CONCEPTS, DEFINITIONS and DATA QUALITY

The Monthly Survey of Manufacturing (MSM) publishes statistical series for manufacturers – shipments, inventories, unfilled orders and new orders. The values of these characteristics represent current monthly estimates of the more complete Annual Survey of Manufactures (ASM) data.

The MSM is a sample survey of approximately 9,500 Canadian manufacturing establishments, which are categorized into over 200 industries. Industries are classified according to the 1997 North American Industrial Classification System (NAICS), which replaced the 1980 Standard Industrial Classification (SIC) system. Reference year 2000 is the last year for which data are released on a SIC basis. The MSM adopted the NAICS for its 2001 reference, while previous years' data have been re-calculated to the new classification system back to 1992. Seasonally adjusted series are available for the main aggregates.

An establishment comprises the smallest manufacturing unit capable of reporting the variables of interest. Data collected by the MSM provides a current 'snapshot' of shipment values by the Canadian manufacturing sector, enabling analysis of the state of the Canadian economy, as well as the health of specific industries in the short- to medium-term. The information is used by both private and public sectors including Statistics Canada, federal and provincial governments, business and trade entities, international and domestic non-governmental organizations, consultants, the business press and private citizens. The data are used for analyzing market share, trends, corporate benchmarking, policy analysis, program development, tax policy and trade policy.

1. Shipments

Shipments are defined as the value of goods manufactured by establishments that have been shipped to a customer. Shipments exclude any wholesaling activity, and any revenues from the rental of equipment or the sale of electricity. Note that in practice, some respondents report financial transactions rather than payments for work done. Shipments are available by 3-digit NAICS, broken down by province.

For the aerospace product and parts, and shipbuilding industries, the value of production is used instead of shipments. This value is calculated by adjusting monthly shipments by the monthly change in goods in process and finished product inventories. Raw materials are not included in the calculation since production tries to measure "work done" during the month. This is done in order to reduce distortions caused by the shipment of high value items as completed sales.

2. Inventories

Measurement of component values of inventory is important for economic studies as well as for derivation of production values. Respondents are asked to report their book values (at cost), of raw materials, any goods-in-process, and finished-product inventories separately. In some cases, respondents estimate a total inventory figure, which is allocated on the basis of proportions reported on the ASM. Inventory levels are calculated on a Canada-wide basis, not by province.

3. Orders

a) Unfilled Orders

Unfilled orders represent a backlog or stock of orders that will generate future shipments assuming that they are not cancelled. As with inventories, unfilled orders and new orders levels are calculated on a Canada-wide basis, not by province.

The MSM produces estimates for unfilled orders for all industries except for those industries where orders are customarily filled from stocks on hand and order books are not generally maintained. In the case of the aircraft companies, options to purchase are not treated as orders until they are entered into the accounting system.

b) New Orders

New orders represent current demand for manufactured products. Estimates of new orders are derived from shipments and unfilled orders data. All shipments within a month result from either an order received during the month or at some earlier time. New orders can be calculated as the sum of shipments adjusted for the monthly change in unfilled orders.

4. Non-Durable / Durable goods

a) Non-durable goods industries

Non-durable goods industries include Food (NAICS 311), Beverage and Tobacco Products (312), Textile Mills (313), Textile Product Mills (314), Clothing (315), Leather and Allied Products (316), Paper (322), Printing and Related Support Activities (323), Petroleum and Coal Products (324), Chemicals (325) and Plastic and Rubber Products (326).

b) Durable goods industries

Durable goods industries include Wood Products (NAICS 321), Non-Metallic Mineral Products (327), Primary Metals (331), Fabricated Metal Products (332), Machinery (333), Computer and Electronic Products (334), Electrical Equipment, Appliance and Components (335), Transportation Equipment (336), Furniture and Related Products (337) and Miscellaneous Manufacturing (339).

SURVEY DESIGN AND METHODOLOGY

Beginning with the August 1999 reference month, the Monthly Survey of Manufacturing (MSM) has undergone an extensive redesign.

Concept Review

It was decided that before any redesign work could begin the basic concepts and definitions of the program would be confirmed.

This was done in two ways: First, a review of user requirements was initiated. This involved revisiting an internal report to ensure that the user requirements from that exercise were being satisfied. As well, another round of internal review with the major users in the National Accounts was undertaken. This was to specifically focus on any data gaps that could be identified.

Secondly, with these gaps or requirements in hand, a survey was conducted in order to ascertain respondent's ability to report existing and new data. The study was also to confirm that respondents understood the definitions, which were being asked by survey analysts.

The result of the concept review was a reduction of the number of questions for the survey from sixteen to seven. Most of the questions that were dropped had to do with the reporting of shipments for work that was partially completed.

Methodology

The new sample design incorporates the 1997 North American Industrial Classification Standard (NAICS) and gives a much higher profile to provincial estimates. Stratification is done by province with equal quality requirements for each province. Large size units are selected with certainty and small units are selected with a probability based on the desired quality of the estimate within a cell.

The opportunity was also taken at this time to allow for the introduction of sample rotation into the survey design. Most of the smaller companies who are asked to participate in the survey will do so only for a set period.

The estimation system generates estimates using the NAICS. The estimates will also continue to be reconciled to the ASM. Provincial estimates for all variables will be produced. A measure of quality (CV) will also be produced.

Components of the Redesigned Survey

Target Population and Sampling Frame

Statistics Canada's business register provides the sampling frame for the MSM. The target population for the MSM consists of all statistical establishments on the business register that are classified to the manufacturing sector (by NAICS). The sampling frame for the MSM is determined from the target population after subtracting establishments that represent the bottom 2% of the total manufacturing shipments estimate for each province. These establishments were excluded from the frame so that the sample size could be reduced without significantly affecting quality.

The Sample

The MSM sample is a probability sample comprised of approximately 9,500 establishments.

Prior to selection, the sampling frame is subdivided into industry-province cells. For the most part, NAICS codes were used. Depending upon the number of establishments within each cell, further subdivisions were made to group similar sized establishments' together (called stratum). An establishment's size was based on its most recently available annual shipments or sales value.

Each industry by province cell has a 'take-all' stratum composed of establishments sampled each month with certainty. This 'take-all' stratum is composed of establishments that are the largest statistical enterprises, and have the largest impact on estimates within a particular industry by province cell. These large statistical enterprises comprise 45% of the national manufacturing shipment estimates.

Each industry - province cell can have at most three 'take-some' strata. Not all establishments within these stratums need to be sampled with certainty. A random sample is drawn from the remaining strata. The responses from these sampled establishments are weighted according to the inverse of their probability of selection.

The initial sample was selected in late 1998 and has been refreshed each month by including a sample of new entrants in the frame.

Data Collection

Only a subset of the sample establishments is sent out for data collection. For the remaining units, information from administrative data files is used as a source for deriving shipment data. For those establishments that are surveyed, data collection, data capture, preliminary edit and follow-up of non-respondents are all performed in Statistics Canada regional offices. Sampled establishments are contacted by mail or telephone according to the preference of the respondent. Data capture and preliminary editing are performed simultaneously to ensure the validity of the data.

In some cases, combined reports are received from enterprises or companies with more than one establishment in the sample where respondents prefer not to provide individual establishment reports. Businesses, which do not report or whose reports contain errors, are followed up immediately.

Use of Administrative Data

Managing response burden is an ongoing challenge for Statistics Canada. In an attempt to alleviate response burden, especially for small businesses, STC has been investigating various alternatives to survey taking. Administrative data files are a rich source of information for business data and STC is working at mining this rich data source to its full potential. As such, effective the August 2004 reference month, the MSM has reduced the number of simple establishments in the sample that are surveyed directly and instead, derives shipments data for these establishments from Goods and Services Tax (GST) files using a statistical model. The model accounts for the difference between shipments and sales (reported for GST purposes) as well as the time lag between the reference period of the survey and the reference period of the GST file.

Inventories and unfilled orders estimates for establishments where shipments are GST-based are derived using the MSM's imputation system. The imputation system applies to the previous month values, the month-to-month and year-to-year changes in similar firms which are surveyed.

Detailed information on the methodology used for modelling shipment from administrative data sources can be found in the '***Monthly Survey of Manufacturing: Use of Administrative Data'*** **(Catalogue no. 31-533-XIE)** document.

DATA QUALITY

Statistical Edit and Imputation

Data are analyzed within each industry-province cell. Extreme values are listed for inspection by the magnitude of the deviation from average behavior. Respondents are contacted to verify extreme values. Records that fail statistical edits are considered outliers and are not used for imputation.

Values are imputed for the non-responses, for establishments that do not report or only partially complete the survey form. A number of imputation methods are used depending on the variable requiring treatment. Methods include using industry-province cell trends, historical responses, or reference to the ASM. Following imputation, the MSM staff performs a final verification of the responses that have been imputed.

Revisions

In conjunction with preliminary estimates for the current month, estimates for the previous three months are revised to account for any late returns. Data are revised when late responses are received or if an incorrect response was reported earlier.

Estimation

Estimates are calculated by multiplying an estimation weight to an establishment's reported responses. The estimation weight is the inverse of the sampled establishment's probability of selection. Take all units are self-representative.

Benchmarking

In April 2006, the Annual Survey of Manufactures (ASM) released preliminary estimates for reference year 2004 and revisions for 2003. To incorporate the new ASM estimates, as of September 2006, the Monthly Survey of Manufacturing data were benchmarked back to January 2002.

Although the historical month-to-month movements were preserved, there were minor adjustments made to the levels.

The Monthly Survey of Manufacturing (MSM) benchmarks to the ASM in order to adjust for new, revised or missing data, to incorporate changes to NAICS classifications and to update the seasonal adjustment factors. Data are revised back to the benchmark year and will affect monthly levels in subsequent months.

Starting with reference year 2000, the ASM incorporated some significant conceptual and methodological changes. The most important change was the expansion to include all manufacturing establishments in Canada. Previously only incorporated establishments that had employees and had sales greater than or equal to $30,000 were covered by the ASM.

Sampling and Non-sampling Errors

The statistics in this publication are estimates derived from a sample survey and, as such, can be subject to errors. The following material is provided to assist the reader in the interpretation of the estimates published.

Estimates derived from a sample survey are subject to a number of different kinds of errors. These errors can be broken down into two major types: sampling and non-sampling.

1. Sampling Errors

Sampling errors are an inherent risk of sample surveys. They result from the difference between the value of a variable if it is randomly sampled and its value if a census is taken (or the average of all possible random values). These errors are present because observations are made only on a sample and not on the entire population.

The sampling error depends on factors such as the size of the sample, variability in the population, sampling design and method of estimation. For example, for a given sample size, the sampling error will depend on the stratification procedure employed, allocation of the sample, choice of the sampling units and method of selection. (Further, even for the same sampling design, we can make different calculations to arrive at the most efficient estimation procedure.) The most important feature of probability sampling is that the sampling error can be measured from the sample itself.

2. Non-sampling Errors

Non-sampling errors result from a systematic flaw in the structure of the data-collection procedure or design of any or all variables examined. They create a difference between the value of a variable obtained by sampling or census methods and the variable's true value. These errors are present whether a sample or a complete census of the population is taken. Non-sampling errors can be attributed to one or more of the following sources:

a) Coverage error: This error can result from incomplete listing and inadequate coverage of the population of interest.

b) Data response error: This error may be due to questionnaire design, the characteristics of a question, inability or unwillingness of the respondent to provide correct information, misinterpretation of the questions or definitional problems.

c) Non-response error: Some respondents may refuse to answer questions, some may be unable to respond, and others may be too late in responding. Data for the non-responding units can be imputed using the data from responding units or some earlier data on the non-responding units if available.

The extent of error due to imputation is usually unknown and is very much dependent on any characteristic differences between the respondent group and the non-respondent group in the survey. This error generally decreases with increases in the response rate and attempts are therefore made to obtain as high a response rate as possible.

d) Processing error: These errors may occur at various stages of processing such as coding, data entry, verification, editing, weighting, and tabulation, etc. Non-sampling errors are difficult to measure. More important, non-sampling errors require control at the level at which their presence does not impair the use and interpretation of the results.

Measures have been undertaken to minimize the non-sampling errors. For example, units have been defined in a most precise manner and the most up-to-date listings have been used. Questionnaires have been carefully designed to minimize different interpretations. As well, detailed acceptance testing has been carried out for the different stages of editing and processing and every possible effort has been made to reduce the non-response rate as well as the response burden.

Measures of Sampling and Non-sampling Errors

1. Sampling Error Measures

The sample used in this survey is one of a large number of all possible samples of the same size that could have been selected using the same sample design under the same general conditions. If it was possible that each one of these samples could be surveyed under essentially the same conditions, with an estimate calculated from each sample, it would be expected that the sample estimates would differ from each other.

The average estimate derived from all these possible sample estimates is termed the expected value. The expected value can also be expressed as the value that would be obtained if a census enumeration were taken under identical conditions of collection and processing. An estimate calculated from a sample survey is said to be precise if it is near the expected value.

Sample estimates may differ from this expected value of the estimates. However, since the estimate is based on a probability sample, the variability of the sample estimate with respect to its expected value can be measured. The variance of an estimate is a measure of the precision of the sample estimate and is defined as the average, over all possible samples, of the squared difference of the estimate from its expected value.

The standard error is a measure of precision in absolute terms. The coefficient of variation, defined as the standard error divided by the sample estimate, is a measure of precision in relative terms. For comparison purposes, one may more readily compare the sampling error of one estimate to the sampling error of another estimate by using the coefficient of variation.

In this publication, the coefficient of variation is used to measure the sampling error of the estimates. However, since the coefficient of variation published for this survey is calculated from the responses of individual units, it also measures some non-sampling error.

The formula used to calculate the published coefficients of variation (CV) in Table 1 is:

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

where X denotes the estimate and S(X) denotes the standard error of X.

In this publication, the coefficient of variation is expressed as a percentage.

Confidence intervals can be constructed around the estimate using the estimate and the coefficient of variation. Thus, for our sample, it is possible to state with a given level of confidence that the expected value will fall within the confidence interval constructed around the estimate. For example, if an estimate of $12,000,000 has a coefficient of variation of 10%, the standard error will be $1,200,000 or the estimate multiplied by the coefficient of variation. It can then be stated with 68% confidence that the expected value will fall within the interval whose length equals the standard deviation about the estimate, i.e., between $10,800,000 and $13,200,000. Alternatively, it can be stated with 95% confidence that the expected value will fall within the interval whose length equals two standard deviations about the estimate, i.e., between $9,600,000 and $14,400,000.

Text table 1 contains the national level CVs, expressed as a percentage, for all manufacturing for the MSM characteristics. For CVs at other aggregate levels, contact the Marketing and Dissemination Section at (613) 951-9497, toll free: 1-866-873-8789 or by e-mail at manufact@statcan.ca.

Text table 1: National Level CVs by Characteristic

MONTH	Shipments %	Raw Material Inventories %	Goods in Process Inventories %	Finished Product Inventories %	Unfilled Orders %
May 2006	0.60	1.00	1.68	1.29	1.92
June 2006	0.62	1.04	1.91	1.30	1.88
July 2006	0.58	1.03	2.11	1.29	1.90
August 2006	0.62	1.06	1.97	1.32	2.02
September 2006	0.64	1.04	1.49	1.41	1.97
October 2006	0.83	1.06	1.04	1.39	2.22
November 2006	0.88	1.04	1.04	1.29	2.22
December 2006	0.61	1.16	1.43	1.26	1.28
January 2007	0.60	1.06	1.50	1.22	1.32
February 2007	0.61	1.07	1.53	1.17	1.27
March 2007	0.58	1.06	1.59	1.35	1.27
April 2007	0.59	1.04	1.65	1.26	1.37
May 2007	0.61	1.05	1.60	1.30	1.48

2. Non-sampling Error Measures

The exact population value is aimed at or desired by both a sample survey as well as a census. We say the estimate is accurate if it is near this value. Although this value is desired, we cannot assume that the exact value of every unit in the population or sample can be obtained and processed without error. Any difference between the expected value and the exact population value is termed the bias. Systematic biases in the data cannot be measured by the probability measures of sampling error as previously described. The accuracy of a survey estimate is determined by the joint effect of sampling and non-sampling errors.

Three sources of non-sampling error in the MSM are non-response error, imputation error and the error due to editing. To assist users in evaluating these errors, weighted rates that are related to these three types of error are given in Text table 2. The following is an example of what is meant by a weighted rate. A cell with a sample of 20 units in which five respond for a particular month would have a response rate of 25%. If these five reporting units represented $8 million out of a total estimate of $10 million, the weighted response rate would be 80%.

The definitions of the three weighted rates noted in Text table 2 follow. The weighted response rate is the proportion of a characteristic's total estimate that is based upon reported data (excluding data that has been edited). The weighted imputation rate is the proportion of a characteristic's total estimate that is based upon imputed data. The weighted editing rate is the proportion of a characteristic's total estimate that is based upon data that was edited (edited data may have been originally reported or imputed).

Text table 2 contains the three types of weighted rates for each of the characteristics at the national level for all of manufacturing. In the table, the rates are expressed as percentages.

Text Table 2: National Weighted Rates by Source and Characteristic

Characteristics	Survey Source			Administrative Data Source		
	Response	Imputation	Editing	Modeled	Imputation	Editing
	%	%	%	%	%	%
Shipments	84.79	4.28	2.69	6.79	1.33	0.11
Raw Materials	77.14	14.85	3.37	0	4.57	0.06
Goods in Process	64.80	8.20	23.00	0	3.90	0.10
Finished Products	81.17	10.81	3.05	0	4.24	0.72
Unfilled Orders	46.06	3.02	49.00	0	1.83	0.09

Joint Interpretation of Measures of Error

The measure of non-response error as well as the coefficient of variation must be considered jointly to have an overview of the quality of the estimates. The lower the coefficient of variation and the higher the weighted response rate, the better will be the published estimate.

Seasonal Adjustment

Economic time series contain the elements essential to the description, explanation and forecasting of the behavior of an economic phenomenon. They are statistical records of the evolution of economic processes through time. In using time series to observe economic activity, economists and statisticians have identified four characteristic behavioral components: the long-term movement or trend, the cycle, the seasonal variations and the irregular fluctuations. These movements are caused by various economic, climatic or institutional factors. The seasonal variations occur periodically on a more or less regular basis over the course of a year. These variations occur as a result of seasonal changes in weather, statutory holidays and other events that occur at fairly regular intervals and thus have a significant impact on the rate of economic activity.

In the interest of accurately interpreting the fundamental evolution of an economic phenomenon and producing forecasts of superior quality, Statistics Canada uses the X11ARIMA/88 seasonal adjustment method to seasonally adjust its time series. This method minimizes the impact of seasonal variations on the series and essentially consists of adding one year of estimated raw data to the end of the original series before it is seasonally adjusted per se. The estimated data are derived from forecasts using ARIMA (Auto Regressive Integrated Moving Average) models of the Box-Jenkins type.

The X-11 part of the X11ARIMA/88 program uses primarily a ratio-to-moving average method. It is used to smooth the modified series and obtain a preliminary estimate of the trend-cycle. It also calculates the ratios of the original series (fitted) to the estimates of the trend-cycle and estimates the seasonal factors from these ratios. The final seasonal factors are produced only after these operations have been repeated several times.

The procedures to determine the seasonal factors necessary to calculate the final seasonally adjusted data are executed every month. This approach ensures that the estimated seasonal factors are derived from an unadjusted series that includes all the available information about the series, i.e. the current month's unadjusted data as well as the previous month's revised unadjusted data.

While seasonal adjustment permits a better understanding of the underlying trend-cycle of a series, the seasonally adjusted series still contains an irregular component. Slight month-to-month variations in the seasonally adjusted series may be simple irregular movements. To get a better idea of the underlying trend, users should examine several months of the seasonally adjusted series.

The Canada seasonally adjusted total is derived indirectly by the summation of the individually seasonally adjusted kinds of business.

Trend

A seasonally adjusted series may contain the effects of irregular influences and special circumstances and these can mask the trend. The short term trend shows the underlying direction in seasonally adjusted series by averaging across months, thus smoothing out the effects of irregular influences. The result is a more stable series. The trend for the last month may be, subject to significant revision as values in future months are included in the averaging process.

Real Manufacturing Shipments, Inventories, and Orders

Changes in the values of the data reported by the Monthly Survey of Manufacturing (MSM) may be attributable to changes in their prices or to the quantities measured, or both. To study the activity of the manufacturing sector, it is often desirable to separate out the variations due to price changes from those of the quantities produced. This adjustment is known as deflation.

Deflation consists in dividing the values at current prices obtained from the survey by suitable price indexes in order to obtain estimates evaluated at the prices of a previous period, currently the year 1997. The resulting deflated values are said to be “at 1997 prices”. Note that the expression “at current prices” refer to the time the activity took place, not to the present time, nor to the time of compilation.

The deflated MSM estimates reflect the prices that prevailed in 1997. This is called the base year. The year 1997 was chosen as base year since it corresponds to that of the price indexes used in the deflation of the MSM estimates. Using the prices of a base year to measure current activity provides a representative measurement of the current volume of activity with respect to that base year. Current movements in the volume are appropriately reflected in the constant price measures only if the current relative importance of the industries is not very different from that in the base year.

The deflation of the MSM estimates is performed at a very fine industry detail, equivalent to the 6-digit industry classes of the North American Industry Classification System (NAICS). For each industry at this level of detail, the price indexes used are composite indexes which describe the price movements for the various groups of goods produced by that industry.

With very few exceptions the price indexes are weighted averages of the Industrial Product Price Indexes (IPPI). The weights are derived from the annual Canadian Input-Output tables and change from year to year. Since the Input-Output tables only become available with a delay of about two and a half years, the weights used for the most current years are based on the last available Input-Output tables.

The same price index is used to deflate shipments, new orders and unfilled orders of an industry. The weights used in the compilation of this price index are derived from the output tables, evaluated at producer’s prices. Producer prices reflect the prices of the goods at the gate of the manufacturing establishment and exclude such items as transportation charges, taxes on products, etc. The resulting price index for each industry thus reflects the output of the establishments in that industry.

The price indexes used for deflating the goods-in-process and the finished goods inventories of an industry are moving averages of the price index used for shipments. For goods-in-process inventories, the number of terms in the moving average corresponds to the duration of the production process. The duration is calculated as the average over the previous 48 months of the ratio of end of month goods-in-process inventories to the output of the industry, which is equal to shipments plus the changes in both goods-in-process and finished goods inventories.

For finished goods inventories, the number of terms in the moving average reflects the length of time a finished product remains in stock. This number, known as the inventory turnover period, is calculated as the average over the previous 48 months of the ratio of end-of-month finished goods inventory to shipments.

To deflate raw materials inventories, price indexes for raw materials consumption are obtained as weighted averages of the IPPIs. The weights used are derived from the input tables evaluated at purchaser’s prices, i.e. these prices include such elements as wholesaling margins, transportation charges, and taxes on products, etc. The resulting price index thus reflects the cost structure in raw materials for each industry.

The raw materials inventories are then deflated using a moving average of the price index for raw materials consumption. The number of terms in the moving average corresponds to the rate of consumption of raw materials. This rate is calculated as the average over the previous 4 years of the ratio of end-of-year raw materials inventories to the intermediate inputs of the industry.