes, les tailles possibles sont strate à tirage complet avec revenu connu, strate à tirage complet avec revenu inconnu et strate à tirage partiel.

Si encore une fois les contraintes ne sont pas satisfaites, on regroupe les industries. On combine par exemple tous les SCIAN-6 d'un même SCIAN-5. On demeure au niveau canadien et à l'intérieur du même groupe de taille. Le niveau le plus agrégé qu'on peut atteindre correspond aux groupements de tous les SCIAN-3 d'un même secteur, au niveau canadien, pour un groupe de taille où au dernier niveau les tailles strate à tirage complet "avec revenu connu" et "inconnu" sont regroupées. Deux exemples permettront de mieux comprendre.

Si un établissement de l'industrie minière canadienne 212114 en Ontario faisant partie du groupe à tirage partiel est à imputer, on a la séquence suivante:

212114 - Ontario - strate à tirage partiel  
 212114 - Canada - strate à tirage partiel  
 21211 - Canada - strate à tirage partiel  
 2121 - Canada - strate à tirage partiel  
 212 - Canada - strate à tirage partiel  
 Secteur de l'extraction minière, de pétrole et gaz -  
 Canada - strate à tirage partiel

Si un établissement du secteur 55 (Gestion de sociétés et d'entreprises) au Québec faisant partie du groupe à tirage complet avec revenu inconnu est à imputer, on a la séquence suivante:

Sector 55-Quebec-take-all stratum (unknown revenues)  
Sector 55-Canada-take-all stratum (unknown revenues)  
Sector 55-Canada-take-all stratum (known and unknown revenues)

We should also point out that a record imputed at a disaggregated level can be used to compute the averages during imputation of another record at a more aggregated level. For example, if we manage to impute all records for Alberta at the first imputation level and must move to the next level for records from New Brunswick, these will be imputed at the Canadian level and the imputed Alberta records will be used in computing the averages at the Canadian level.

Once the missing values for establishments are imputed, we can move on to the estimation stage.

### 3.5.8 Estimation

The ratio estimator is used for estimation with revenue being the auxiliary variable. This method ensures that the final weight multiplied by the income for each unit in the sample matches the known total for the income variable for the entire population in the group. The groups used in this instance correspond to the lowest industry level published within a single size group at the Canadian level. The difference from the original stratum is the grouping at the Canadian level. The following example provides a better understanding.

For an establishment for which the stratum corresponds to NAICS-3 323 of the Manufacturing sector in Nova Scotia for the take-some stratum, we use the estimation group

323 - Canada - take-some stratum

During the survey, an establishment may be reclassified into a new industry or province. This new classification is used to define the domain of publication and it is this classification that will determine where the investments will appear in the final table. The following example provides a better understanding.

If an establishment sampled in Quebec under NAICS-3 411 is found in Ontario under NAICS-3 444, it will have the following characteristics:

stratum: 411 - Quebec  
group for computing outliers: 444 - Ontario  
initial imputation group: 444 - Ontario  
estimation group: 411 - Canada  
domain of publication: 444 - Ontario

Here is the ratio estimator formula

$$\hat{Y}_d = \sum_h \sum_{i \in S_h} \frac{N_h}{n_h} G_i y_i(d)$$

where for each unit  $i$  of a group  $g$ ,

Secteur 55-Québec-strate à tirage complet (revenu inconnu)  
Secteur 55-Canada-strate à tirage complet (revenu inconnu)  
Secteur 55-Canada-strate à tirage complet (revenu connu et inconnu)

Notons aussi qu'un enregistrement imputé à un niveau désagrégé peut servir dans le calcul des moyennes lors de l'imputation d'un autre enregistrement à un niveau plus agrégé. Par exemple, si on réussit à imputer tous les enregistrements de l'Alberta au premier niveau d'imputation et qu'on doit passer au niveau suivant pour les enregistrements du Nouveau-Brunswick, ceux-ci seront imputés au niveau canadien et les enregistrements imputés de l'Alberta seront utilisés dans le calcul des moyennes au niveau canadien.

Une fois les valeurs manquantes des établissements imputées, on peut passer à l'étape d'estimation.

### 3.5.8 Estimation

L'estimation se fait par la méthode du quotient avec le revenu comme variable de contrôle. Cette méthode assure que le poids final multiplié par le revenu de chaque unité de l'échantillon, respectera le total connu de la variable revenu pour la population complète du groupe. Les groupes utilisés correspondent cette fois-ci au plus bas niveau industriel publié, à l'intérieur d'un même groupe de taille, au niveau canadien. La différence avec la strate originale est le groupement au niveau canadien. L'exemple suivant permet de mieux comprendre.

Pour un établissement dont la strate correspond au SCIAN-3 323 du secteur de la fabrication dans la Nouvelle-Écosse pour le groupe de strate à tirage partiel, on utilise le groupe d'estimation

323 - Canada - strate à tirage partiel

Lors de l'enquête, il est possible qu'un établissement soit reclassifié dans une nouvelle industrie ou dans une nouvelle province. On utilise cette nouvelle classification pour définir les domaines de publication et c'est cette classification qui déterminera où les investissements seront présents dans les tableaux finaux. L'exemple suivant permet de mieux comprendre.

Si un établissement, échantillonné au Québec selon le SCIAN-3 411, est retrouvé en Ontario selon le SCIAN-3 444, il aura les caractéristiques suivantes:

strate : 411 - Québec  
groupe pour calcul de données aberrantes : 444-Ontario  
groupe d'imputation initial : 444 - Ontario  
groupe d'estimation : 411 - Canada  
domaine de publication : 444 - Ontario

Voici la formule de l'estimateur par le quotient utilisé

$$\hat{Y}_d = \sum_h \sum_{i \in S_h} \frac{N_h}{n_h} G_i y_i(d)$$

où pour chaque unité  $i$  d'un groupe  $g$ ,

$$G_i = \frac{\sum_{j \in P_s} x_j}{\sum_{j \in P_s} \frac{x_j}{p_j}} \text{ and } y_i(d) = \begin{cases} y_i & \text{if } i \in d \\ 0 & \text{otherwise} \end{cases}$$

where  $y$  is the variable of interest (investment),  
 $x$  is the auxiliary variable (revenue),  
 $h$  denotes the stratum,  
 $g$  denotes the estimation group,  
 $d$  denotes the domain of publication,  
 $n$  denotes the sample size,  
 $N$  denotes the population size,  
 $s$  denotes the sample,  
 $P$  denotes the population,  
 $G$  denotes the control weight ("G-weight") and  
 $p$  denotes the selection probability.

Note that the G-weight calculation is done in such a way that the final weight cannot be lower than one. In doing that, we ensure that a respondent's value will be at least that value once it is weighted.

#### Estimation of Variance and Calculation of CV

Variance is estimated using Taylor's linearization formula in the case of ratio estimator. This is available in Estevao (1991). Using the same notation as before:

$$\hat{V}(\hat{Y}(d)) = \sum_h \frac{N_h - n_h}{n_h - 1} \frac{n_h}{N_h} \sum_{i \in s_h} (u_{hi} - \bar{u}_h)^2$$

$$\text{Where } u_{hi} = \frac{N_h}{n_h} G_i \left( y_i(d) - x_i * \frac{\sum_{i \in s_g} y_i / p_i}{\sum_{i \in s_g} x_i / p_i} \right)$$

$$\text{and } \bar{u}_h = \frac{\sum_{i \in s_h} u_{hi}}{n_h}$$

The coefficient of variation (CV) is computed using the ratio:

$$CV(\hat{Y}(d)) = \frac{\sqrt{\hat{V}(\hat{Y}(d))}}{\hat{Y}(d)}$$

When the estimates are published, a scale distinguishes between the various qualities of accuracy. It combines the effect of sampling (since we did not do a census) and the imputation rate (each imputation (other than historical imputation) adds to the uncertainty of the results). The scale is presented in Table 6.

$$G_i = \frac{\sum_{j \in P_s} x_j}{\sum_{j \in P_s} \frac{x_j}{p_j}} \text{ et } y_i(d) = \begin{cases} y_i & \text{si } i \in d \\ 0 & \text{si non} \end{cases}$$

où  $y$  est la variable d'intérêt (l'investissement),  
 $x$  est la variable auxiliaire (le revenu),  
 $h$  désigne la strate,  
 $g$  désigne le groupe d'estimation,  
 $d$  désigne le domaine de publication,  
 $n$  désigne la taille échantillonnale,  
 $N$  désigne la taille de la population  
 $s$  désigne l'échantillon,  
 $P$  désigne la population,  
 $G$  désigne le poids de contrôle ("G-weight") et  
 $p$  désigne la probabilité de sélection.

Notons que le calcul du poids-g est de telle sorte que le poids final est borné inférieurement à 1. Ainsi on évite que pour des raisons mathématiques les valeurs de certains répondants valent moins que les valeurs rapportées une fois pondérées.

#### Estimation de la variance et calcul de CV

La variance est estimée à l'aide de la formule de linéarisation de Taylor dans le cas de l'estimateur par le quotient. Celle-ci est disponible dans Estevao (1991). En suivant la même notation que précédemment:

$$\hat{V}(\hat{Y}(d)) = \sum_h \frac{N_h - n_h}{n_h - 1} \frac{n_h}{N_h} \sum_{i \in s_h} (u_{hi} - \bar{u}_h)^2$$

$$\text{où } u_{hi} = \frac{N_h}{n_h} G_i \left( y_i(d) - x_i * \frac{\sum_{i \in s_g} y_i / p_i}{\sum_{i \in s_g} x_i / p_i} \right)$$

$$\text{et } \bar{u}_h = \frac{\sum_{i \in s_h} u_{hi}}{n_h}$$

Le coefficient de variation (CV) est calculé à l'aide du quotient:

$$CV(\hat{Y}(d)) = \frac{\sqrt{\hat{V}(\hat{Y}(d))}}{\hat{Y}(d)}$$

Lors de la publication des estimations, une échelle permet de distinguer entre les différentes qualités de précision. Celle-ci combine l'effet dû à l'échantillonnage (puisqu'on n'a pas effectuée de recensement) et le taux d'imputation (chaque imputation (autre qu'historique) ajoute à l'incertitude des résultats). L'échelle est reproduite au tableau 6.



**Table 6**  
**Quality indicator interpretation**

CV	Imputation Rate			
	0.00 - 0.10	0.10 - 0.33	0.33 - 0.60	0.60 - + + +
0.00 - 0.01	A	B	C	F
0.01 - 0.05	B	C	D	F
0.05 - 0.25	C	D	E	F
0.25 - 0.33	D	E	F	F
0.33 - 0.50	E	F	F	F
0.50 - + + +	F	F	F	F

A: Excellent      B: Very good      C: Good  
D: Acceptable      E: Use with caution      F: Unpublishable

Unpublishable data will be hidden with the code \* and class E data will be identified with a superscript E. The rest of the data will have no distinguishing marks.

*Due to some technical considerations, the quality indicator will not be implemented for the present publication.*

### Confidentiality

Some confidentiality rules obviously are used to suppress any information that might lead to disclosure of the data supplied by a respondent. These rules allow Statistics Canada to comply with its mandate of non-disclosure of information supplied by respondents. The rules themselves are confidential and are not available for consultation.

### 3.5.9 Sampling error and non-sampling error

The difference between an estimate based on sample data and the value obtained by surveying the entire population is called the sampling error. This difference varies with sample size, expenditure variability, sampling scheme, and estimation method. In general, the larger a sample, the smaller its sampling error. If the population is very heterogeneous, a larger sample size is required to produce a reliable estimate. The sampling error is measured by a quantity known as the standard deviation. The latter indicates the expected variability of the estimate that will be produced if the expenditures are sampled repeatedly. The actual value of the standard deviation is unknown, but it can be estimated from the sample.

Another measure of precision is the coefficient of variation (CV). The CV is simply the standard deviation expressed as a percentage of the estimate. Hence it is a relative measure of precision and can be used for comparisons across industries or provinces. The smaller the CV, the more reliable the estimate. (See Quality Measures table in Section 5).

Another kind of error is non-sampling error. Although every effort is made to keep such errors to a minimum (section 4.4), they always exist. They are not taken into

**Tableau 6**  
**Interprétation de la cote de qualité**

CV	Taux d'imputation			
	0.00 - 0.10	0.10 - 0.33	0.33 - 0.60	0.60 - + + +
0.00 - 0.01	A	B	C	F
0.01 - 0.05	B	C	D	F
0.05 - 0.25	C	D	E	F
0.25 - 0.33	D	E	F	F
0.33 - 0.50	E	F	F	F
0.50 - + + +	F	F	F	F

A: Excellent      B: Très bon      C: Bon  
D: Acceptable      E: Utiliser avec réserve      F: Non-publiables

Les données non-publiables seront cachées par le code \* alors que les données de la classe E seront identifiées avec un exposant E. Le reste des données n'auront aucun caractère spéciaux.

*À cause de problèmes techniques, la cote de qualité ne sera pas disponible pour cette publication.*

### Confidentialité

Certaines règles de confidentialité sont évidemment utilisées pour supprimer toute information qui pourrait mener à la divulgation des données fournies par un répondant. Ces règles permettent à Statistique Canada de respecter son mandat de non-divulgation d'information fournie par les répondants. Les règles elles-mêmes sont confidentielles et ne sont pas disponibles pour consultation.

### 3.5.9 Erreur d'échantillonnage et non-due à l'échantillonnage

La différence entre l'estimation produite à partir de données échantillonnées et de données recensées est appelée erreur d'échantillonnage. Cette différence varie plus ou moins selon la taille de l'échantillon, la variabilité des dépenses, le plan de sondage et la méthode d'estimation. En général, un échantillon plus grand produit une erreur d'échantillonnage plus petite. Si la population est très hétérogène, une taille d'échantillon plus grande est requise pour produire une estimation fiable. L'erreur d'échantillonnage est mesurée par une quantité appelée écart-type. Cette quantité mesure la variabilité anticipée de l'estimation produite si on fait un échantillonnage répété des dépenses. La vraie valeur de l'écart-type est inconnue mais peut être estimée à partir de l'échantillon.

Une deuxième mesure de précision est le coefficient de variation (CV). Ce coefficient est simplement l'écart-type exprimé en pourcentage de la valeur de l'estimation. Il donne donc une mesure de précision relative et comparable entre différentes industries ou provinces. Notons qu'un plus petit CV indique une plus grande fiabilité de l'estimation. (Voir tableau sur les Mesures de la qualité à la section 5).

En plus de l'erreur d'échantillonnage, il existe des erreurs non-dues à l'échantillonnage sur lesquelles on tente de conserver un contrôle des plus stricts (section 4.4). Néanmoins, celles-ci

account in computing the CV, nor are they measured by the CV. Measures such as response rate, coverage rate and imputation rate can be used as indicators of the possible extent of non-sampling errors.

### 3.5.10 References

Cholette, P.A. (1984). Adjusting Sub-Annual Series to Yearly Benchmarks. *Survey Methodology*, vol 10. No. 1, 39-53.

Corneau, L. (1995). Spécifications des règles de vérification dans le cadre de l'enquête sur les dépenses en immobilisations. Internal Statistics Canada document, December 1995.

Cuthill, I. (1996). The Statistics Canada Business Register. Internal Statistics Canada Document, 1996.

Estevao, V. (1991). Generalized Estimation System, Methodology Review. Internal Statistics Canada document, September 1991.

Hidiroglou, M.A. (1986). The Construction of a Self-Representing Stratum of Large Units in Survey Design. *The American Statistician*, 40, 27-31

Hidiroglou, M.A. and Berthelot, J.-M. (1986). Contrôle statistique et imputation dans les enquêtes-entreprises périodiques. *Survey Methodology* 12, 79-89.

Lacroix, J. (1991). Capital and Repair Expenditures - Surveys Overview. Internal Statistics Canada document, January 1991.

Latouche, M. (1988). Détermination, allocation et sélection de l'échantillon. Cahier de travail BSMD-88-021 de Statistique Canada, May 1988

Pandher G.H. (1995). Population asymétrique: construction optimale de groupes "à tirage complet" et "échantillons", avec application au remaniement de l'enquête sur les finances des administrations locales. Cahier de travail SSMD-95-001 de Statistique Canada, March 1995.

### 3.5.11 Glossary

AD: Agriculture Division  
 BR: Business Register  
 BRD: Business Register Division  
 CC: Capital expenditures for new construction  
 CES: Capital Expenditure Survey  
 CM: Capital expenditures for new machinery and new equipment  
 CV: Coefficient of variation  
 ICSD: Investment and Capital Stock Division  
 IP: Integrated Portion  
 NIP: Non-integrated portion  
 NAICS: North American Industrial Classification System

existent toujours et ne sont pas comptabilisées lorsque l'on calcule le coefficient de variation et ne sont pas incluses dans cette mesure de précision. Certaines mesures telles que les taux de réponses, de couverture et d'imputation peuvent être utilisées comme indicateurs du niveau potentiel des erreurs non-liées à l'échantillonnage.

### 3.5.10 Références

Cholette, P.A. (1984). L'ajustement des séries infra-annuelles aux repères annuels. *Technique d'enquête*, vol 10. No. 1, 39-53.

Corneau, L. (1995). Spécifications des règles de vérification dans le cadre de l'enquête sur les dépenses en immobilisations. Document interne de Statistique Canada, décembre 1995.

Cuthill, I. (1996). The Statistics Canada Business Register. Document interne de Statistique Canada, 1996.

Estevao, V. (1991). Generalized Estimation System, Methodology Review. Document interne de Statistique Canada, septembre 1991.

Hidiroglou, M.A. (1986). The Construction of a Self-representing Stratum of Large Units in Survey Design. *The American Statistician*, 40, 27-31.

Hidiroglou, M.A. et Berthelot, J.-M. (1986). Contrôle statistique et imputation dans les enquêtes-entreprises périodiques. *Techniques d'enquêtes* 12, 79-89.

Lacroix, J. (1991). Capital and Repair Expenditures - Surveys Overview. Document interne de Statistique Canada, janvier 1991.

Latouche, M. (1988). Détermination, allocation et sélection de l'échantillon. Cahier de travail BSMD-88-021 de Statistique Canada, mai 1988.

Pandher G.H. (1995). Population asymétrique: construction optimale de groupes "à tirage complet" et "échantillons", avec application au remaniement de l'enquête sur les finances des administrations locales. Cahier de travail SSMD-95-001 de Statistique Canada, mars 1995.

### 3.5.11 Glossaire

CC: Capitaux immobilisés pour de la nouvelle construction  
 CM: Capitaux immobilisés pour de la nouvelle machinerie et du nouvel équipement  
 CTI: Classification type des industries  
 CV: Coefficient de variation  
 DA: Division de l'agriculture  
 DIP: Division des institutions publiques  
 DISC: Division des investissements et stocks de capitaux  
 DRE: Division du registre des entreprises  
 EDI: Enquête sur les dépenses en immobilisations  
 EP: Enquête sur les perspectives  
 EPR: Enquête sur les perspectives révisées  
 ER: Enquête sur les données réelles

PID: Public Institution Division  
RC: Repair expenditures on construction  
RM: Repair expenditures on machinery and equipment  
SA: Survey on Actual Data  
SI: Survey on Intentions  
SIC: Standard Industrial Classification  
SPA: Survey on Preliminary Actual Data  
SRI: Survey on Revised Intentions  
SS: Sub-sector

ERP: Enquête sur les données réelles provisoires  
PI: Portion intégrée  
PNI: Portion non-intégrée  
RC: Capitaux immobilisés pour la réparation reliée à la construction  
RE: Registre des entreprises  
RM: Capitaux immobilisés pour la réparation reliée à la machinerie et à l'équipement  
SCIAN: Système de classification ind. de l'Amérique du Nord  
SS: Sous-secteur

### 3.6 Users and Uses

Within Statistics Canada, data collected by capital expenditures surveys are used by the System of National Accounts, National Accounts and Environment Division, to benchmark the quarterly projections of gross fixed capital formation by government and businesses. The Investment and Capital Stock Division, National Wealth and Capital Stock Section, uses the investment series to produce estimates of the gross and net capital stock as well as depreciation. In turn, the estimates of capital stock are used in the calculation of productivity estimates. Other Statistics Canada divisions using the investment series in the production of various statistics include Industry Measures and Analysis and Input-Output Division.

In the public sector, aggregated capital investment data are used by the Department of Finance in the development of fiscal policy and to calculate equalization payments to the provinces. The Bank of Canada uses the capital expenditures series in the development of monetary policy while Industry, Science and Technology uses the series in regional industrial policy development.

In the private sector, aggregated capital expenditures data are used in the development of economic policy by institutions such as the chartered banks and consulting firms. Analysis of market demands can be conducted using capital expenditures data, while investment intentions and revised intentions can be used for projecting demands on labour and materials. Through special tabulations, suppliers of machinery and equipment can determine market share through an evaluation of the capital expenditures for the identified machinery and equipment within a particular industry.

Provincial/territorial statistical agencies and departments use the expenditures series in micro data form for the production of various provincially based statistics. Information is shared under Section 11 of the Statistics Act with; the Newfoundland Statistics Agency, the New Brunswick Statistics Agency, the "Bureau de la statistique du Québec", the Statistics Unit, Ontario Ministry of Treasury and Economics, the Manitoba Bureau of Statistics, the Saskatchewan Bureau of Statistics, and the British Columbia Central Statistics Bureau for respondents in each of the respective provinces. Furthermore, sharing of information is conducted under Section 12 of the Statistics Act with; the Prince Edward Island Department of Finance, the Nova Scotia Department of Development Statistics and Research Services Branch, the Budget Planning and Economics Division of the Alberta Treasury, the Northwest Territories Bureau of Statistics and the Yukon Bureau of Statistics.

Also through Section 12 data sharing agreements, the following provincial departments are given access to mining related micro data; the Newfoundland Department

### 3.6 Utilisateurs et utilisations

À Statistique Canada, la Division des comptes nationaux et de l'environnement de la Direction du système de comptabilité nationale se sert des données recueillies lors des enquêtes sur les dépenses en immobilisations pour appuyer les projections de formation brute de capital fixe au gouvernement et dans l'entreprise. Pour sa part, la Section de la richesse nationale et des stocks de capital de la Division de l'investissement et du stock de capital emploie la série sur les investissements pour estimer le stock de capital brut et net ainsi que l'amortissement pour dépréciation. Ces estimations de stock de capital sont ensuite utilisées dans les estimations de productivité. D'autres divisions de Statistique Canada emploient la série des investissements pour produire diverses statistiques, notamment la Division des mesures et de l'analyse des industries et la Division des entrées-sorties.

Dans le secteur public, l'ensemble des données sur les investissements sont utilisées par le ministère des Finances dans l'élaboration de la politique fiscale et dans le calcul des paiements de péréquation aux provinces. La Banque du Canada se sert de la série des dépenses en immobilisations dans la formulation de sa politique monétaire tandis qu'Industrie, Sciences et Technologie Canada en fait usage pour élaborer sa politique industrielle régionale.

Dans le secteur privé, l'ensemble des données sur les investissements permettent à des établissements comme les banques à charte et les sociétés d'experts-conseils d'établir leurs politiques économiques. Les données sur les dépenses en immobilisations servent aussi à analyser la demande sur le marché, tandis que les perspectives d'investissements et les perspectives révisées peuvent servir à prévoir la demande portant sur la main-d'oeuvre et les matériaux. En consolidant des totalisations, les fournisseurs de matériel et d'outillage peuvent connaître les parts de marché en évaluant les dépenses en immobilisations consacrées aux différents types de matériel et d'outillage dans chaque secteur.

Les ministères et organismes chargés des statistiques provinciales et territoriales emploient aussi la série des dépenses sous forme de microdonnées dans l'exécution de leur mandat. Conformément à l'article 11 de la *Loi sur la statistique*, les renseignements se rapportant aux diverses provinces sont communiqués respectivement à la Newfoundland Statistics Agency, à l'Agence de la statistique du Nouveau-Brunswick, au Bureau de la statistique du Québec, à l'Unité de la statistique du ministère du Trésor et de l'Économie de l'Ontario, au Bureau de la statistique du Manitoba, au Saskatchewan Bureau of Statistics, et au British Columbia Central Statistics Bureau. Des données sont aussi transmises conformément à l'article 12 de la *Loi sur la statistique* au Department of Finance de l'Île-du-Prince-Édouard, à la Statistics and Research Services Branch du Department of Development de la Nouvelle-Écosse, à la Budget Planning and Economics Division de la Treasury Branch de l'Alberta, au Bureau de la statistique des Territoires du Nord-Ouest et au Bureau de la statistique du Yukon.

Toujours dans le cadre d'ententes conclues en vertu de l'article 12, les ministères provinciaux suivants ont accès aux microdonnées portant sur les mines : le Department of Mines

of Mines and Energy, the Nova Scotia Department of Natural Resources, the New Brunswick Department of Natural Resources, Ministry of Northern Development and Mines of Ontario, the Manitoba Department of Energy and Mines, and the British Columbia Ministry of Energy, Mines and Petroleum Resources. The Mineral Policy and Energy Policy Sectors of Natural Resources Canada have access to micro data related to the mining industry and the petroleum and natural gas industry, respectively.

and Energy de Terre-Neuve, le Department of Natural Resources de Nouvelle-Écosse, le ministère des Ressources naturelles du Nouveau-Brunswick, le ministère du Développement du Nord et des Mines de l'Ontario, le ministère de l'Énergie et des Mines du Manitoba et le Ministry of Energy, Mines and Petroleum Resources de la Colombie-Britannique. Les secteurs de la Politique minérale et de la Politique énergétique de Ressources naturelles Canada ont accès respectivement aux microdonnées portant sur l'industrie minière et sur l'industrie du pétrole et du gaz naturel.

### 3.7 Expenditure Series Chronology

In 1941 the Dominion Bureau of Statistics initiated the first actual capital expenditure series with the collection of, among other information, capital expenditure data on selected industries. The first forecast of investment was released to the public in the fall of 1946 as *Capital, Repair and Maintenance Expenditures of Business Enterprises in Canada: Forecast 1946*.

In 1947, the scope of the capital expenditure series was expanded to include capital items charged to operating expenses. The addition of this type of capital expenditure increased the accuracy of the reported data by providing an estimate of all those items which add to the capital stock of the country, but were not capitalized by the reporting industries.

Since 1946, the coverage of capital expenditure survey has grown to encompass more sectors of the economy. Capital expenditures for the mining and manufacturing sub-industries were presented in the *Service Bulletin: Investment Statistics* (Catalogue 61-007) starting in 1975, followed by the first appearance of energy related data in 1976. The release of energy related data in volume 2, number 2 of the *Service Bulletin: Investment Statistics* included current year data as well as estimates dating back to 1955.

In 1978 the first issue of *Capital and Repair Expenditures: Manufacturing Sub-Industries, Canada* (catalogue 61-214) was released with estimates for 1976 and 1977.

The introduction of *The Daily* (catalogue 11-001), in 1980, signified the replacement of the Service Bulletin as the primary vehicle for disseminating mining industry and energy related industries capital expenditure data. Expenditures for the mining sector appeared in this format from 1980 to 1982.

Energy related data was incorporated into *Capital and Repair Expenditures: Manufacturing Sub-Industries, Canada* (catalogue 61-214) in 1981. Further developments in the production of manufacturing sub-industry data were achieved in 1982 with the publication of the historical series from 1960 to 1967, for 20 major groups and sub-industries, in *Investment Statistics: Manufacturing Sub-Industries, Canada* (catalogue 61-518).

The definition of capital expenditures, related to exploration and development in the mining sector, was expanded in 1982 to include field expenditures on all physical work and surveys and other related costs such as applied administration costs, general overhead and lease rental costs. *Investment Statistics: Exploration, Development,*

### 3.7 Chronologie de la série des dépenses en immobilisations

En 1941, le Bureau fédéral de la statistique donne le coup d'envoi à la première véritable série sur les dépenses en immobilisations en recueillant, entre autres, des données sur les dépenses en immobilisations provenant de certains secteurs. Les premières prévisions d'investissements sont publiées à l'automne 1946 sous le titre *Dépenses d'immobilisation, de réparation et d'entretien des entreprises commerciales au Canada : prévisions 1946*.

En 1947, le champ de la série des dépenses en immobilisations s'élargit avec l'inclusion des biens d'investissement imputés aux dépenses d'exploitation. Cette nouveauté accroît la précision des données en permettant l'estimation des biens qui comptent dans le stock de capital du pays sans être immobilisés par les entreprises répondantes.

Depuis 1946, la portée de l'enquête sur les dépenses en immobilisations s'est accrue pour englober de nouvelles branches de l'économie. Les dépenses en immobilisations des sous-industries minières et manufacturières sont publiées dans *Statistique des investissements - Bulletin de service* (catalogue 61-007) depuis 1975; pour les données sur l'énergie, la première parution remonte à 1976. Les données sur l'énergie publiées dans le numéro 2 du volume 2 de *Statistique des investissements - Bulletin de service* portaient sur l'année en cours et présentaient des estimations remontant à 1955.

En 1978 paraît le premier numéro de *Dépenses d'immobilisations et de réparations - Sous-industries manufacturières, Canada* (catalogue 61-214), avec des estimations pour 1976 et 1977.

Le *Quotidien* (catalogue 11-001), en 1980, remplace le *Bulletin de service* à titre de principal organe de diffusion des données sur les dépenses en immobilisations des secteurs des mines et de l'énergie. Les dépenses du secteur minier paraissent dans cette publication de 1980 à 1982.

Les données sur l'énergie sont greffées à *Dépenses d'immobilisations et de réparations - Sous-industries manufacturières, Canada* (catalogue 61-214) en 1981. La publication des données portant sur les sous-industries manufacturières connaît un nouveau tournant en 1982 avec la parution de la série historique portant sur les années 1960 à 1967, pour 20 grands groupes et sous-industries, dans *Statistique des investissements - Sous-industries manufacturières, Canada* (catalogue 61-518).

La définition de *dépenses en immobilisations*, en rapport avec l'exploration et la mise en valeur dans le secteur minier, est élargie en 1982 pour comprendre les frais de travaux physiques et de levés exécutés sur le terrain ainsi que les frais connexes tels que les frais de location à bail, les frais généraux et les frais d'administration passés en charges. *Statistique des*

*Capital and Repair Expenditures by Mining and Exploration Companies* (catalogue 61-216) was released for the first time in 1983.

In 1986, the 1985 Actual Survey was expanded to include asset detail on new assets, used assets, renovations/retrofit for both construction and machinery and equipment. This new survey format also included other data items such as the reason for disposal/sale/write-downs of fixed assets, age of assets, lives of assets, reasons for expenditure and gross book value of asset. In addition, non-military machinery and equipment expenditures were now included under Department of National Defence expenditures.

Catalogue 61-216 was expanded in 1987 to include detailed data from the petroleum and natural gas industry (dating back to 1985) and energy related industries, which were previously included in catalogue 61-214.

In line with the National Accounts capital expenditure requirements and the movement toward streamlined operations, Statistics Canada stopped collecting and publishing data on non-producing exploration companies in 1990. These data are now surveyed by Natural Resources Canada.

In 1993, the survey adopted the 1980 Standard Industrial Classification and merged catalogues 61-214 and 61-216 into *Private and Public Investment in Canada* (catalogues 61-205-XPB and 61-205-XPB).

The most recent changes start with the current 1995 Revised Forecast where a probability sample was almost entirely selected from the the Central Frame Data Base of the Business Register Division.

Data prior to 1956 are only available in hard copy form, while subsequent historical data are available on CANSIM or from the Investment and Capital Stock Division of Statistics Canada.

*investissements - Dépenses d'exploration, de mise en valeur, d'immobilisations et de réparations par les compagnies minières et d'exploration* (catalogue 61-216), paraît pour la première fois en 1983.

En 1986, l'enquête sur les données réelles de 1985 s'élargit pour comprendre des renseignements détaillés sur les biens nouveaux, les biens usagés, les rénovations et les réfections des constructions, du matériel et de l'outillage. Cette nouvelle formule vise aussi d'autres objets, notamment les motifs d'aliénation, de vente ou de réduction des immobilisations, l'âge des biens, la durée de vie des biens, les motifs des dépenses et la valeur comptable brute. De plus, le matériel ou l'outillage non militaires sont maintenant imputés au ministère de la Défense.

La portée de la publication n° 61-216 commence en 1987 à englober des données détaillées provenant des secteurs du pétrole et du gaz naturel (remontant à 1985) ainsi que de l'énergie, qui étaient auparavant comprises dans la publication n° 61-214.

Conformément aux exigences de la comptabilité nationale en matière de dépenses en immobilisations et à la tendance en faveur d'une simplification des opérations, Statistique Canada cesse en 1990 de recueillir et de publier des données sur les compagnies d'exploration non productrices. Ces données sont maintenant recensées par Ressources naturelles Canada.

En 1993, l'enquête adopta la Classification type des industries de 1980 et fusionna les publications n°s 61-214 et 61-216 dans *Investissements privés et publics au Canada*, (cat. 61-205-XPB et 61-205-XPB).

Les derniers changements commencent avec l'enquête courante les Prévisions révisées pour 1995 dont l'échantillon aléatoire fut presque entièrement sélectionné de la Base de données du registre central de la Division du Registre des entreprises.

Les données antérieures à 1956 existent sur papier seulement, mais les données historiques ultérieures peuvent être obtenues sur CANSIM ou à la Division de l'investissement et du stock de capital de Statistique Canada.