

THE REDESIGNED CANADIAN MONTHLY WHOLESALE AND RETAIL TRADE SURVEY: A POSTMORTEM OF THE IMPLEMENTATION

Julie Trépanier, Statistics Canada

Julie Trépanier, Business Survey Methods Division, Statistics Canada, Ottawa, Canada K1A 0T6

Keywords: Retail Trade, Wholesale Trade, Size Stratification, NAICS, Births, Deaths, Misclassifications, Evaluation.

1. INTRODUCTION

The Monthly Wholesale and Retail Trade Survey (MWRTS) is one of the mission critical surveys conducted by Statistics Canada (SC). It collects monthly retail and wholesale sales and inventories at the industrial and geographical levels. On a value-added basis, these two industries constitute approximately 12% of the Gross Domestic Product (GDP). It uses the same sample month after month with the exception of a sample of births that is added monthly.

The last major redesign of the MWRTS was completed in 1988 using the 1980 Standard Industrial Classification (SIC) system. With the implementation of the North American Industry Classification System (NAICS) in 1997, it was obvious that the MWRTS needed to be redesigned as a NAICS-based survey. Other objectives were to improve the coverage of non-employer businesses, to take advantage of the new Goods and Services Tax (GST) data and to incorporate innovative solutions at the sampling and estimation stages to maintain the quality of the survey over time. In addition, the processing systems, which were based mostly on a mainframe computer environment, were becoming costly and inefficient. Under the Project to Improve Provincial Economic Statistics (PIPES) and its generic Unified Enterprises Survey (UES), SC's annual business survey program (including retail and wholesale annual surveys) had also been redesigned in the late 1990s. As well, consistency between the MWRTS and the UES methodology needed to be improved. Reducing response burden where possible was a priority to compensate for the expansion, both in coverage and details, of the annual surveys.

The redesign started with a feasibility study that was launched in the fiscal year 1999-2000. The newly available Goods and Services Tax (GST) data were studied to determine how they could improve sampling and estimation for the MWRTS. The SC GST processing system was still being improved at that time. Consequently, the study concluded that, for

the moment, annual GST sales should only be used in the size stratification, an area needing improvement in MWRTS. The sampling design methodology (Bérard 2001, Majkowski 2001) was developed in 2000-01 followed by the development of the editing and imputation processes in 2001-02. Estimation, including an outlier detection and treatment strategy (Matthews and Bérard 2002), was also developed. The fiscal year 2002-03 served to fine-tune the processes and develop a backcasting approach for NAICS estimates (Fortier 2003). The new MWRTS sample was selected in April 2003. The fiscal year 2003-04, the fourth and final year of the redesign project, also permitted the completion of diagnostic tools to monitor processes (Majkowski et al. 2004). A complete test of the new MWRTS in parallel with the old MWRTS was performed from December 2003 to April 2004 reference months. The new MWRTS released in June 2004 its first estimates for the April 2004 reference month.

This paper describes the sampling design implementation and assumptions, as well as the estimation in section 2. The new estimates are compared to the old MWRTS estimates in section 3 followed by an assessment of the assumptions made at the sampling design stage. The conclusion presents general observations on the redesign and aspects that we would keep and those we would change if we were to start over the project.

2. SAMPLING DESIGN AND ESTIMATION

2.1 Sampling Design

The April 2003 sampling design is described in this section. The resulting sample is used to collect retail sales, wholesale sales and wholesale inventories, but the sampling design focuses on sales primarily. Retail inventories are collected for a portion of the sample only.

The frame is extracted from SC's Business Register (BR). The BR provides a four-level hierarchical statistical structure starting at the top with the enterprise, then the company, the establishment and the location. The MWRTS *frame* is made of *establishments, both employers and non-employers*, for which the first two digits of their NAICS (i.e., sector) are 41 (wholesale), 44 or 45 (retail). Excluded

from this frame are retail establishments with NAICS 4541 (electronic shopping and mail-order houses), 4542 (vending machine operators), 45431 (fuel dealers) or 45439 (other direct selling establishments), and wholesale establishments with NAICS 41112 (oilseed and grain), 412 (petroleum products) or 419 (agents and brokers). Apart from being NAICS-based, the two major differences between the new and old MWRTS frames are:

- 1) non-employer establishments are now included. In 1988 no frame existed for them. A constant and outdated adjustment for their undercoverage was made to the retail estimates only;
- 2) the frame is built at the establishment level for consistency with the UES (previously built at the location level for retail).

Sampling units are defined as *clusters of establishments within the same stratification trade group (TG), the same stratification geographical region (GEO) and under the same enterprise*. This differs from the old MWRTS where the sampling units were the companies of the retail locations or wholesale establishments. For retail, 24 TGs based on 3- to 5-digit NAICS and 16 GEOs, i.e., the 13 provinces and territories with 3 separate regions for Montreal, Toronto and Vancouver, are used. For wholesale, there are 16 TGs based on 3- to 5-digit NAICS and 13 GEOs (provinces and territories).

The clusters of establishments are first stratified by **TG and GEO** as defined above. An annual size measure is then associated with each cluster. In the old MWRTS, an estimated measure of the Gross Business Income (GBI) provided by the BR was used to stratify the companies by size. Although the GBI is overall of good quality, the GBI of some large companies was sometimes underestimated. As a result, some large companies were assigned to take-some strata with large weights, creating an undesired impact on the estimates. An objective of the new MWRTS was to improve size stratification by using other data sources such as annual sales provided by GST. A study performed in 2000-2001 (Bérard 2001) showed that the maximum of the GBI, annual GST sales and revenue available for the corporate income tax data (T2-revenue) – called TRIO, would produce a more efficient size stratification. A 10% reduction in size misclassification for employers was expected. TRIO seemed particularly efficient in identifying large units. Survey data from the old MWRTS and the UES could also be used for size stratification since the April 2003 sample was to be selected independently from the samples of these two surveys (no control of overlap). In summary, the **size measure** for a given establishment is set to: *I*)

annual sales from the 2002 MWRTS; else 2) annual sales from the 1999 or 2000 UES; else 3) TRIO=Max(GBI, GST sales, T2-revenue). The cluster size measure is then equal to the sum of the size measures of its establishments. Statistics on the use of each source are presented in Table 3 later in this section. Section 3 presents the performance of this size measure now that monthly sales are available for the sampled clusters.

Once stratified by TG and GEO, the cluster's size measure is used to identify the non-surveyed portion of the survey frame or, if one prefers, to build the **"take-none" strata**. This portion of the population will be estimated by other data sources such as GST. Using a pre-determined set of thresholds (often referred to at SC as the Royce-Maranda thresholds), a threshold is chosen from this set for each combination of TG x GEO so that the clusters with a size measure below this threshold do not represent more than 5% of the total size measure of the TG x GEO. These clusters comprise the take-none stratum of that TG x GEO. When multi-cluster enterprises have clusters that fall both in and out of the take-none strata, all of the clusters are forced into the surveyed portion. This is to ensure that multi-cluster enterprises are either totally excluded or totally surveyed. Tables 2A and 2B show that overall 45.1% of the retail clusters and 68.1% of the wholesale clusters are in the non-surveyed portion. This is a significant effort to reduce response burden.

Before stratifying the surveyed portion of the frame, particular clusters such as those in multi-cluster enterprises are **"pre-specified" take-all** in their TG x GEO. Based on the remaining number of clusters in the surveyed portion of the TG x GEO, the following **number of size strata** is usually formed:

- 10 clusters or less : 1 take-all stratum
- 11 to 50 clusters : 1 take-all stratum and 1 take-some stratum
- More than 50 clusters: 1 take-all stratum and 2 take-some strata ("large" and "small").

An algorithm developed by Lavallée and Hidirolou (L-H) (1988) is widely used for the **size stratification** of SC business surveys. It provides an iterative algorithm that can stratify a highly skewed population into a take-all stratum and a number of take-some strata while minimizing the sample size required for a given level of relative precision. It assumes simple random sampling without replacement (SRSWOR) within the take-some strata, and any form of power allocation of the total sample size to the strata. The L-H algorithm needs a size measure highly correlated with the survey variables

of interest. The new MWRTS uses the size measure as described earlier (referred to as X) and a sample allocation proportional to the square root of the size measure is specified. In the old MWRTS, the same sample allocation was used but the size measure was the GBI. The take-all strata were established by a method from Hidirolou (1986). Boundaries between the two take-some strata were not optimal, being based on some fixed BR boundaries.

For the new MWRTS, a modified version of the L-H algorithm (Ferland 2003) is used in each TG x GEO. It is very similar to the original L-H method but it takes into account **expected rates of out-of-business (OOB) units** in the sample. OOB units, still unknown to the BR and thus present as active on the BR, can fall on the MWRTS frame and sample. They will be found inactive by the survey and perhaps by signals from administrative sources. In the interim, one can only estimate their percentage. These rates are **10% for large take-some strata** and **20% for small take-some strata** and are based on old MWRTS results. (Please note that the out-of-scope rate (the rate of active clusters that are expected to be found out of retail and wholesale) could have been considered here as well but it was not.) The L-H algorithm is modified so that the function to minimize in each TG x GEO in the Lavallée and Hidirolou paper:

$$n = N_L + \frac{\sum_{h=1}^{L-1} N_h^2 S_h^2 / a_h}{(Nc\bar{Y})^2 + \sum_{h=1}^{L-1} N_h S_h^2}$$

is replaced by :

$$n = N_L + \frac{\sum_{h=1}^{L-1} p_h N_h^2 (S_h^2 + (1 - p_h) \bar{Y}_h^2) / a_h}{c^2 (\sum_{h=1}^L p_h \bar{Y}_h)^2 + \sum_{h=1}^{L-1} p_h N_h (S_h^2 + (1 - p_h) \bar{Y}_h^2)}$$

where n is the total sample size to be minimized, h represents the stratum with L being the take-all stratum and others being take-some strata, N_h is the population size of h , N is the total population size, S_h^2 is the population variance of h , a_h represents the sample size allocation formula to the take-some

stratum h , i.e., $a_h = \sqrt{X_h} / \sum_{h=1}^{L-1} \sqrt{X_h}$ in the

MWRTS case, p_h is the expected OOB rate in h , and c is the target coefficient of variation. This form of n taking into account the OOB rates was mentioned in Latouche (1988). It takes into account the effect of such units (\$0) on the resulting variance of the estimate. Integrating it into the L-H algorithm ensures that size stratification and sample size

determination are optimally obtained to achieve the expected CV c although OOB rates p_h could be observed in the sample once surveyed.

The **target coefficients of variation (CVs)** for the new MWRTS sales are presented in Table 1. The old MWRTS had higher target CVs for wholesale (e.g., 1.7% national).

Table 1: Target CVs for Sales

| Level | CV (%) |
|--------------------------|--------|
| National All Industries | 1.2 |
| Provincial / Territorial | 2.5 |
| TG Published | |
| - Priority | 2.5 |
| - Non- Priority | 3.5 |
| TG Stratification X GEO | 16.5 |

The 24 and 16 stratification TGs for retail and wholesale make respectively 19 and 15 published TGs. There are more stratification than published TGs in order to create more homogeneous strata and to satisfy needs of the Quarterly Retail Commodity Survey (a second-phase sample selected from the first-phase retail sample that collects commodity distribution). Some TGs, like new car dealers and grocery stores in retail, are considered priority TGs due to their importance in the national estimate.

Two **adjustments** to the **sample sizes** of take-some strata provided by the modified L-H algorithm are planned to avoid undesired impact on the estimates or their variances: 1) establishing minimum sample size; and 2) capping design weights. Since these adjustments increase the final sample size and consequently decrease the expected CVs, higher CVs than those presented in Table 1 are used in the modified L-H. Before describing these adjustments, one must understand a second series of adjustments that is performed: oversampling for frame misclassifications and nonresponse.

Contrary to minimum sample size and maximum weight rules that decrease the expected CVs, **oversampling for frame misclassifications and nonresponse** is applied to prevent the observed CVs from being greater than expected due to misclassifications and nonresponse. Although they artificially increase the sample sizes given by the modified L-H method, one should not adjust the expected CVs used as an input to the L-H method. Ideally, like the expected OOB rates, expected misclassification and nonresponse rates should be integrated into the modified L-H method so that size stratification and sample size determination remain optimally obtained. However, in the new MWRTS, expected frame misclassification and nonresponse

rates are only used to inflate the sample sizes after the modified L-H. This inflation is the last step and its rates are derived as follows.

- 1) **Frame misclassifications:** Out-of-scope cases and movements between TG and GEO are handled in an unbiased way by domain estimation. However, they can affect the CVs of the estimates as a result of stratification inefficiency. Based on UES 2000, the expected rate of out-of-scope units (live units moving out of retail and wholesale) is set to 15%. Movements between TGs are on average 2.7% for retail and 0.6% for wholesale but vary by TG. GEO misclassifications are expected to rarely occur. ***An overall frame misclassification rate of 17% (15% for out-of-scopes plus 2% for TG movements)*** is assumed. The rates vary from 10% to 35% by TG.
- 2) **Nonresponse:** A ***10% nonresponse rate*** was observed on average in the old MWRTS. That rate is assumed for the new MWRTS.

Knowing the above oversampling rates, the adjustments for minimum sample size and maximum design weights are first performed as follows.

- 1) Ultimately ***a minimum sample size of 8 clusters*** (10 in some cases) is desired. Knowing the combined oversampling rate in a stratum h for misclassifications and nonresponse, say r_h , the temporary minimum sample size to use before the oversampling in h is $8 / (1 + r_h)$, truncated to the integer below. For example, if $r_h=30\%$, the minimum sample size is the integer of $(8/1.30)=6$.
- 2) Ultimately sample sizes should ensure that ***the design weight does not exceed 10 for large take-some strata and 30 for small take-some strata.*** (The maximum is lower for some TG x GEO.) As in 1), knowing r_h , the temporary maximum design weight to use before the oversampling for h is $10 * (1 + r_h)$, rounded up to the next integer (assuming h is a large take-some stratum). For example, if $r_h=30\%$, the maximum weight is round $(10*1.30)=13$.

In the old MWRTS, final design weights of 15 in the large take-some strata and 50 in the small take-some strata were not rare, causing a large impact on the estimates when associated with a misclassified company. Lower weights were desirable in the new MWRTS.

In summary, the size stratification and sample size determination of the new MWRTS are performed in the following iterative way.

- 1) The modified L-H (including the OOB rates) is run by TG x GEO based on :
 - Prespecified take-all clusters
 - CVs for the TG x GEO
 - 1 take-all stratum, 1 or 2 take-some strata
 - Sample allocation proportional to the square root of the size measure
 - SRSWOR within strata
- 2) Minimum sample size and maximum weight adjustments are made to the resulting modified L-H sample sizes.
- 3) Expected CVs by TG, GEO and TG x GEO are computed given the sample sizes in 2) for all TG x GEO and taking into account OOB rates used in 1) and their impact on the variance of the estimates. If any of the expected CVs are above the target CVs (see Table 1), steps 1), 2) and 3) are repeated with appropriate new CVs by TG x GEO.
- 4) The TG x GEO population distribution and size stratification thresholds are examined graphically. For example, if the large take-some stratum population is still highly skewed, the threshold between the large take-some and take-all strata may be lowered. The revised threshold is then forced in step 1). Steps 2) and 3) are repeated. (Please note that 46% of the thresholds were lowered during this graphical analysis).
- 5) Oversampling for frame misclassifications and nonresponse is finally performed.

Systematic sampling within strata was in fact used to select the April 2003 sample although stratified SRSWOR is assumed at estimation. Before sample selection, clusters were ordered based on their size measure. This technique was adopted to avoid an extreme sample (i.e., a sample either made of very large units or very small units). This is particularly important when the sample is not changed frequently except for the addition of a monthly sample of births.

The results of the April 2003 sampling are presented below in Tables 2A and 2B. A large proportion of the sample (46.1% for retail and 72.2% for wholesale) is made of take-all clusters. This is largely due to the prespecification of take-all clusters, mostly from multi-cluster enterprises. The distribution of employer and non-employer clusters is different in the surveyed and non-surveyed portions. Non-employer clusters are more present in the non-surveyed portion (smaller clusters). This explains why in Table 3 the source of the size measure is less often based on the old MWRTS data in the non-surveyed portion since the old MWRTS did not survey non-employers.

Table 2A: April 2003 Retail Sampling

| Strata | Population | | Sample | |
|------------------------|----------------|-------------|---------------|-------------|
| | Counts | % | Counts | % |
| Take-All : | | | | |
| -Pre-specified | 4,045 | 2.1 | 4,045 | 32.0 |
| -Modified L-H | 1,784 | 0.9 | 1,784 | 14.1 |
| Take-All Total | 5,829 | 3.1 | 5,829 | 46.1 |
| Large Take-Some | 19,462 | 10.3 | 3,171 | 25.1 |
| Small Take-Some | 78,181 | 41.5 | 3,659 | 28.9 |
| Surveyed - Total | 103,472 | 54.9 | 12,659 | 100 |
| -Employer | 83,480 | 80.7 | | |
| -Non-Employer | 19,992 | 19.3 | | |
| Non-Surveyed | 85,041 | 45.1 | 0 | 0 |
| -Employer | 25,941 | 30.5 | | |
| -Non-Employer | 59,100 | 69.5 | | |
| Total | 188,513 | 100 | 12,659 | 100 |

Table 2B: April 2003 Wholesale Sampling

| Strata | Population | | Sample | |
|------------------------|----------------|-------------|--------------|-------------|
| | Counts | % | Counts | % |
| Take-All : | | | | |
| -Pre-specified | 4,991 | 5.0 | 4,991 | 58.7 |
| -Modified L-H | 1,148 | 1.1 | 1,148 | 13.5 |
| Take-All Total | 6,139 | 6.1 | 6,139 | 72.2 |
| Large Take-Some | 5,632 | 5.6 | 1,296 | 15.3 |
| Small Take-Some | 20,259 | 20.2 | 1,063 | 12.5 |
| Surveyed – Total | 32,030 | 31.9 | 8,498 | 100 |
| -Employer | 28,911 | 90.3 | | |
| -Non-Employer | 3,119 | 9.7 | | |
| Non-Surveyed: | 68,350 | 68.1 | 0 | 0 |
| -Employer | 22,221 | 32.5 | | |
| -Non-Employer | 46,129 | 67.5 | | |
| Total | 100,380 | 100 | 8,498 | 100 |

Table 3: Source of the Size Measure Used for the Population of Establishments (%)

| Source | Retail | | Wholesale | |
|---------------|----------|--------------|-----------|--------------|
| | Surveyed | Non Surveyed | Surveyed | Non Surveyed |
| <u>Survey</u> | | | | |
| MWRTS | 18.6 | 1.3 | 22.0 | 1.2 |
| UES | 6.7 | 0.8 | 5.8 | 0.4 |
| <u>Admin.</u> | | | | |
| GBI | 24.0 | 44.8 | 22.8 | 44.6 |
| GST | 35.4 | 37.4 | 31.2 | 32.8 |
| T2 | 15.2 | 15.7 | 18.3 | 21.0 |

For collection purposes, the clusters are converted into collection entities, i.e., the company tied to the establishments within the same TG. The collection entity, although at the company level (between the enterprise and the establishment), does not consider the geographical dimension like the clusters. Instead, sales by GEO are collected from each collection entity. The 21,157 clusters in the retail and wholesale samples (12,659+8,498) are expected to translate into 16,238 collection entities. Because of expected death rates (out-of-business and out-of-scope) and special collection arrangements, 11,285 collection entities are expected to be collected in a regular month. This

is a significant reduction in collection costs as the old MWRTS had 16,275 to collect.

2.2. Sample Update

Since April 2003, births have been added to the sample monthly. A birth in the context of the new MWRTS is a brand new cluster of establishments. A new establishment joining an existing cluster (the enterprise already has establishments in the same TG x GEO) is not considered a birth. Birth clusters are stratified by TG, GEO and size using the same definitions and stratum boundaries as in April 2003. They are sampled using the same sampling fractions. So far, the typical number of clusters birthed monthly in the population is 900 for retail and 400 for wholesale. Approximately 60% of them fall in the non-surveyed portion. Typically 90 retail and 80 wholesale clusters are added to the sample.

Deaths have also been identified monthly since April 2003. A death in the context of the new MWRTS is a cluster that does not have live and in-scope establishments anymore, i.e., all its establishments have died, have ceased retail or wholesale activities or have moved to a different enterprise. Surveys identify deaths more quickly than administrative sources. Both surveys and administrative sources update the BR. Surveys then extract their latest universe from the BR to update their frame. The source of the updates is difficult to establish (either a survey, an administrative source or both can identify a death). It is difficult to establish if the source of the update is independent or not from the survey. For that reason, MWRTS does not remove its deaths monthly from either its sample or its population. Instead, they are imputed with \$0 and contribute to the variance estimate. Every 6 months or so, an unbiased death removal is performed (Trépanier et al. 1998).

Other than births and deaths, the same sample is used month after month. In 2000, monthly sample rotation was investigated for the new MWRTS (Majkowski 2001). Results showed that any “reasonable” monthly sample rotation will affect the month-to-month change estimate by at least 0.2 percentage points. This was considered too high. To reduce response burden, a new sample should be selected every 4-5 years simultaneously with a full restratification of the population. In the interim, unbiased death removals and partial restratifications are planned to maintain the efficiency of the sample. MWRTS is also analysing options to replace sales obtained from direct data collection by modelled sales from the GST data for “simple” units (direct link with GST). This may be possible due to the improved processing of GST data (Dubreuil et al. 2003).

2.3. Estimation

The new MWRTS uses a simple expansion domain estimator for its surveyed portion

$$\hat{Y}(d) = \sum_h \frac{N_h}{n_h} \sum_{i \in h} y_{hi}(d),$$

where d is the domain of interest (e.g., TG, GEO), $Y(d)$ is the parameter of interest (e.g., total sales) in domain d , h denotes the strata, N_h and n_h are the population and sample sizes of h , i denotes the clusters, and $y_{hi}(d) = y_{hi}$ if $i \in d$, else $y_{hi}(d) = 0$ where y_{hi} represents the value of Y for cluster i in h . Estimated variances are also computed for the surveyed portion. The non-surveyed portion's contribution is added to the estimates through a multiplicative adjustment factor applied to $\hat{Y}(d)$. For the first year of the new MWRTS, it has been decided that the adjustment factor will be the ratio of the total April 2003 size measure in the non-surveyed portion over the total April 2003 size measure in the surveyed portion within a given TG x GEO. As a result, the month-to-month change estimate will be completely driven by the surveyed portion. This approach has some weaknesses but it has the advantage of allowing analysts and methodologists to concentrate their efforts on analysing and stabilising the surveyed portion. The adjustment factor methodology will be reviewed once the survey processes are stable.

Matthews and Bérard (2002) implemented an outlier detection and treatment methodology in the new MWRTS that provides a compromise between variance reduction and bias increase. It is based on a modified version of the Fuller (1991) "Test and Treat" method and another method referred to as the Deflation Factor method. Its objective is to identify influential units within suspicious domains (domains for which the simple expansion domain estimate is greater than an expected value (e.g., the forecasted estimate based on the time series)). These influential units are deemed to have true survey response values but their observed size is much larger than expected. A large observed value combined with a large weight seriously impacts the estimate. The outlier detection and treatment module proposes a factor which, if accepted by the analyst, is applied to the y -value to reduce its contribution to the domain estimate.

3. RESULTS AND EVALUATION

The new MWRTS sample was selected in April 2003. From May to November 2003, the sample was updated for births and deaths but data were not collected from the units. In the summer of 2003, a pre-contact was made with the sampled units to

introduce them to the survey and verify that they were in-scope for the survey. Then survey data were collected for the new MWRTS for December 2003 to April 2004 reference months in parallel with the old MWRTS data collection.

The first estimates from the new MWRTS (for April 2004 reference month) were released in June 2004. At the same time, the time series were revised and converted to NAICS-based TGs back to 1991 for retail and 1993 for wholesale. To do so, NAICS-based TG domain estimates were first produced from the old MWRTS from 1998 to 2003 (period for which the NAICS classification was carried on the old MWRTS). A subset of these estimates, the 1998 to 2001 estimates, also served in establishing macro-level conversion coefficients for each SIC-based TG to each NAICS-based TG. Backcasted NAICS-based TG estimates were produced using these coefficients for the period prior to 1999 (Fortier 2003). A NAICS-based TG time series was then available from the old MWRTS. Using January 2004 estimates for retail and February 2004 for wholesale final linkage was accomplished to reflect the new level provided by the new MWRTS.

3.1 Estimates : Old vs. New

The total retail and wholesale trade estimates are presented from both the old and new MWRTS for March 2004. Such differences were expected as wholesale trade lost units to retail trade and manufacturing under NAICS.

Table 4 : National Estimates from the Old and New MWRTS (in \$ billions) – March 2004

| | Old (SIC) | New (NAICS) | % Diff. |
|------------------|-----------|-------------|---------|
| Retail | 26.07 | 27.00 | +3.6% |
| Wholesale | 40.53 | 39.92 | -1.5% |

3.2. Evaluation

This section evaluates some of the assumptions made at the sampling design stage as well as the performance of the new size measure. The evaluation is however based on limited data, i.e., Dec. 2003 to March 2004.

3.2.1. Death Rates

The death rates take into account both out-of-business and out-of-scope clusters. Although they were dealt with separately at the sampling design, they are evaluated globally here. As mentioned in 2.1, the out-of-business rates were set at 10% for large take-some strata and 20% for small take-some strata. The out-of-scope rate was 15%. Consequently, the expected death rates in the sample are 25% and 35%, respectively for large take-some and small take-

some strata. Results observed for March 2004 reference month are presented in Table 5. The death rates in the sample appear higher than those expected at the sampling stage. However, if an unbiased death removal were performed, i.e., by removing the same proportion of deaths in and out of the sample, there would still be a significant proportion of deaths left in the sample but it would be in line with the initial assumption.

Table 5: Dead Clusters in the Take-Some Surveyed (TS) Portion (%) – March 2004

| Strata | In-Sample | Out-of-Sample |
|-------------------|-----------|---------------|
| Retail: | | |
| Large TS | 26.3 | 4.3 |
| Small TS | 36.6 | 3.1 |
| Wholesale: | | |
| Large TS | 28.9 | 5.5 |
| Small TS | 39.0 | 3.4 |

3.2.2. TG and GEO Movements

Based on March 2004 survey data, the TG and GEO movements are presented in Table 6.

Table 6 : Overall TG and GEO Movements (in %) for March 2004

| | TG | GEO |
|------------------|-----|-----|
| Retail | 2.9 | 5.1 |
| Wholesale | 1.9 | 5.2 |

As mentioned in 2.1, the movements between TGs were expected to be 2.7% for retail and 0.6% for wholesale. The observed rates are slightly higher for TGs. Movements between GEOs were deemed negligible based on UES. This made sense since industrial misclassification is known to be more frequent than geographical misclassification on the BR. The observed GEO rates are quite large, even larger than the TG rates. One explanation can be that MWRTS estimated the misclassification rate at the design stage on old information, i.e., 2000 UES information.

3.2.3. Response Rates

Table 7 presents December 2003 to March 2004 response rates, i.e., number of responding units divided by the total number of in-scope units. A 90% response rate was assumed at the sample design stage (or 10% nonresponse). As time goes by and the processes become more stable, that target response rate seems achievable and is in fact met for retail.

Table 7: Response Rates (%) from December 2003 to March 2004

| | Dec.03 | Jan.04 | Feb.04 | Mar.04 |
|------------------|--------|--------|--------|--------|
| Retail | 84.9 | 89.7 | 88.8 | 90.0 |
| Wholesale | 83.9 | 87.8 | 86.8 | 87.9 |

3.2.4. Efficiency of the Size Stratification

A preliminary analysis of the size stratification efficiency was performed. Micro-level annualised sales were produced for the sample using the sum of the seasonally adjusted December 2003 to March 2004 sales multiplied by 3. The annualised sales of each cluster were compared to the size stratification thresholds to see if the cluster appears to be correctly or incorrectly stratified. When it was incorrectly stratified, we looked if the cluster was “under stratified” (its size measure underestimated its real size) or “over stratified”. The “under stratified” cases are more problematic at estimation as their sales value, which is larger than expected, may be combined with a large design weight. The efficiency based on the April 2003 size measure source (MWRTS, UES, TRIO, and then for the individual components of TRIO) was examined. Table 8 shows the results. Note that the prespecified take-all clusters were excluded from that comparison as their criteria to be stratified take-all was not size. Multi-establishment clusters where the establishments’ size measure came from different sources were excluded.

Table 8: Efficiency of the April 2003 Size Stratification - Preliminary Evaluation

| Source | Retail | | Wholesale | |
|----------------|------------------------|------------------|------------------------|------------------|
| | Incorrectly Stratified | Under stratified | Incorrectly stratified | Under stratified |
| MWRTS | 11.8 | 2.2 | 15.3 | 3.2 |
| UES | 19.4 | 5.7 | 30.1 | 7.1 |
| TRIO | 25.5 | 2.5 | 28.5 | 3.0 |
| <i>GBI</i> | 33.8 | 2.5 | 42.2 | 3.0 |
| <i>GST</i> | 20.2 | 3.0 | 20.9 | 3.3 |
| <i>T2</i> | 27.4 | 1.1 | 27.4 | 2.4 |
| Overall | 21.2 | 2.8 | 24.2 | 3.4 |

Although a non-negligible proportion of the clusters appear to be incorrectly stratified, the size measure meets its first objective: avoid the “under stratified” situations. UES does not perform as well as other sources, but note that 1999 and 2000 information was used. Other results indicate that if MWRTS had restratified its population in April 2004 based on the latest size measure (but with same stratum boundaries), only 9.1% and 5.1% of the clusters in the retail and wholesale population respectively would have changed size strata. This seems to indicate that the incorrectly stratified cases in Table 8 are not due only to a global increase in the clusters’ size since April 2003. However, a more in-depth evaluation needs to be performed before drawing any conclusion.

3.2.5. Coefficients of Variation

Finally, the estimated CVs for March 2004 are compared to the target CVs (see Table 1).

Table 9 : Number of Domain Estimates within or not within Target CVs – March 2004

| | Retail | | Wholesale | |
|--------------------|--------------------|----|--------------------|----|
| | Within target CVs? | | Within target CVs? | |
| | Yes | No | Yes | No |
| National | 1 | 0 | 1 | 0 |
| TG | 16 | 3 | 12 | 3 |
| Prov./Terr. | 13 | 0 | 13 | 0 |

Target CVs are not achieved in some TGs for the following reasons: 1) These TGs often have the highest TG misclassification rates; 2) The variability in the annual size measure appears to underestimate the variability in the monthly sales.

4. CONCLUSION

The new MWRTS predicted quite accurately frame imperfections, such as deaths and TG movements, and nonresponse, but GEO misclassification was underestimated. Provisions were made at the sampling design to account for these in a modified Lavallée-Hidiroglou algorithm that incorporates expected out-of-business rates, and by inflating sample sizes. A size measure based on survey and administrative sources improved size stratification. As a result, most CVs were within the target CVs.

In the future, efforts could also be put on incorporating as well the impact of frame misclassifications and nonresponse in the Lavallée-Hidiroglou algorithm. This would better preserve the optimality provided by the Lavallée-Hidiroglou algorithm. A size measure that better reflects monthly variability of the survey data could also be an area of investigation. The 8-month delay between sample selection and actual data collection should be shortened to allow the use of more up to date information to estimate frame imperfection rates. If budget permits, the 5-month period to test the new survey in parallel with the old one should be longer to allow more time for fine-tuning and analysis.

5. ACKNOWLEDGEMENT

The author would like to express her sincere thanks to Catherine Dufour, Michel Ferland, Susie Fortier and Mark Majkowski for their precious inputs into this paper, and to Richard Evans and Chris Mohl for their thorough review of the paper.

6. REFERENCES

Bérard, H. (2001), “The Redesign of the Monthly Wholesale and Retail Trade Survey of Statistics Canada”, *Proceedings of the Survey Methods Section*, Statistical Society of Canada, pp. 81-86.

Dubreuil, G., Hidiroglou, M. A. and Pierre, L. (2003), “Use of Administrative Data in Modeling of Monthly Survey Data”, *Proceedings of the Survey Methods Section*, Statistical Society of Canada [CD-ROM].

Ferland, M. (2003), “Enhanced Lavallée-Hidiroglou Algorithm”, *Unpublished Manuscript*, Ottawa : Statistics Canada.

Fortier, S. (2003), “La conversion de données historiques selon un nouveau système de classification pour l'Enquête mensuelle sur le commerce de gros et de détail”, *Proceedings of the Survey Methods Section*, Statistical Society of Canada [CD-ROM].

Fuller, W. A. (1991), “Simple Estimators for the Mean of Skewed Populations”. *Statistica Sinica*, 1, 137-158

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

Latouche, M. (1988), “Détermination, allocation et sélection de l'échantillon”, *Recueil des textes des présentations du colloque sur les méthodes et domaines d'application de la statistique*, Bureau de la statistique du Québec, pp. 69-78.

Lavallée, P. and Hidiroglou, M. A. (1988), “On the Stratification of Skewed Populations”, *Survey Methodology*, June 1988, Vol. 14, No. 1, pp. 33-43.

Majkowski, M. (2001), “Maintaining Estimate Quality and Easing Response Burden in a Sub-Annual Business Survey”, *Proceedings of the Survey Research Methods Section*, American Statistical Association [CD-ROM].

Majkowski, M., Bérard, H., Dufour, C. and Fortier, S. (2004), “Monitoring Survey Processes of the Canadian Monthly Wholesale and Retail Trade Survey”, *Proceedings of the Survey Research Methods Section*, American Statistical Association [CD-ROM].

Matthews, S. and Bérard, H. (2002), “The Outlier Detection and Treatment Strategy for the Monthly Wholesale and Retail Trade Survey of Statistics Canada”, *Proceedings of the Survey Methods Section*, Statistical Society of Canada [CD-ROM].

Trépanier, J., Babyak, C., Marchand, I., Bissonnette, J. and St-Pierre, M. (1998), “Enhancements to the Canadian Monthly Wholesale and Retail Trade Survey”, *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp. 487-492.