Estimating the Variance of Seasonally Adjusted Series of Monthly Statistics Canada Surveys

François Verret[1], Catalin Dochitoiu[1]
[1]Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Ontario K1A 0T6

Abstract
Seasonal adjustment is done using the X-12-ARIMA method at Statistics Canada. Appropriate variance estimates are needed to support inference based on seasonally adjusted estimates. In their absence, cross-sectional quality indicators of precision of the unadjusted series estimates are sometimes provided to data users along with the seasonally adjusted series estimates. Those indicators are often based on the estimated coefficient of variation of the unadjusted estimates. Statistics Canada is currently exploring options for estimating the variance of the seasonally adjusted series for its sub-annual surveys. A method suitable for every statistical program is sought. This paper discusses the methodologies considered and recent progress.

Key Words: X-12-ARIMA, X-11 filters, replication methods, linearization techniques, sample surveys

1. Introduction

In repeated infra-annual surveys (e.g. monthly or quarterly), estimates of the raw series and estimates of the seasonally adjusted (SA) series can be disseminated. With either of those, estimates of the variance of the point estimates are needed to support inference. For most survey programs, estimates of variance of the raw series are produced and quality indicators based on those estimates are disseminated. However, it is less customary to produce variance estimates for the SA series. No quality indicators based on such variance estimates are published at Statistics Canada. For most statistical programs, cross-sectional quality indicators of the unadjusted/raw series estimates based on their estimated coefficient of variation (CV) are provided to data users along with the SA series estimates. When analysing the SA estimates, one could make the assumption that the CV of the SA estimate is comparable to the CV of the raw estimate of the same reference period, but this assumption needs to be validated and could potentially be questionable.

Statistics Canada is thus exploring options for estimating the variance of SA series of its infra-annual statistical programs. A method suitable for all programs and applicable to various sample designs (such as household and business sample surveys and censuses) is sought, given that the X-12-ARIMA method (Findley et al., 1998) is used for seasonal adjustment. This exploration is done on two surveys: the Labour Force Survey (LFS) and the Monthly Retail Trade Survey (MRTS). They respectively are a household and a business survey. This paper presents results of this exploration.

Furthermore, comparison of estimates of different reference periods of a series (e.g. the study of the month-to-month movement) is usually the focus of analysis for repeated

Disclaimer: This paper describes theoretical approaches and does not reflect currently implemented methods at Statistics Canada.

surveys. Appropriate estimates of the covariance between the estimates being compared are needed to support this kind of inference. At the moment, no lag-covariance estimates are disseminated at Statistics Canada. This means one must use alternatives such as the overlapping confidence intervals method to compare estimates of different reference periods of a given series. The correlations between the two estimates being compared are typically positive in a repeated survey and alternatives which do not require the estimation of the covariance between the estimates, such as the overlapping intervals method, are in general conservative. This is also studied in this paper for both the raw and SA series.

In Section 2, estimation of the variance and covariance of the raw series is discussed. In Section 3, the methodology for estimating the variance and covariance of the SA series considered by Statistics Canada is presented. This methodology is based on an unpublished paper by Benoît Quenneville and from the comments of collaborators and reviewers. In Section 4, empirical results obtained from studying variance estimation of the MRTS are presented. The paper ends with a conclusion on variance estimation, on the dissemination of quality indicators based on the estimated variance and on possible next steps.

2. Variance Estimation of the Raw Series

Let the vector of the observed raw series be

$$y = Y + e,$$

where Y is the raw series vector one would observe had a census took place and e is the vector of sampling errors. When studying the raw series, the parameter of interest is Y (or a function of it) and the estimate is y (or the corresponding function of y). When no unit or item non-response is present, the variability of y is typically viewed as being design-based. Under design-based inference, Y is viewed as fixed. Let Σ_e denote the (design-)variance matrix of y and e. Design variance comes from sampling and depends on the sampling design and on the estimation methodology used to produce the estimator y. Särndal et al. (1992) refer to the combination of sampling design and estimator as a *strategy*. Variance due to non-response will be ignored in the rest of this paper. However, the results presented could be extended to include variance components due to non-response and adjustments done to compensate for non-response such as imputation.

2.1 Strategy of the Raw Series of the Labour Force Survey

The Labour Force Survey (LFS) is a monthly survey which provides estimates of employment and unemployment. It has a stratified multi-stage design, with a small sampling fraction of clusters at the first stage and where households are selected at the last stage. Rotating panels are used to create overlap between samples of neighbouring months. This is done to increase the precision of estimates of the difference between two neighbouring monthly estimates. A given household is in the LFS sample for six consecutive months before rotating out.

In terms of estimation, the monthly sample data are weighted to account for probability of selection in the sample. Survey non-response is compensated both by imputation and adjustment of the weights. Weights are also calibrated to correct for coverage errors. A regression composite estimator is used to further improve the estimates of level and change (Gambino et al., 2001).

The Rao-Wu bootstrap (Rao & Wu, 1988; Rao, Wu & Yue, 1992) is used to produce variance estimates of monthly estimates of the LFS. The bootstrap weights are generated in a coordinated fashion longitudinally using the coordinated bootstrap method proposed by Roberts et al. (2001) to allow for estimation of variance of point estimates involving multiple survey months.

2.2 Strategy of the Raw Series of the Monthly Retail Trade Survey

The Monthly Retail Trade Survey (MRTS) collects sales, e-commerce sales and number of locations from a sample of retailers. The same sample of retailers (establishments) is surveyed for a period of approximately 5 years. A given 5-year sample is selected independently of previous samples using a stratified simple random cluster sampling design. Sampling fractions can be large in some strata. Every month, births may be observed on the sampling frame (Statistics Canada's Business Register). To make sure they are covered by the survey, a sample of births is drawn every month using again a stratified simple random cluster sampling design.

Estimation in the MRTS is performed monthly in a cross-sectional fashion. The sample data is weighted to account for the sampling design. Non-response is compensated using imputation. The sampling weights are calibrated to Goods and Services Tax (GST) totals with a ratio adjustment and the ratio estimator (which is a special case of the calibrated estimator and of the Generalized Regression (GREG) estimator) is used. Additionally, correction factors might be applied to the survey variables to minimise response error.

To prevent a break in the series the first month a new 5-year sample is used, a "linkage" of the old and new samples is performed. To make linkage possible, the old 5-year sample and the new 5-year sample are surveyed concurrently until the new sample monthly estimates are of quality that is judged adequate. Only the new sample is surveyed the following months. Linking consists of modifying the old sample estimates using time series benchmarking under the following constraints:

1. the estimate of the first month of the old sample remains unchanged;
2. the estimate for the last month the old sample is surveyed is made equal to the new sample estimate for that month.

The benchmarking process ensures that the two constraints are satisfied while preserving the original month-to-month movement as much as possible. Some theoretical details on the linkage process are provided in Section 2.4.

As is the case for point estimation, variance estimation is performed monthly in a cross-sectional fashion in the MRTS (i.e. variance estimates of the raw point estimates of the reference month are produced, but no estimates of lag-covariances are produced). A linearization approach is used to estimate the variance of the ratio estimator. Variance estimates are not updated after a linkage is performed. More information on the MRTS strategy can be found in Renaud & Laroche (2016) and Laroche (2019).

2.3 Impact of not Estimating the Variance in a Longitudinal Fashion

Estimating the variance in a longitudinal fashion, such as with the LFS bootstrap, makes it possible to get an estimate of the full variance matrix of the raw series Σ_e and not just its diagonal. This enables the data user to compare estimates of two reference periods using a variance estimate for the difference between the two estimates which takes into account the covariance between them.

Let s and t be the indices of the two different months for which the estimates need to be compared. The 95% confidence interval one would typically use for the difference between two totals Y_s and Y_t is

$$y_s - y_t \pm 1.96\sqrt{\hat{V}(y_s - y_t)} = y_s - y_t \pm 1.96\sqrt{\hat{V}(y_s) + \hat{V}(y_t) - 2\hat{C}(y_s, y_t)}.$$

This approach will be referred to as the "ideal" approach or method in the rest of the paper. When no estimate of the covariance is provided, a first alternative could be to use the overlapping intervals method. It consists of verifying if the two confidence intervals $y_s \pm 1.96\sqrt{\hat{V}(y_s)}$ and $y_t \pm 1.96\sqrt{\hat{V}(y_t)}$ overlap and concluding that the two totals are significantly different when the two intervals do not overlap. This is equivalent to using the following confidence interval for the difference:

$$y_s - y_t \pm 1.96\sqrt{\hat{V}(y_s) + \hat{V}(y_t) + 2\sqrt{\hat{V}(y_s)\hat{V}(y_t)}}.$$

It is in turn equivalent to assuming the most conservative value possible for the correlation between the two estimates: -1. A less conservative alternative would be to assume a null correlation between the estimates. The confidence interval would then be:

$$y_s - y_t \pm 1.96\sqrt{\hat{V}(y_s) + \hat{V}(y_t)}.$$

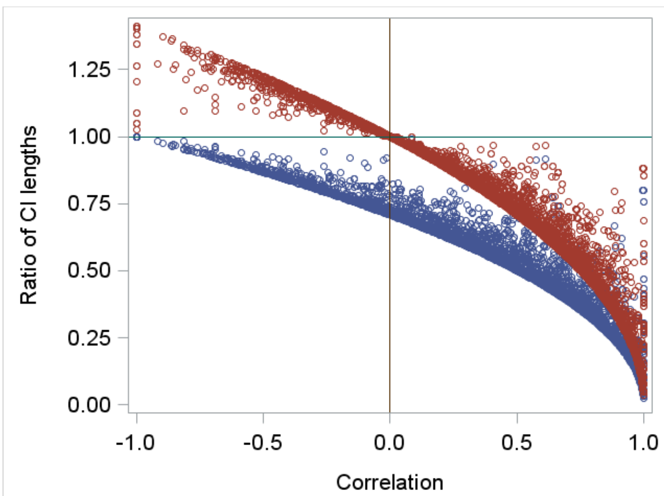

Figure 1: Ratio of confidence interval lengths for MRTS month-to-month movements as a function of the correlation. In blue, the ideal approach vs. the overlapping interval method. In red, the ideal approach vs. assuming no correlation.

Let $\hat{\rho} = \widehat{corr}(y_s, y_t)$. In the rest of this paper, the "month-to-month movement" will correspond to the absolute month-to-month movement: the difference between two consecutive monthly estimates of a given series. Figure 1 gives for various MRTS raw month-to-month movements the ratio of the length of the ideal approach confidence interval and of the length of either of the alternative confidence intervals as a function of

$\hat{\rho}$, when at least one of the two estimated variances is greater than 0. Those ratios are respectively given by

$$\frac{1.96\sqrt{\hat{V}(y_s)+\hat{V}(y_t)-2\hat{C}(y_s,y_t)}}{1.96\sqrt{\hat{V}(y_s)+\hat{V}(y_t)+2\sqrt{\hat{V}(y_s)\hat{V}(y_t)}}}$$

and

$$\frac{1.96\sqrt{\hat{V}(y_s)+\hat{V}(y_t)-2\hat{C}(y_s,y_t)}}{1.96\sqrt{\hat{V}(y_s)+\hat{V}(y_t)}}.$$

For the first alternative, the ratio is bounded by $\sqrt{(1-\hat{\rho})/2}$ when $\hat{V}(y_s)=\hat{V}(y_t)>0$ and by 1 when one of the variances is 0. The first alternative is thus more conservative than the ideal approach unless $\hat{\rho}=-1$ or unless at least one of the estimated variance is 0.

Similarly, the ratio for the second alternative is bounded by $\sqrt{1-\hat{\rho}}$ and 1. The second alternative is thus more conservative than the correct approach if the estimated correlation is positive and more liberal if the estimated correlation is negative.

2.4 Estimating the Variance in a Longitudinal Fashion for the MRTS

A longitudinal application of the Rao-Wu bootstrap was developed for the study presented in this paper to estimate the variance of MRTS estimates in a longitudinal fashion. It follows the principles of the coordinated bootstrap method of Roberts et al. (2001). This bootstrap can be applied independently to each 5-year sample since before a linkage process is applied the estimates of a given 5-year sample are independent of estimates of other 5-year samples. This implies that Σ_e is block-diagonal in the MRTS. For each one of those samples, the following steps were followed:

1. A set of B = 500 Rao-Wu bootstrap weights were generated for the sample selected the first month the 5-year sample was surveyed.
2. For each of the following months, 500 Rao-Wu bootstrap weights were generated for the monthly sample of births. Those bootstrap weights can be generated independently of the other bootstrap weights because sampling of births is done independently of selection of the rest of the sample. Those bootstrap weights are combined to those of Step 1 and to those of previous applications of Step 2 to create the set of bootstrap weights used for the month.
3. An independent ratio-adjustment/calibration of bootstrap weights was performed for each month to the corresponding monthly GST totals. After this step, because the calibration adjustments vary every month, each month has its own set of calibrated bootstrap weights.
4. Bootstrap estimates were calculated for each month using that month's set of calibrated bootstrap weights.

Once Steps 1 to 4 are applied to each 5-year sample, a set of bootstrap (vector) estimates $y^{(b)}$, b = 1, …, 500 is obtained. The usual bootstrap formula could be used to estimate the matrix Σ_e:

$$\hat{\Sigma}_e = \frac{1}{B}\sum_{b=1}^{B}\left(y^{(b)} - \bar{y}.\right)\left(y^{(b)} - \bar{y}.\right)^T.$$

However, since it is known that Σ_e is block-diagonal, it is preferable to use the corresponding formula independently by 5-year sample to estimate each of the blocks and assign values of 0 to all other elements of the $\hat{\Sigma}_e$ matrix.

For the current 5-year sample, when a new month of data becomes available, Steps 2 to 4 could be applied. Step 2 would further increase the size of the bootstrap weights file. In Step 4, new records would be created for the new month's births. The value of the variable of interest should be set to 0 for those records for months before their birth. Since this would not change previous months' bootstrap estimates, in Step 4 it is more efficient to simply

1. calculate the bootstrap estimates for the new month;
2. update $y^{(b)}$ accordingly by making it one element longer;
3. calculate the new row and column of the bootstrap variance matrix estimate $\hat{\Sigma}_e$.

2.5 Estimating the Variance of a Linked Series

Time series linking is a special case of time series benchmarking. The general subject of benchmarking is discussed in Dagum & Cholette (2006). Let *y* and *z* respectively be the vector of values of the old sample and of the new sample, and τ and λ be predetermined linking parameters. Linking of the series consists of removing the break in the series caused by the change of sample. It is done by replacing the values of *y* with those of θ, the vector of values which minimises the following distance function:

$$\left(1-\tau^2\right)\left(\frac{\theta_1 - y_1}{y_1^{\lambda}}\right)^2 + \sum_{t=2}^{T}\left(\frac{\theta_t - y_t}{y_t^{\lambda}} - \tau\frac{\theta_{t-1} - y_{t-1}}{y_{t-1}^{\lambda}}\right)^2,$$

under the constraints $\theta_1 = y_1$ and $\theta_T = z_T$, where t = 1 is the first reference period the old sample is surveyed, $t = T$ is the last time it is surveyed and the first time the new sample estimates are used (this would be the last overlapping month in the MRTS). Minimizing the objective function aims at preserving the relative month-to-month movement in the old series when $\lambda = 1$. Linkage is then multiplicative. When $\lambda = 0$, minimizing the objective function aims at preserving the original absolute month-to-month movement. Linkage is then additive. The solution to the minimization problem can be shown to be

$$\theta_t = y_t + \frac{\tau^{T-t}\left[\sum_{j=0}^{t-1}\tau^{2j} - \tau^{2(t-1)}\right]}{\sum_{k=0}^{T-2}\tau^{2k}}\left(\frac{z_T - y_T}{y_T^{\lambda}}\right)y_t^{\lambda}; \quad t = 1,\ldots,T.$$

Assuming that the linking parameters are not sample-dependent, when a replication approach is used to estimate the variance of y, one could apply a linkage to each of the series replicate estimates to take linkage into account in variance estimation.

Alternatively, one could approximate the linkage process using a linearization approach and adjust the pre-linkage variance matrix estimate. It can be shown that $\underline{\theta} \approx \mathbf{U}\left[y_1,\ldots,y_T,z_T\right]^T$, where

$$\mathbf{U}(t,) = \begin{cases} \Bigg[\underbrace{0,\ldots,0}_{t-1}, 1 + \lambda f_t(\tau) Y_t^{\lambda-1}\dfrac{\Theta_T - Y_T}{Y_T^{\lambda}}, \underbrace{0,\ldots,0}_{T-t-1}, \\ \qquad -f_t(\tau)\dfrac{Y_t^{\lambda}}{Y_T^{\lambda+1}}\left[Y_T + \lambda\left(\Theta_T - Y_T\right)\right], f_t(\tau)\dfrac{Y_t^{\lambda}}{Y_T^{\lambda}}\Bigg], \quad \text{if } t = 1,\ldots,T-1, \\ \left[0,\ldots,0,1\right], \text{ if } t = T, \end{cases}$$

where Θ is the census equivalent of θ and $f_t(\tau) = \tau^{T-t}\left[\sum_{j=0}^{t-1}\tau^{2j} - \tau^{2(t-1)}\right] \Big/ \sum_{k=0}^{T-2}\tau^{2k}$. A linearization variance estimator is thus given by $\hat{V}\left(\underline{\theta}\right) = \hat{\mathbf{U}}\hat{V}\left(\left[y_1,\ldots,y_T,z_T\right]^T\right)\hat{\mathbf{U}}^T$, where $\hat{\mathbf{U}}$ corresponds to $\mathbf{U}$ with Y replaced by y and Θ replaced by θ. It is straightforward to extend this linearization approach to the full span of the series since linkage does not affect the data for $t < 1$ and $t > T$. In the case of the MRTS, either approach to taking linkage into account in variance estimation would make Section 2.4's estimator of the block-diagonal variance of the raw series not block-diagonal, as they should since the target variance matrix to estimate is not block-diagonal after linkage.

3. Variance Estimation of the Seasonally Adjusted Series

3.1 Estimating the Variance of the Seasonally Adjusted Series Under Additive Decomposition

When Y is decomposed additively for seasonal adjustment, one has:

$$\begin{aligned} y &= Y + e \\ &= S + C + I + e, \end{aligned}$$

where S is the seasonal and calendar component, C is the trend and I is the irregular. In this paper, S and C will be viewed as deterministic components and I and e will be viewed as stochastic components. The SA series of interest is given by $N = C + I = Y - S$. It is not directly observable, even when a census of the finite population of interest is performed. When seasonally-adjusting y with X-12-ARIMA, y is decomposed as $y = \hat{S} + \hat{C} + \widehat{I+e}$. It is not possible to get separate estimates of I and e. The estimated SA series is given by $\hat{N} = \hat{C} + \widehat{I+e} = y - \hat{S}$. The sampling error is thus part of the SA estimate (except in the unlikely event that e is absorbed by the estimate of the seasonal component in X-12-ARIMA). Furthermore, there are two sources of error in the SA estimate: the sampling error and the model error which is made when decomposing the series into its components.

As was mentioned in the introduction, this section of the paper is for the most part based on an unpublished paper by Quenneville (2013) and on the comments of collaborators and reviewers. Quenneville (2013) shows that the X-12 ARIMA seasonal adjustment method can be viewed as a nonparametric smoothing method and estimates the variance of the SA series under the assumption that I is a white noise process by following and adapting a method of Cleveland (1979, Section 6). The method is adapted for the presence of sampling error and for estimation of the variance of smoothed values.

The main steps of the seasonal adjustment of X-12-ARIMA are:

1. The fitting of a reg-ARIMA model for calendar effects and other priors such as outliers;
2. The estimation of calendar effects, application of prior adjustments to the original series, and extension of this series with forecasts;
3. The estimation of seasonal factors using the X-11 method on the prior adjusted series extended with forecasts.

More details on X-12-ARIMA and the X-11 method are respectively given in U.S. Census Bureau (2017) and Ladiray and Quenneville (2001).

The following steps describe how Quenneville (2013) linearly approximates the X-12-ARIMA method. At the end of those steps, matrices **W** and **R** are obtained such that $\hat{N} \approx \mathbf{W}y$ and $\widehat{I+e} \approx \mathbf{R}y$. The steps are:

1. Obtain and consider fixed for the following step the orders of differentiation of the ARIMA model, the reg-ARIMA regressors, the estimated values of the ARMA parameters (and by extension the autocorrelation matrix), the order of the trend moving average, the selected seasonal filters and the identified extreme values and their weights (in Table C17 of X-12-ARIMA).
2. Obtain the matrix that transforms y into the prior adjusted series.
3. For a fixed number of forecasts, using the impulse response method described in Ladiray and Quenneville (2001, Section 3.4) or in Findley and Martin (2006), obtain the two weight matrices which transform a series (extended with its forecasts) into the SA series and into the estimated irregular component. This method consists of seasonally adjusting the columns of the identity matrix of order equal to the length of the original series plus the number of forecasts, holding fixed the items listed in Step 1, but by bypassing the extreme observations identification step by increasing sufficiently the sigma limits of the X-11 algorithm.
4. Matrix **W** is obtained by multiplying the matrix of Step 2 with the first matrix of Step 3 and matrix **R** is obtained by multiplying the matrix of Step 2 with the second matrix of Step 3.

See Bell (2012) for properties of the X-12-ARIMA smoothing matrices.

An alternative to the approach of Quenneville (2013) is to do the following:

1. For chosen seasonal adjustment options, apply X-12-ARIMA to the original series and obtain the estimated SA series (Table D11) and the estimated irregular (Table D13).
2. Using the orders of differentiation, the estimated ARMA parameters and seasonal adjustment options of Step 1, except for the sigma limits which should be large enough to bypass the extreme values identification step, successively apply X-12-ARIMA to the original series plus a small value (say δ) times each vector of the identity matrix of order equal to the length of the series. Obtain each resulting pairs of SA series (D11) and irregular (D13) vectors.
3. The *i*th column of **W** is given by the difference between the *i*th D11 vector of Step 2 minus the D11 vector of Step 1, divided by δ. The *i*th column of **R** is given by the difference between the *i*th D13 vector of Step 2 minus the D13 vector of Step 1, divided by δ.

This is the approach that was used in the empirical study of Section 4. Note that to get proper non-zero estimates of the columns of **W**, δ should be chosen to be large enough to produce D11 and D13 vectors in Step 2 that are different from those of Step 1.

Under the assumptions that I results from a white noise process and that seasonal adjustment can be approximated by a linear transformation **W**, i.e. $\hat{N} \approx \mathbf{W}y$, one has

$$V\left(\hat{N}-N\right) \approx \mathbf{W}\Sigma_e\mathbf{W}^T + \sigma_I^2\left(\mathbf{I}-\mathbf{W}\right)\left(\mathbf{I}-\mathbf{W}\right)^T,$$

where **I** is the identify matrix and σ_I^2 is the variance of *I*. The first component of variance of the estimated SA series $\mathbf{W}\Sigma_e\mathbf{W}^T$ is design-based. The second component $\sigma_I^2\left(\mathbf{I}-\mathbf{W}\right)\left(\mathbf{I}-\mathbf{W}\right)^T$ is model-based.

In general, the variance matrix of the raw series Σ_e is either estimated using

- direct/classic formulas combined with the Taylor linearization technique if the GREG or the calibration estimator is used,
- or a replication approach.

In both cases, the first component of variance of the estimated SA series can be estimated by $\mathbf{W}\hat{\Sigma}_e\mathbf{W}^T$. When a replication approach is used to estimate Σ_e, one can instead replicate the seasonal adjustment process and the resulting design variance matrix estimated would be an estimate of the design-based component of variance of the estimated SA series. It might be preferable to use this latter approach if it is thought that the linear approximation of the seasonal adjustment process is questionable. Some steps of the X-12-ARIMA process make it non-linear, namely the estimation of auto-correlations and the forecasts in the ARIMA/first half of the process and the correction for extreme values in the X-12/second half of the process.

To estimate the model-based component of the variance of the SA series $\sigma_I^2\left(\mathbf{I}-\mathbf{W}\right)\left(\mathbf{I}-\mathbf{W}\right)^T$ requires to estimate σ_I^2. Quenneville (2013) suggests the following estimator of the variance of the irregular component *I*:

$$\breve{\sigma}_I^2 = \max\left\{0, \frac{1}{\text{trace}\left(\mathbf{R}\mathbf{R}^T\right)}\left[r^T r - \text{trace}\left(\mathbf{R}\hat{\Sigma}_e\mathbf{R}^T\right)\right]\right\},$$

where $r = \mathbf{R}y$ is the linear approximation of the estimate $\widehat{I+e}$. He also suggests the following robust version of the estimator:

$$\ddot{\sigma}_I^2 = \max\left\{0, \frac{1}{\text{trace}\left(\Delta\mathbf{R}\mathbf{R}^T\Delta\right)}\left[r^T\Delta^2 r - \text{trace}\left(\Delta\mathbf{R}\hat{\Sigma}_e\mathbf{R}^T\Delta\right)\right]\right\},$$

where Δ is a diagonal matrix with values between 0 and 1 in the diagonal that is used to down-weight extreme values of the irregular (down-weighting corresponds to values smaller than 1).

3.2 Estimating the Variance of the Seasonally Adjusted Series Under Multiplicative Decomposition

Multiplicative decomposition corresponds to:

$$\begin{aligned} y &= Y + e \\ &= S \times C \times I + e. \end{aligned}$$

The SA series of interest is given by $N = C \times I = Y/S$. Its X-12-ARIMA estimator is given by $\hat{N} = \widehat{C'} \times \widehat{I'} = y/\widehat{S'}$. Primes are used in the preceding formula because e potentially contaminates the estimation of each of the three components of Y in a more significant fashion under multiplicative decomposition. Because of the following approximation

$$\begin{aligned} \log y_t &\approx \log Y_t + e_t/Y_t \\ &= \log S_t + \log C_t + \log I_t + e_t/Y_t. \end{aligned}$$

and assuming that the linearization matrix $\mathbf{W}$ is such that $\log \hat{N} \approx \mathbf{W}\log y,$ an estimator of $V\left(\log\hat{N} - \log N\right)$ is given by

$$\mathbf{W}\Psi^{-1}\hat{\Sigma}_e\Psi^{-1}\mathbf{W}^T + \hat{\sigma}_{\log I}^2\left(\mathbf{I}-\mathbf{W}\right)\left(\mathbf{I}-\mathbf{W}\right)^T,$$

where $\Psi = \text{diag}\left(y\right)$. The latter matrix is needed in order to approximate the variance of the transformed data using the variance of the original data. The estimate of the irregular variance $\hat{\sigma}_{\log I}^2$ is equal to either

$$\breve{\sigma}_{\log I}^2 = \max\left\{0, \frac{1}{\text{trace}\left(\mathbf{R}\mathbf{R}^T\right)}\left[r^T r - \text{trace}\left(\mathbf{R}\Psi^{-1}\hat{\Sigma}_e\Psi^{-1}\mathbf{R}^T\right)\right]\right\},$$

where $r = \mathbf{R} \log y$, or to the corresponding robust version of this estimator $\ddot{\sigma}^2_{\log I}$. To find the variance of the untransformed data $V\left(\hat{N} - N\right)$ requires to use the following approximations: $\log \hat{N}_t \approx \log E\left[\hat{N}_t\right] + \hat{N}_t / E\left[\hat{N}_t\right] - 1$ and $\log N_t \approx \log E\left[N_t\right] + N_t / E\left[N_t\right] - 1$. In both expressions, only the second term contributes to variance and an estimator of $V\left(\hat{N} - N\right)$ is thus given by

$$\Omega \mathbf{W} \Psi^{-1} \hat{\Sigma}_e \Psi^{-1} \mathbf{W}^T \Omega + \hat{\sigma}^2_{\log I} \Omega \left(\mathbf{I} - \mathbf{W}\right)\left(\mathbf{I} - \mathbf{W}\right)^T \Omega, \tag{1}$$

where Ω is a diagonal matrix with either $\hat{N}$ or $\hat{C}$ in its diagonal (both are estimates of $E\left[\hat{N}\right]$ and $E\left[N\right]$). As in the additive decomposition case, when a replication method is used to estimate the design variance, the design-based/first component of (1) can be estimated by replicating seasonal adjustment.

4. Empirical Study

This section presents an empirical study of variance estimation of the raw and SA series of the MRTS. Focus is put on this survey because its design features vary more from domain to domain than those of the LFS. In particular, the sampling fraction can be large for some domains. Moreover, since the LFS estimates primarily counts, no population unit is influential on the total to estimate, whereas in the MRTS some population units might be a lot more influential than others due to large weights, large values of the variable of interest or a combination of the two.

Results presented in this section were obtained using part of the span of the MRTS, namely the data of reference periods covering January 2013 to January 2019. This involves two "5-year" samples, the first spanning from January 2013 to December 2016 and the second spanning from December 2016 to January 2019. A total of 21 retail North American Industry Classification System (NAICS) sales series where studied. Those are NAICS 44111, 44112, 4412, 4413, 4421, 4422, 443, 444, 44511, 44512, 4452, 4453, 446, 447, 4481, 4482, 4483, 451, 4521, 4529 and 453 (excluding 453993 – Cannabis stores which was only recently introduced). Geographic sales series were also studied. Those geographies are the 10 provinces, the 3 territories, the 3 largest Census Metropolitan Areas (Montreal, Toronto, Vancouver) and their 3 respective provincial complements.

Variance estimation of the raw series was performed using the longitudinal bootstrap described in Section 2.4. Linkage was taken into account in variance estimation by replicating the linkage for each bootstrap replicate as described in Section 2.5. It was furthermore verified that a comparable estimate of Σ_e is obtained when a linearization approach is used to take linkage into account.

Variance estimation of the estimated SA series was performed as described in Section 3. Most MRTS series are decomposed in a multiplicative fashion to perform seasonal adjustment. In that case the diagonal of Ω was composed of $\hat{N}$. The design-based component of variance was obtained by replicating seasonal adjustment on each bootstrap series. The model-based component was estimated using the non-robust version of the estimator of the variance of the irregular component I (or of the log of the irregular component in the multiplicative decomposition case).

4.1 Comparing the Variance of the Raw Series and of the SA Series

When only estimators of the variance of the raw series are available, it is of interest to know if the variance of the raw series is a good proxy for the variance of the estimated SA series. In the case of a census, $\Sigma_e = 0$ and unless the irregular component is negligible the model-based component of the variance of the estimated SA series dominates. Figures 2 and 3 illustrate this.

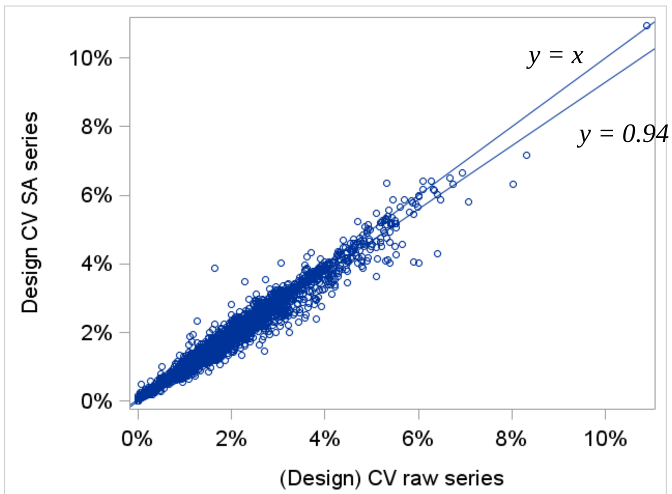

Figure 2: Scatterplot of the design-based component of the CV of the SA series (y-axis) and of the (design) CV of the raw series (x-axis). The two lines shown on the graph correspond to $y = x$ and to the Deming (1943) regression line $y = 0.94\ x$.

Figure 2 shows that the design-based component of the CV of the SA series and the CV of the raw series are very highly correlated. The design-based component of the CV of the SA series is globally 6% smaller than the CV of the raw series. This means if the model-based component was negligible, using the CV of the raw series to qualify the estimated SA series would be a bit conservative.

Figure 3 gives the more complete picture by comparing the CV of the estimated SA series taking both variance components into account with that of the raw series. The CV of the SA series is in general greater than that of the raw series. Taking both components of variance into account, using the quality indicators of the unadjusted series to qualify the adjusted series is thus in general liberal. It can be very liberal because the model-based variance component can be relatively important for some domains. Three domains are highlighted on Figure 3. NAICS 4521 – Department stores is at the far left in black. A census is conducted for this domain, which means the CV of the raw series is 0, but the CV of the estimated SA series is positive. Next to this domain, in red, are the 3 territory series, which correspond to near censuses and are volatile in nature. On top in turquoise is the NAICS 443 – Electronics and appliances stores series. The sales weighted sampling fraction of this domain is around 20%, which is not that large. Short seasonal filters are used for this domain to pick up evolving seasonal effects found in this industry (related to Black Friday). This favors a larger model variance component. This series also has had more correction factors applied than most series because of a recent redefinition of the NAICS. Those factors where not applied to artificially reduce the irregular component and most likely have inflated the estimated variance resulting from $I + e$. Because they have the same effect on each bootstrap replicate, the inflation should result in an increased estimated model-based component of the variance only.

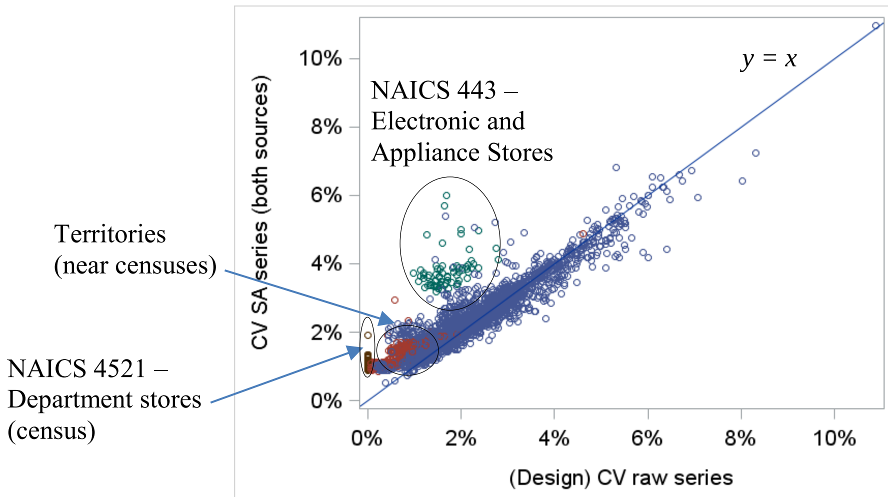

Figure 3: Scatterplot of the CV of the SA series (y-axis) and of the (design) CV of the raw series (x-axis). The line shown on the graph corresponds to $y = x$.

4.2 Study of the Month-to-month Movement

The month-to-month movement in the raw series and in the SA series is studied in this section. More specifically, tests of hypothesis that the movement is different from 0 are performed using the ideal approach (i.e. by making use of the estimated covariance between the two estimates being compared) or using the two alternatives described in Section 2. Hypothesis tests at the 5% level were performed for every month-to-month movement.

Table 1: Percentage of Significant Month-to-month Movements in the Raw Series at the 5% Level as a Function of the Correlation Between the Two Estimates Being Compared

Correlation	*Total number of movements*	*Ideal method*	*Overlapping intervals method*	*Assuming no correlation*
[-1, -0.5]	1	100%	100%	100%
(-0.5, 0]	204	74%	69%	74%[1]
(0, 0.5]	346	66%	52%	63%
(0.5, 1]	2329	83%	45%	55%

Table 1 presents the results for the movement in the raw series as a function of the correlation between the two estimates being compared. The vast majority of the month-to-month movements in the raw series involve a large correlation (greater than 0.5). Most of

[1] The result of the hypothesis tests of the "Assuming no correlation" alternative were always in agreement with that of the ideal method at the 5% level (they could have been more liberal).

the movements are also significant at the 5% level with the ideal method. This is not a surprising result since the movement could be due for the largest part to the seasonal component. The two alternatives most often agree with the ideal method for correlations of 0.5 or less. However for correlations greater than 0.5, the two alternatives often fail to detect the movements.

Table 2: Percentage of Significant Month-to-month Movements in the SA Series at the 5% Level as a Function of the Correlation Between the Two Estimates Being Compared – Using the Non-robust Estimator of the Irregular Component Variance

Correlation	*Total number of movements*	*Ideal method*	*Overlapping intervals method*	*Assuming no correlation*
(-0.5, 0]	421	7%	3%	7%[2]
(0, 0.5]	949	6%	1%	4%
(0.5, 1]	1510	7%	0%	0%

Table 2 presents the results for the movement in the SA series when the non-robust version of the estimator of the variance of the irregular component is used. Compared to Table 1, the correlations are reduced quite a bit because of the contribution of the model variance component. The percentage of significant movements is very close to the 5% level. Note that it was verified that almost every outlier identified in the reg-ARIMA procedure of X-12-ARIMA are involved in at least one significant month-to-month movement. One reason for the small percentage might be the span of the series studied. The series are only 6 years long, which is pretty short to estimate the seasonal patterns. Expanding the span further should improve the quality of the decomposition of the series and could make more month-to-month movements significant. Table 2 also shows that the two alternatives are not effective at detecting the significant movements, especially for large correlations.

Table 3: Percentage of Significant Month-to-month Movements in the SA Series at the 5% Level as a Function of the Correlation Between the Two Estimates Being Compared – Using a Robust Estimator of the Irregular Component Variance

Correlation	*Total number of movements*	*Ideal method*	*Overlapping intervals method*	*Assuming no correlation*
(-0.5, 0]	339	9%	5%	10%[3]
(0, 0.5]	858	10%	3%	7%
(0.5, 1]	1683	13%	0%	0%

Table 3 presents the results corresponding to those of Table 2, but using instead a robust version of the estimator of the variance of the irregular component. The down-weights used in matrix Δ are those of Table C17 of X-12-ARIMA, except for reference periods where an outlier was found by the reg-ARIMA procedure, in which case the weight was set to 0. When using such an estimator, there is a risk of under-estimating the variance if down-weighting is too pronounced. Table 3 can thus be seen as results obtained with a liberal estimator of the model-based component of variance and the truth may lie between the

[2] The results of the hypothesis tests of the “Assuming no correlation” alternative were more liberal than the ideal method at the 5% level only twice.

[3] The results of the hypothesis tests of the “Assuming no correlation” alternative were more liberal than the ideal method at the 5% level only thrice.

results of this table and those of Table 2. Though the percentages in the table are higher, it can be seen that the results of Table 3 are not drastically different from those of Table 2.

Table 4: Percentage of Significant Month-to-month Movements in the SA Series at the 5% Level as a Function of the Correlation Between the Two Estimates Being Compared, Taking Into Account the Model-based Variance Component Only – Using a Robust Estimator of the Irregular Component Variance

Correlation	*Total number of movements*	*Ideal method*	*Overlapping intervals method*	*Assuming no correlation*
[-1, -0.5]	11	0%	0%	0%
(-0.5, 0]	2576	24%	15%	27%[4]
(0, 0.5]	93	28%	22%	28%
(0.5, 1]	200	100%	100%	100%

Finally, Table 4 presents the results when the design-based component of the variance of the estimated SA series is set to 0. This is done to show how many significant month-to-month movements should be expected if a census of the population were done instead of a sample survey. Note that under a census, every raw movement would be significant in Table 1. Table 4 shows that most correlations would become null or negative and that between 2 or 3 times more significant month-to-month movements could be observed. The two alternatives to the ideal method would be detecting the significant movements well, especially the one assuming no correlation.

5. Conclusion

Some conclusions can be made in light of the results presented in this paper with respect to the quality indicators provided to time series data users at Statistics Canada. The first one is that providing quality indicators based on the estimated variance of the raw series along with the SA estimates might lead the data user to underestimate the variance of the SA estimates, in particular when the sampling fraction is large. However, this underestimation might be negligible for household surveys with small sampling fractions. Indeed, although no empirical results were presented in the paper for the LFS, our studies indicate that the model component of variance represents at most 5% of the total variance of the SA estimates for this survey.

The second conclusion that can be made is that when comparing estimates of two different reference periods of a series, using an approach not requiring the estimation of the covariance between the two estimates might be very conservative. In the case of the MRTS, a small fraction of the month-to-month movements were found to be significant. When this is the case, one would not want to be too conservative in his or her conclusions.

It would be straightforward to extend some of the developments presented in this paper to three Statistics Canada surveys that have a strategy very similar to that of the MRTS: the Monthly Wholesale Trade Survey, the Monthly Survey of Manufacturing and the Monthly Survey of Food Services and Drinking Places. Extension to other Statistics Canada surveys could also be considered. This might be more difficult to achieve, however, for surveys

[4] The results of the hypothesis tests of the "Assuming no correlation" alternative were more liberal than the ideal method at the 5% level 80 times.

that do not have monthly sample overlap schemes which yield variance matrices of the raw series as simple as those of the MRTS.

Statistics Canada's exploration of variance estimation of the SA series will be continued with the study of more volatile series (e.g. the domains formed by crossing a geography with an industry/NAICS in the MRTS). One would expect the model-based variance component to be a decreasing function of the population size and the design-based variance component to be a decreasing function of the sample size. Moreover, the analytical potential of drilling down series to smaller domains might be limited if variance increases and power decreases too much. Increasing the span to its full length (i.e. adding years 2004 to 2012) in the MRTS study is also of interest. This should improve the quality of the decomposition of Y into its components and help lower the model-based variance component. Furthermore, other variance structures of the irregular component that were suggested in the literature could be studied. For example, Pfeffermann and Scott (1997) suggest modeling I using an MA model.

Acknowledgements

The authors would like to thank Zdenek Patak, Steve Matthews and Susie Fortier for their direct and indirect contribution to this paper through their previous work on variance estimation of seasonally adjusted series (see Fortier and Matthews, 2016).

References

Bell, W.R. (2012). Unit root properties of seasonal adjustment and related filters. *Journal of Official Statistics*, 28, 441-461.

Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74, 829-836.

Dagum, E.B., and Cholette, P. (2006). *Benchmarking, Temporal Distribution and Reconciliation Methods of Time Series*. Springer-Verlag, New York, Lecture notes in Statistics, #186.

Deming, W.E. (1943). *Statistical Adjustment of Data*. New York: Wiley.

Findley, D.F., and Martin, D.E.K. (2006). Frequency domain analysis of SEATS and X-11/X-12-ARIMA seasonal adjustment filters for short and moderate-length time series. *Journal of Official Statistics*, 22, 1-34.

Findley, D., Monsell, B., Bell, W., Otto, M. and Chen, B. (1998). New capabilities and methods of the X-12-ARIMA seasonal-adjustment program. *Journal of Business & Economic Statistics*, 16(2), 127-152.

Fortier, S., and Matthews, S. (2016). Estimating the variance of seasonal adjustment: evaluating available methods. *Statistics Canada*, Internal document presented to the Advisory Committee on Statistical Methods of Statistics Canada.

Gambino, J., Kennedy, B. and Singh, M.P. (2001). Regression composite estimation for the Canadian Labour Force Survey: evaluation and implementation. *Survey Methodology*, 27, 65-74.

Ladiray, D., and Quenneville, B. (2001). *Seasonal Adjustment with the X-11 Method*. New York: Springer-Verlag.

Laroche, R. (2019). Methodology of the Monthly Retail Trade Survey. *Statistics Canada*, Internal document.

Pfeffermann, D., and Scott, S. (1997). Variance measures for X-11 seasonally adjusted estimators: some new developments with application to labor force series. *Proceedings of the Business and Economic Section of the American Statistical Association*.

Quenneville, B. (2013). Estimating variance in X-12-ARIMA. *Statistics Canada*, Unpublished manuscript.

Rao, J.N.K., and Wu, C.F.J. (1988), Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83, 231-241.

Rao, J.N.K., Wu, C.F.J. and Yue, K. (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology*, 18(2), 209-217.

Renaud, M., and Laroche, R. (2016). The use of administrative data in business surveys: the Statistics Canada experience. *Proceedings of the Fifth International Conference on Establishment Surveys* (*ICES-V*), June 20-23, 2016.

Roberts, G., Kovacevic, M., Mantel, H. and Phillips, O. (2001). Cross-sectional inference based on longitudinal surveys: some experiences with Statistics Canada surveys. *Proceedings of the Federal Committee on Statistical Methodology Research Conference*, November 14-16, 2001.

Särndal, C.-E., Swensson, B. and Wretman. J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag, Inc.

U.S. Census Bureau (2017). X-13ARIMA-SEATS Reference Manual, Version 1.1, Washington, DC.