Workplace Employee Survey – Data Accuracy

While considerable effort is made to ensure a high standard throughout all survey operations, the resulting estimates are inevitably subject to a certain degree of error. This is true in every survey. The total survey error can be divided into two main components: the sampling error and the non-sampling errors. The sampling error is due to the fact that estimates are computed using only a sample of the whole population instead of a complete census while the non-sampling errors are due to all other causes such as an imperfect frame, measurement errors or non-response. For instance, measurement errors can arise from mistakes made by respondents or interviewers during the collection of data, from errors made in keying in the data, or from other sources. This type of error may lead to the imputation of consistent but not necessarily correct values.

In WES, the sampling error and part of the non-response and frame errors are dealt with by attaching an estimation weight, called the final weight, to each sampled unit (workplace or employee) for which we have data; be they imputed or not. The remaining of the non-response error is dealt with through the imputation of missing data. The editing stage of the survey attempts to minimize the effect of measurement errors. This stage involves outlier detection and different validation steps. The boundary between editing and data quality is fuzzy. The former is performed to improve the latter.

If there were no non-sampling error, the weighting strategy would ensure that the estimates are approximately design unbiased in the sense that the expectation over all possible samples of the survey error would be approximately equal to zero. To evaluate the quality of an estimate and to obtain valid inferences, measures of precision, such as the estimated coefficient of variation, are usually computed. The estimated coefficient of variation is defined as the square root of the estimated design variance of an estimate over the estimate itself. The design variance is the hypothetical variability of the estimates taken over all possible samples that could have been drawn under the sampling design. Since only one sample is selected in practice, the design variance is unknown. However, it can be estimated using only one sample (in WES, the mean bootstrap technique is used), which allows the desired measures of precision to be obtained. Note that smaller coefficients of variation imply better quality of the estimates.

The WES sample was designed to be efficient for estimating totals at an industry by region by size level within the available budget. The projected coefficients of variation were around 5% for industry and 10% for industry by region for variables highly correlated with employment. When estimates are produced, they are compared to the projected precision. Approximately 60% of all estimates of totals exceeded expectation with another 25% being within the Statistics Canada publishable cut-off of 33%. The remaining 15% were not publishable by our

standards. These were mostly estimates not highly correlated with employment. All estimates falling into the unpublishable category are validated.   Estimates with a cv in the range of 25 to 33% are published with a cautionary flag, denoting their relatively high variability.