

Labour Statistics: Technical Papers

Transition of Labour Force Survey Data Processing to the Social Survey Processing Environment (SSPE)

Release date: February 8, 2019



How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at STATCAN.infostats-infostats.STATCAN@canada.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- | | |
|---|----------------|
| • Statistical Information Service | 1-800-263-1136 |
| • National telecommunications device for the hearing impaired | 1-800-363-7629 |
| • Fax line | 1-514-283-9350 |

Depository Services Program

- | | |
|------------------|----------------|
| • Inquiries line | 1-800-635-7943 |
| • Fax line | 1-800-565-7757 |

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "[Standards of service to the public](#)."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

© Her Majesty the Queen in Right of Canada as represented by the Minister of Industry, 2019

All rights reserved. Use of this publication is governed by the Statistics Canada [Open Licence Agreement](#).

An [HTML version](#) is also available.

Cette publication est aussi disponible en français.

Table of Contents

Overview of LFS processing.....	4
LFS processing systems.....	4
Development and testing of LFS processing using SSPE.....	5
Results and analysis	7
Estimates of particular interest.....	11
Conclusions	11
Appendix A.....	12

Transition of Labour Force Survey Data Processing to the Social Survey Processing Environment (SSPE)

Overview of LFS processing

Each month, the production of statistics from the Labour Force Survey (LFS) involves three broad sets of activities: 1) data collection, during which interviews are conducted with approximately 50,000 households over a 10-day period, 2) data processing, during which data are verified and corrected when required to produce microdata files, and 3) the production of statistical tables and analytical products, during which products including *The Daily* are prepared and loaded to the Statistics Canada website within 10 days after the end of collection.

LFS data processing consists of several steps. Some of these are carried out on a daily basis, concurrently with data collection, while others are done once per month, on the day after the completion of data collection. The daily steps include:

1. Receipt of data from collection.
2. Coding of responses to questions with “Other – specify” answer categories.
3. Editing of demographic data (age, sex, family relationships) to ensure validity and consistency.
4. Imputation of educational data.
5. Editing of labour force data.
6. Coding of industry and occupation responses.

Once collection is complete, the following monthly processing steps are carried out:

1. Imputation of remaining demographic data.
2. Imputation of labour force data.
3. Calculation of derived variables.
4. Weighting and variance estimation.

Imputation is the replacement of missing data with plausible values. When individual data items are missing, an effort is made to find a suitable donor. Donors are records that are judged similar based on a series of characteristics, and that have valid and consistent data that could be used to reasonably replace the recipient record's missing value. If a suitable donor cannot be found, a larger block of the recipient's data may be stricken out and replaced with a donor's block of data; for this 'whole-record imputation', the criteria of similarity is typically relaxed until the best possible donor is found.

LFS processing systems

The computer system currently involved in LFS data processing - known as the Head Office Processing System (HOPS) - has been in place for more than 25 years. During this time, a number of modifications have been introduced and the system has been maintained to ensure the continued accuracy and robustness of LFS production. Beginning in January 2019, LFS processing will be transitioned from HOPS to the Social Survey Processing Environment (SSPE), a generalized survey data processing system used by a number of other Statistics Canada surveys. This transition of LFS data processing to SSPE offers a number of advantages:

- Whereas the design of HOPS is relatively complex and integrated, with changes in one step having potential impacts on others, SSPE allows for a flexible and modular approach to LFS processing, with individual steps being easier to isolate, analyse and enhance as required.
- While HOPS is used only by LFS, SSPE is used by most of Statistics Canada's social surveys and experience in data processing using SSPE is relatively broad. As a result, transition to SSPE reduces LFS reliance on system-specific and person-specific knowledge, thereby reducing risks related to staff retention and turnover.

- Because SSPE is by design integrated with other generalized systems—including the Canadian Census Edit and Imputation System (CANCEIS) for imputation and the Common Coding Environment (CCE) for coding—overall resources required for systems development and maintenance are reduced.

Development and testing of LFS processing using SSPE

The project to transition LFS data processing began in April 2015. The major phases of the project included:

1. Designing 'processing maps' for daily and monthly LFS processing. Within the SSPE generalized environment, a generic set of instructions is available for each step in data processing (for example, receiving data from the collection systems into the processing system). For each of these steps, survey-specific instructions and rules must be developed and a set of steps or a 'map' must be designed to determine and manage the sequencing and execution of steps.
2. Modular testing of each processing step, to ensure that at each processing step, data was input, modified or transformed, and output as expected.
3. Integrated testing, to evaluate the cumulative impact of all processing steps on LFS estimates.
4. Operational readiness testing, to ensure that LFS data processing could be completed within existing time and resource constraints; for example, that each morning of daily processing, data could be received from collection and transferred to the group responsible for industry and occupation coding within an acceptable timeframe.

Design of LFS processing map

The LFS has a very high standard for data quality, and unique editing challenges and opportunities exist since respondents are interviewed for six months in a row. For example, in cases where a sampled household responds to LFS in a given month and is a non-respondent the following month, data from the first month may, if specific conditions are satisfied, be used to complete the missing data in the second month.

In designing the SSPE processing maps, every effort was made to retain these underlying conditions and rules, or processing methodology, thereby minimising the impact of the transition on the quality of LFS data and the level of LFS estimates. In some situations, opportunities existed to streamline or improve the underlying methodology without affecting LFS estimates (for example, the replacement of some manual editing of records with deterministic edit rules) while in others it was necessary to slightly modify some aspects of the methodology (for example, the specific conditions under which a partial record as opposed to an entire record is corrected through imputation).

Modular testing of LFS processing steps

Each step in the SSPE daily and monthly processes map was developed and tested in a modular fashion. The objective of modular testing was to ensure that each for each step data was input, transformed and output according to specifications.

The LFS Redesign team conducted modular testing from April 2016 to April 2017. During this period, each module was tested independently and validated against specification documentation. After the success of the modular testing, approval was granted to begin fully integrated testing in May 2017.

Integrated testing of LFS processing maps

The primary objective of integrated testing was to generate test data by running all processing steps in sequence and then evaluating whether transitioning LFS processing would introduce any discontinuities in LFS time series. Doing so involved overcoming a number of operational challenges, including:

- Comparing a given month of LFS estimates using two processing systems (HOPS and SSPE) involves a non-negligible investment of time to copy and configure large datasets and to then execute a large number of sequential processing steps. It was necessary to develop test strategies which made efficient use of the time available.

- Some daily LFS processing steps—notably coding of “Other – specify” responses, manual correction of demographic data and manual coding of industry and occupation responses—are manual processes which: 1) involve significant human input and decision-making each month, and 2) have a non-negligible impact on LFS estimates. It was necessary to develop test strategies which included an evaluation of the impact of changing the systems used to execute these steps while respecting time and resource constraints.
- Under some conditions some information (such as industry and occupation of employment) is collected in the first month that a household is interviewed and is carried forward to subsequent months. The impact of SSPE transition therefore had to be evaluated both in terms of its impact on the first month post-transition and on possible impacts in the months following transition.
- Many key LFS estimates are seasonally adjusted to permit analyses controlling for the impact of large seasonal variations. It was necessary to develop approaches to evaluate not only the short-term impact of transition to SSPE but the long-term impact on seasonal patterns.

To overcome these challenges, a combination of types of integrated testing were completed, including:

- **Multi-year testing**, in which more than nine years of data (January 2008 to May 2017) were produced by using outputs of HOPS daily processing as inputs to the monthly SSPE processing map. This was evaluated as being an efficient and effective means of evaluating the impact of SSPE transition on:
 - ▶ the imputation of labour force information
 - ▶ long-term seasonal patterns.

Key findings from this testing included:

- ▶ During initial testing, some differences in the level of some LFS estimates were observed. This resulted in further examination and adjustment of imputation parameters. Testing following these changes indicated no unacceptable impact on estimates. More detail on this matter is presented in the “**Results and analysis**” section of this document.
- ▶ No impact of transition month (data were produced using various transition months and carried forward for six months each; there was no evidence that SSPE behaved differently depending on the starting month).
- ▶ No impact on seasonal patterns.
- **Multi-month testing – Phase 1**, in which data were produced by executing both the daily and monthly SSPE processing maps, with the exception of steps involving manual inputs or human decision-making. For these steps, inputs and outputs were simulated using HOPS values. Twenty consecutive months of data (from January 2016 to August 2017) were produced (data produced in winter 2018).

Key findings from this testing included:

- ▶ Some differences in HOPS and SSPE estimates of wages. This led to examination of causes of these differences and adjustments to editing and imputation parameters.
- ▶ Differences between estimates obtained using SSPE and those obtained using HOPS were acceptable.
- ▶ Differences between SSPE and HOPS were not systematically positive or negative and did not exhibit a pattern, indicating that transition to SSPE does not introduce bias.
- ▶ No impact on the sampling variances associated with LFS estimates.
- **Multi-month testing – Phase 2**, in which data were produced in the same way as Multi-month Testing - Phase 1 but with adjustments made to editing and imputation parameters affecting the estimation of wages. Nine consecutive months of data (October 2017 to April 2018) were produced in summer 2018.

Key findings from this testing included:

- ▶ No significant differences in wage estimates produced by HOPS and SSPE.
- ▶ No change in the impact associated with SSPE transition introduced by a change in transition date.

- **Fully-integrated testing**, in which data were produced by executing both the daily and monthly SSPE processing maps, including steps involving manual inputs or human decision-making.

The first of these tests was conducted in the summer of 2017 and produced data for three months (March 2017 to May 2017). The primary objective of this testing was to test the flow of data through all steps in the daily and monthly processing maps. Testing led to identification of a number of areas where refinements and enhancements were required.

A second round of fully-integrated testing was completed in the summer of 2018. Four separate months of LFS data were processed using the full SSPE daily and monthly processing maps, with the exception of industry and occupation, with approximately half of records being coded using SSPE and the remainder being simulated using HOPS data. Whereas in multi-month testing a single transition month was simulated and SSPE data were used in subsequent months, in each of the four months of fully-integrated testing, the first month of SSPE was simulated.

Key findings of this second round of fully-integrated testing included:

- ▶ Differences in estimates obtained using HOPS and SSPE were acceptable (further details are presented in the “**Results and analysis**” section).
- ▶ Consistency in the impact of SSPE transition across the 4 months.

Operational readiness testing

The fully-integrated testing conducted in the summer of 2018 also served as an operational readiness test, to evaluate whether, from an operational point of view, LFS data processing could be transitioned to SSPE without jeopardising the timeliness or robustness of the LFS production process. Specific criteria were evaluated including the adequacy of documentation and procedures manuals, the training of staff to execute LFS processing using SSPE, and the stability and performance of IT infrastructure.

This testing revealed that all conditions were satisfied and that from an operational point of view LFS data processing could be transitioned to SSPE beginning in January 2019.

Results and analysis

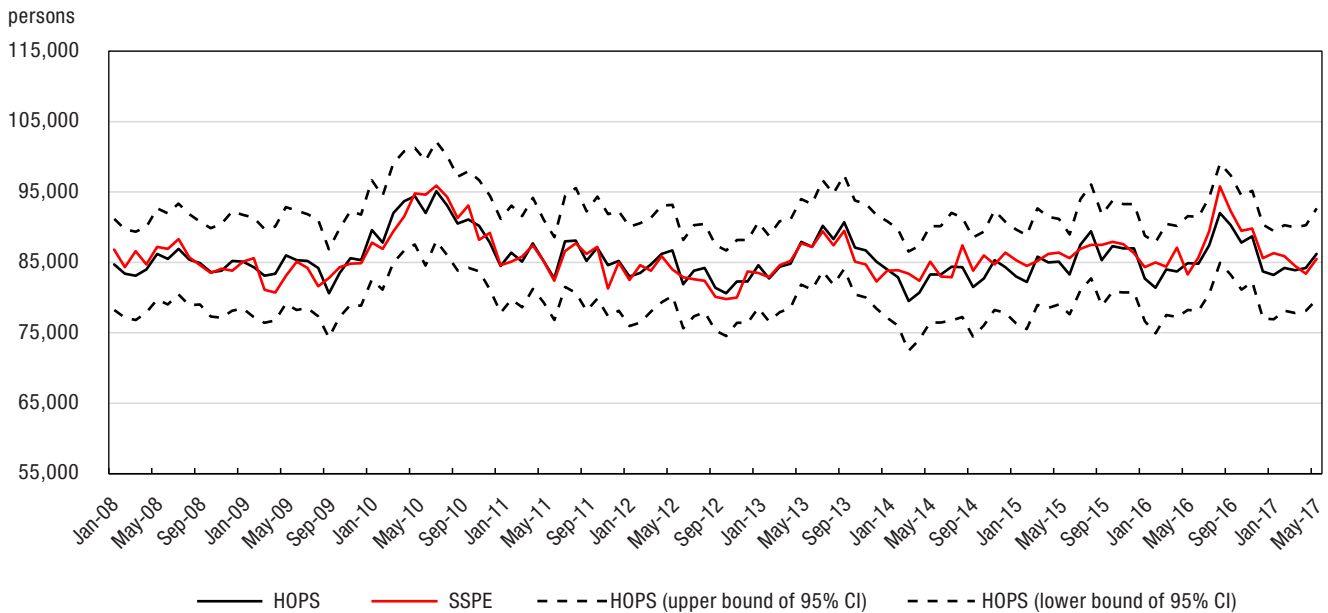
Using the data from the various types of integrated testing, the impact of transitioning LFS processing to SSPE on the following was evaluated:

- For a given month, estimates of levels (e.g., total number of self-employed persons) and of rates (e.g., unemployment rate);
- Estimates of month-to-month changes;
- Estimates of year-over-year changes;
- Seasonally adjusted time series;
- Measures of the precision of estimates.

Some estimates, like the unemployment rate and the total number of employed persons, are considered key estimates and were therefore scrutinized in greater detail. Their results were compared at more disaggregated levels, based on key analytical domains. The list of key estimates, as well as the list of key domains, are provided in “**Appendix A**”.

The results presented here are typical of the very large number of analyses that were carried out (see “**Appendix A**” for a list of key estimates and domains). Figure 1 shows the number of self-employed persons in Manitoba from January 2008 to May 2017. This data series is based on Multi-year testing (where the main goal was to evaluate the impact of SSPE transition on the imputation of labour force information). The estimates have not been seasonally adjusted. The solid black line represents the HOPS estimates, and the dashed black lines represent the 95% confidence intervals of the HOPS estimates. The solid red line represents the corresponding SSPE estimates. Since the confidence intervals around the SSPE estimates had the same width as the ones around the HOPS estimates, they are not shown.

Figure 1
Number of self-employed in Manitoba, multi-year testing, January 2008 to May 2017

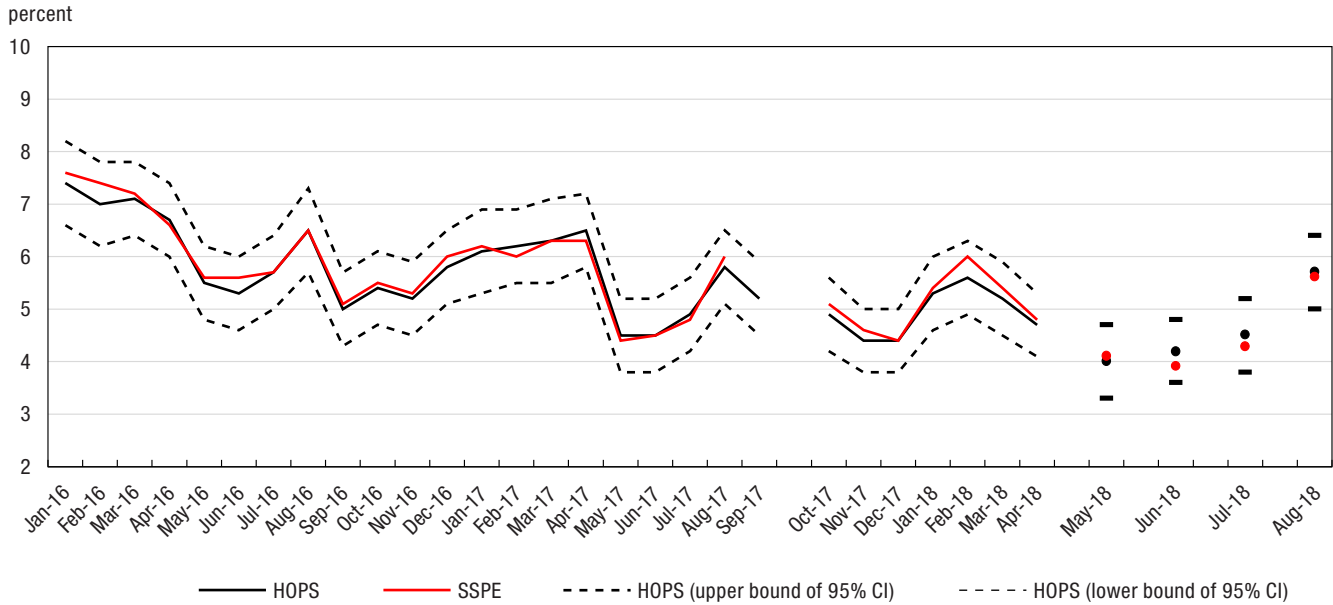


Source: Statistics Canada, Labour Force Survey.

Figure 1 demonstrates that the labour force imputation did not have a dramatic effect on the estimates, as the SSPE estimates track well with those from HOPS, staying within the 95% confidence intervals throughout the 9-year series.

Figure 2 represents the unemployment rate in Quebec, using data from Multi-month testing. Once again, the black lines represent the HOPS estimates and their 95% confidence intervals, while the red line represents the SSPE estimates. The first section, from January 2016 to August 2017, represents Phase 1 of Multi-month testing. Following the gap in September 2017, data is from Phase 2 and represents a fresh start in October 2017 (previous month data generated from HOPS), carried through until April 2018. The final four sets of points represent four individual starts associated with Fully-integrated testing (May 2018 until August 2018).

Figure 2
Unemployment rate in Quebec, multi-month testing and fully-integrated testing, January 2016 to August 2018

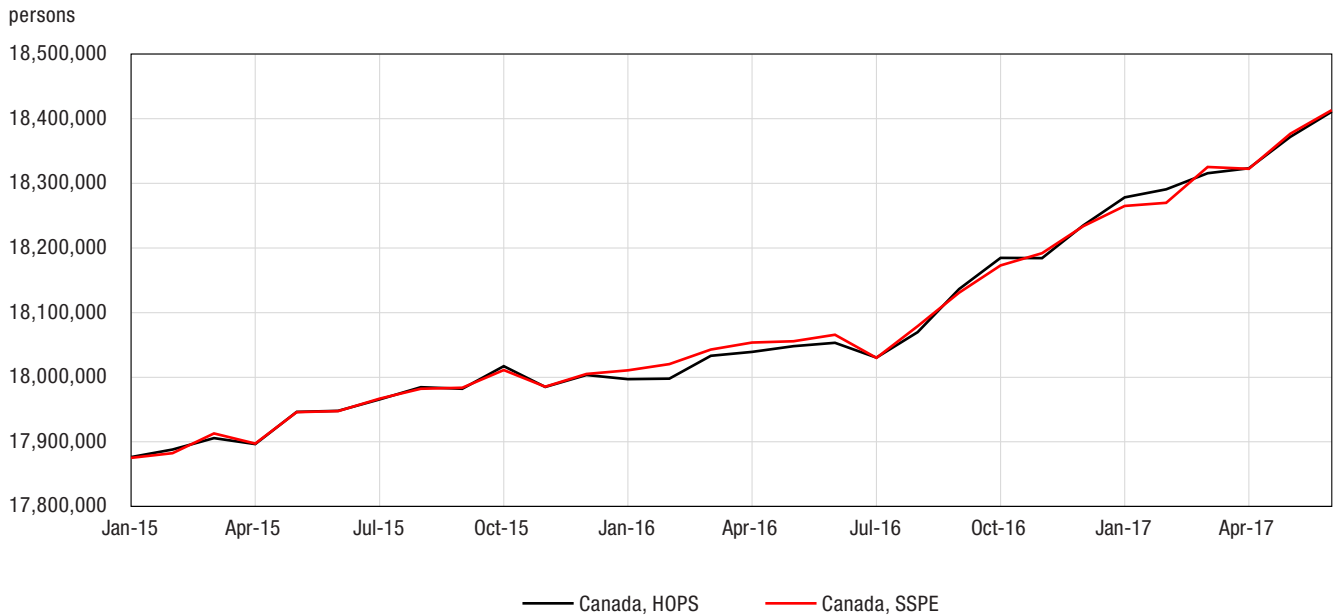


Note: Data represent the results from six distinct tests.
Source: Statistics Canada, Labour Force Survey.

The conclusions are the same as with the data shown in Figure 1: the SSPE data is not significantly different from the HOPS data, staying within the 95% confidence intervals generated by the bootstrap method on the HOPS data.

Some LFS data are seasonally adjusted, in order to remove seasonal patterns and provide a better representation of trends. The associated seasonal adjustment models—which are external to HOPS and therefore not affected by the transition to SSPE—were applied to the SSPE test data, and a series of comparisons were carried out. The analysis revealed no structural changes in the data, and the seasonality of the SSPE data was captured well by the existing seasonal adjustment models. Figure 3 shows that the seasonally-adjusted series of employment in Canada from January 2015 to June 2017, based on SSPE, is statistically equivalent to the analogous LFS data.

Figure 3
Seasonally adjusted level of employment in Canada, January 2015 to June 2017



Source: Statistics Canada, Labour Force Survey.

Month-to-month changes are of particular interest for many analysts of LFS data. Figure 4 shows the monthly change in total full-time employment in Canada in 2016. The graphic demonstrates that the discrepancies between the two series are not significant enough to imply a change in the story told by the data.

Figure 4
Month-to-month change in full-time employment, Canada, 2016



Source: Statistics Canada, Labour Force Survey.

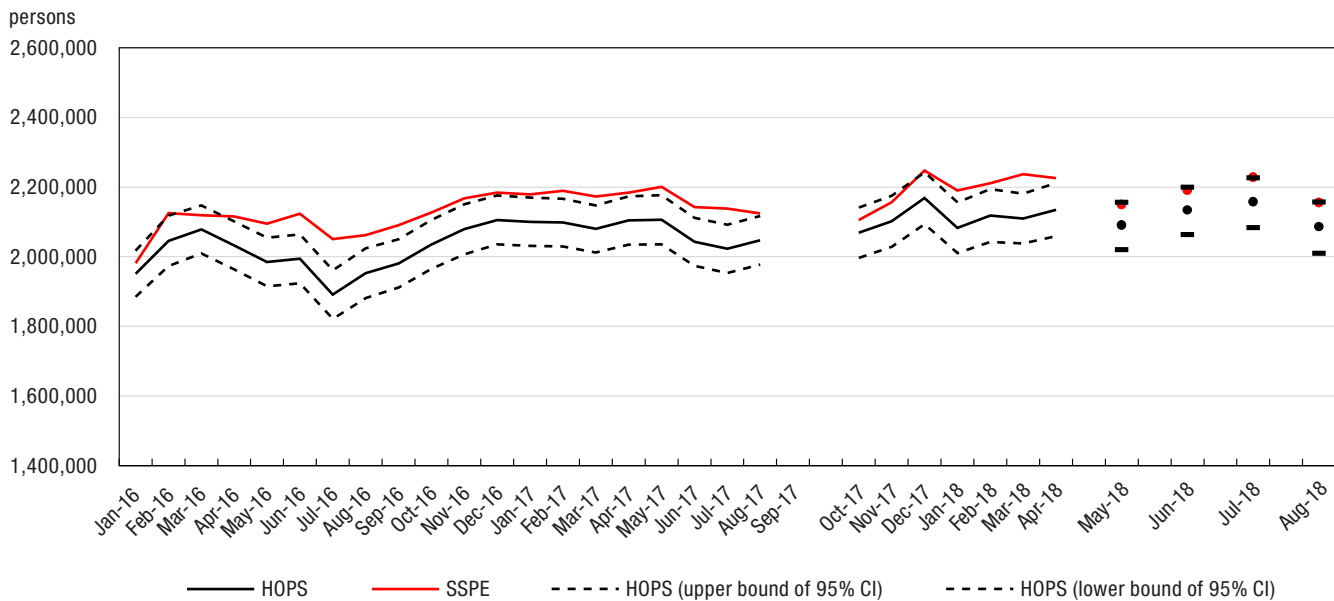
Estimates of particular interest

Although integrated testing indicates that for most LFS estimates differences between HOPS and SSPE estimates are very small, users of LFS data should be aware of some specific testing results:

Full-time and part-time employment: While the underlying LFS processing methodology will not change with the transition to SSPE, the CANCEIS system used for imputation with SSPE is more efficient than HOPS at finding appropriate imputation donors. As a result, SSPE tends to preserve more respondent data while HOPS tends to resort more often to whole-record imputation. This results in more efficient imputation in SSPE for respondents with lower-prevalence characteristics, such as part-time employment. Testing indicates that this could result in a slight increase in part-time employment and a corresponding decrease in full-time employment. Related estimates such as the average number of hours worked could decrease slightly.

Firm size: As a result of correcting a specific edit in HOPS, the distribution of employment by firm size could be affected by the transition to SSPE, with an increase of approximately 3% in the number of employees working for firms having more than 500 employees. As shown in Figure 5, estimates calculated using data from Multi-month testing are consistently higher than those from HOPS, often falling just outside of the 95% confidence intervals.

Figure 5
Number of persons 15 years and over working in firms of 500 or more employees, Canada, January 2016 to August 2018



Note: Data represent the results from six distinct tests.
Source: Statistics Canada, Labour Force Survey.

The impact of transition to SSPE on these estimates will be monitored and, if appropriate, an historical revision of the series will be included in the planned 5-year rebasing of LFS estimates, currently planned for January 2020.

Conclusions

Based on extensive testing, the transitioning of LFS data processing from the current system (HOPS) to the Social Survey Processing Environment (SSPE) is expected to have minimal impact on LFS estimates and to be transparent to users of LFS data.

Appendix A

Key estimates

These estimates are considered the most important estimates published from LFS data.

Table 1
Key estimates

Estimate	Group
Total, labour force	1
Total, employment	1
Total, unemployment	1
Total, not in labour force	1
Unemployment rate	1
Participation rate	1
Employment rate	1
Total, full-time employment	2
Total, part-time employment	2
Total, employees	2
Total, public sector employees	2
Total, private sector employees	2
Total, self-employed	2
Total, employees by industry (16 levels)	2
Total, employees by occupation (10 levels)	2
Average wages	2
Average hours	3
Total, multiple job holders	3
Total, employees by unionization status (2 levels)	3
Total, employees by job permanence (2 levels)	3
Total, employees by firm size (4 levels)	3

Source: Statistics Canada, Labour Force Survey.

Key variables labelled 'group 1' were evaluated for all key domains. Those labelled 'group 2' or 'group 3' were formally evaluated at aggregate levels and for limited levels of disaggregation.

Key domains

These are the major geographic and demographic domains for which estimates are published.

Table 2
Key domains

Domain	Components
Geography	Provinces and territories
	Economic regions (ERs)
	Census metropolitan areas (CMAs)
	Employment Insurance economic regions (EIERS)
Age	15 years and over
	Youth: 15 to 24 years, 15 to 19 years, 20 to 24 years
	Core-aged: 25 to 54 years
	Senior: 55 years and over, 55 to 64 years, 65 to 69 years, 70 years and over
Sex	Men, 15 years and over
	Women, 15 years and over
Aboriginal and Immigrant status	Aboriginal population, 15 years and over
	Immigrant population, 15 years and over

Source: Statistics Canada, Labour Force Survey.