

The accuracy of employment indicators using bootstrapped variance estimation

Ólafur Már Sigurðsson and Anton Örn Karlsson

April 21, 2017

Introduction

A core duty of National Statistical Institutes is to provide users with high quality data to produce official statistics that are fit for use. Fit for use in the case of official statistics can mean a number of things but one important factor is the accuracy of the statistics, i.e. to what extent a particular measure actually measures what it is supposed to.

A recent user satisfaction survey conducted by Statistics Iceland suggested that while user satisfaction with official statistics is first and foremost influenced by the extent statistics fulfilled their needs and the comparability of the statistics. Accuracy of the figures was not a predictor of satisfaction. However, when asked about satisfaction with the service provided by Statistics Iceland, the accuracy of published statistics was the second strongest predictor of users satisfaction (only behind comparability). This indicates that users view accuracy of official statistics as an important factor in the service performed by Statistics Iceland. Users of official statistics get information from a variety of sources but they expect Statistics Iceland to provide a specific type of service which is to provide them with accurate statistics. To increase users satisfaction with the service provided by Statistics Iceland it is therefore important to focus on statistical accuracy in its production processes and dissemination. To do this it is important to apply sound methods in assessing accuracy. If the methods used to gauge accuracy are biased, any increase or decrease in quality of the statistics will go unnoticed - users will not realize the true quality of the statistics.

Statistics Iceland recently conducted a study aimed at estimating the variance of the yearly difference of annual unemployment in the Icelandic Labor Force Survey (ISLFS). Comparing different estimation methods, the study indicated that a bootstrapped variance estimation procedure produced low variance figures and seemed robust enough to be used in a variety of settings. Furthermore, the methodological advisory committee of Statistics Iceland suggested that this methodological framework should be used as a default method for variance estimation within the institute.

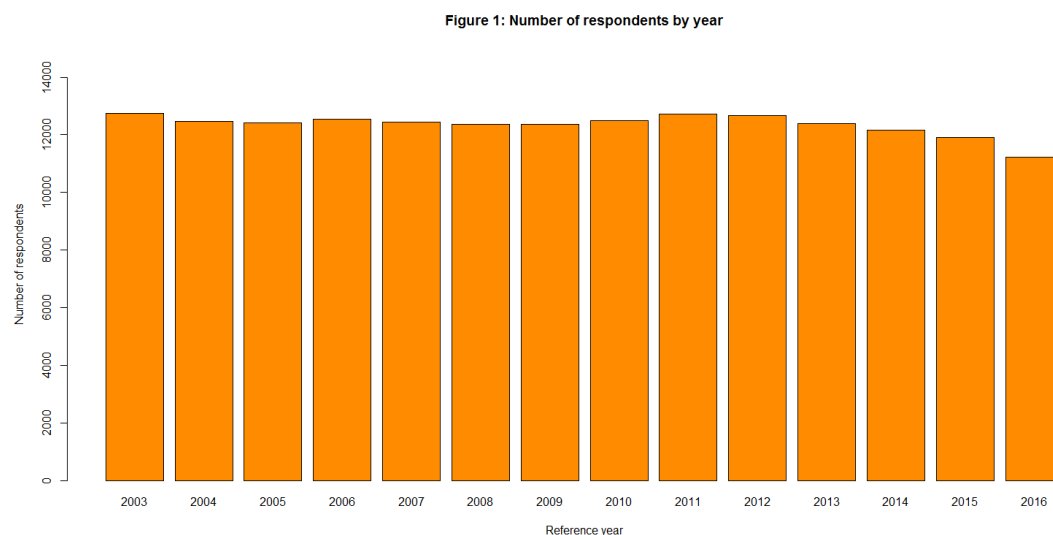
In this paper we would like to go further with this method and use it to estimate the variance of the yearly differences for other indicators of the ISLFS with the ultimate aim of providing users of official statistics with better information on the accuracy which should be helpful when using the statistics, e.g. for policy formulations based on differences in annual estimations from the ISLFS. Also important for Statistics Iceland; sound methods for variance estimation gives a better understanding of the accuracy of the estimators. In statistical software, often the estimation of variance is based on simple, exact methods. While these methods deliver unbiased unbiased estimates of variance under very specific

conditions, i.e. simple random samples with no item or unit non-response, in more complex situations, e.g. where non response weights or imputations have been applied, such forms of exact variance estimation are not suitable. In such cases the analyst has to use other types of variance estimation methods, like the bootstrap.

The bootstrap method for variance estimation is based on selecting a number of random samples (e.g. 1000), with replacement from the original sample. The size of each bootstrap sample is the same as the size of the original sample which is being sampled from. For each of the bootstrap samples, the statistic in question is calculated and then averaged over the entire set of bootstrapped samples, in effect creating a sampling distribution for this particular statistic. It can be shown that this creates a maximum likelihood estimation of the population distribution.

In this case, the statistics in question were taken from an internal variable calculated by Statistics Iceland which is called New-ILO (detailed ILO). New-ILO further divides the categories of the traditional labour market definition of the population; employed, outside the labour market and unemployed. The categories in New-ILO are: * Working full time - temporary hire * Outside the labour market * Part time employment * Ready to work, not searching * Searching for work, not ready to start * Underemployed - wants to work more * Unemployed * Working full time - indefinite, standard employment relationship

Data and results



The data for the study comes from the entire period of the longitudinal form of the ISLFS, from 2003-2016. Figure 1 shows the number of yearly responses recorded from the survey over the time period. The response rate of the survey has traditionally been in the range between 80-85%. Recent quarters in the data collection have, however, struggled to reach 70%.

Figure 2: Number of respondents by New-ILO categories

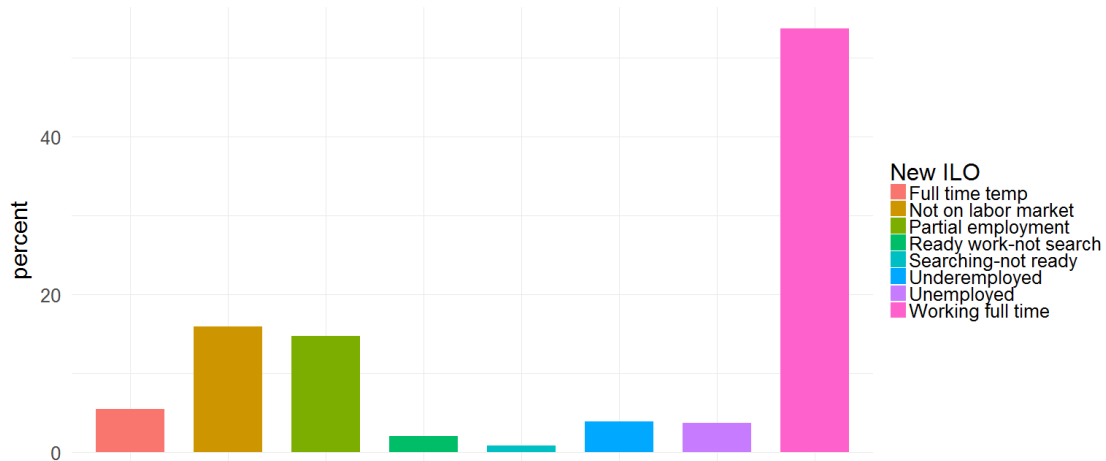


Figure 2 shows the total number of respondents (unweighted) for each of the categories of the new ILO variable. By far the largest group is individuals working full time indefinitely which constitutes roughly 53.73% of the total over the entire timeframe of the ISLFS.

Figure 3: Number of respondents by New-ILO categories, yearly percentages

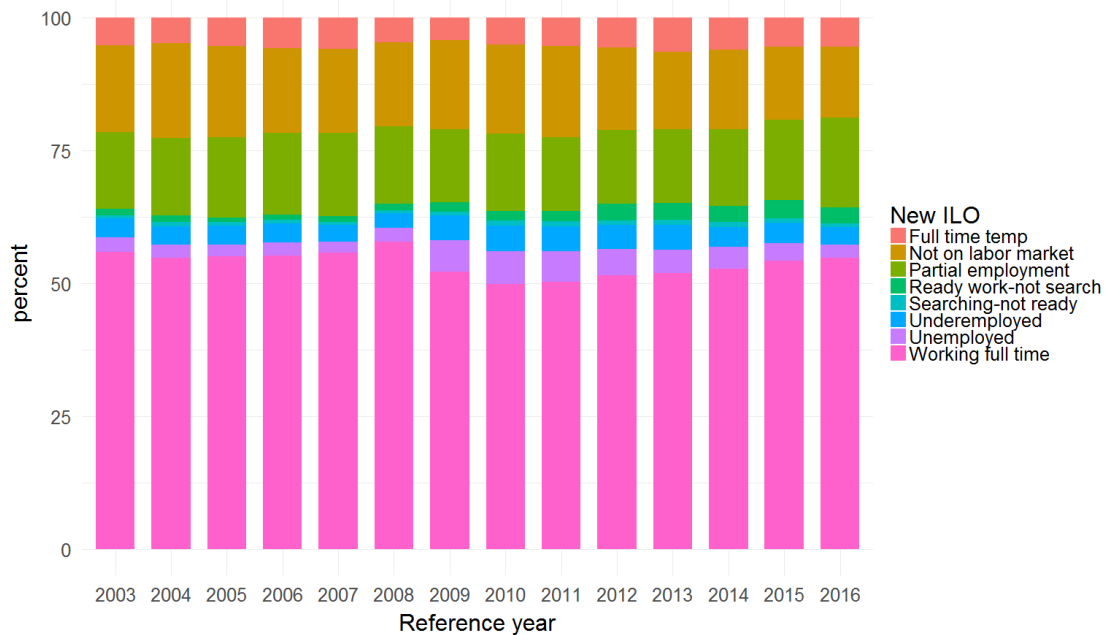


Figure 3 shows a stacked plot (in very fancy colors) indicating the yearly composition of the Icelandic population according to the New ILO variable from 2003 until 2016. The Icelandic financial crash is clearly evident in the figure, with changes in the labor market composition of the Icelandic population happening in 2009. For instance, the estimated percentage of working full time dropped from 57.9% in 2008 to 52.3% in 2009; the number of unemployed rose from 2.5% in 2008 to 5.8% in 2009, the estimated percentage of underemployed rose from 2.8% in 2008 to 5.8% in 2009 and a modest rise in the

percentage of individuals ready to work but not actively searching from 1.2% in 2008 to 1.7% in 2009. Additionally, most of these indicators did also increase from 2009 to 2010. The question remains, though, if these changes between years are significant or if they are merely an artifact of the survey. In other words, do these measures reflect actual changes in the population or do they indicate random variation from year to year. We proceeded to test this for four different population groups: Full-time employment, full-time temporary employment, part-time employment and underemployment. In all cases, weighted estimates of yearly percentages for each category in the population was compared to the percentage in the previous year with the variance estimated by bootstrap, using 1000 replicates for each comparison between years. For each comparison a confidence interval using the normal approximation was calculated as there was almost no difference between that and other types of confidence intervals that were examined (studentized bootstrap method, bootstrap percentile method, adjusted bootstrap percentile (BCa)) - which are all available in the boot package in R.

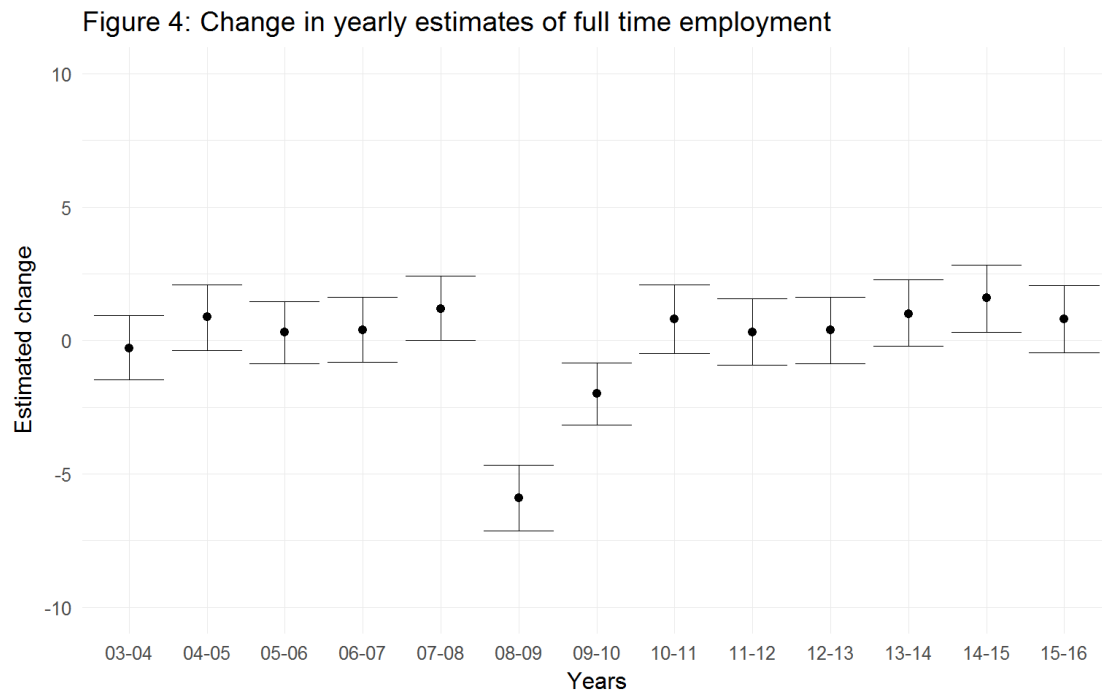
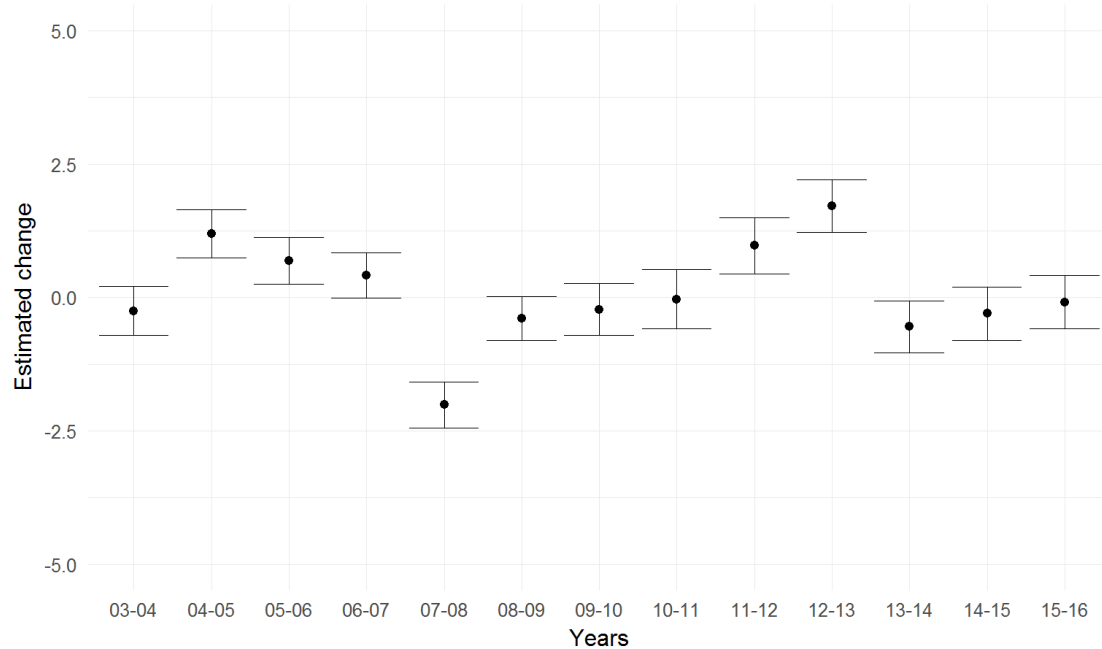


Figure 4 shows the yearly change in estimates of full-time employment, with 95% confidence intervals around the estimates. From 2003-2008 the annual change was not significant - indicating that the percentage of the Icelandic population in full time employment was relatively stable. The most dramatic change in the figure is between 2008 and 2009 where the percentage fell by 4.7-7.1 percentage points, and again by 0.8-3.2 percentage points between 2009-2010. The percentage has been stable since then, with only a modest increase of 0.3-2.8 percentage points increase between 2014-2015.

Figure 5: Change in yearly estimates of full time, temporary employment



In figure 5 are the bootstrapped annual changes in estimates for full time temporary employment with 95% normal confidence intervals. Among the main trends the figure shows is that there seem to be significant annual fluctuations in full time temporary employment in Iceland. While there is no difference noticeable between 2003 and 2004, there is a significant increase between 2004 and 2005, as well as 2005 and 2006, while this increase seems to be similar. Between 2007 and 2008 there is a significant downturn in the number of full time temporary workers in Iceland. Not until 2011-2012 are there indications of a rise in full time temporary and again between 2012-2013 - while the difference in the annual changes are not significant. Over the past years (from 2014 and onward) there has not been a significant annual change in full time temporary employment.

Figure 6: Change in yearly estimates of part-time employment

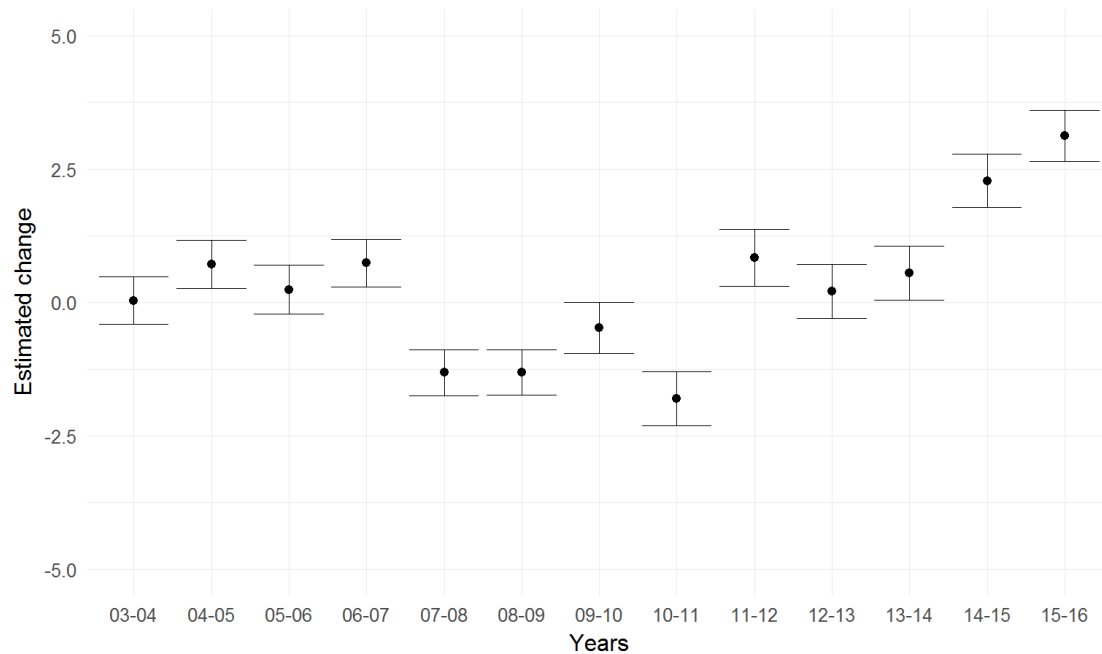


Figure 6 shows the bootstrapped changes in yearly estimates of part time employment with the corresponding 95% confidence intervals. For the first four years of the ISLFS the annual difference in part time employment is either close to zero or a nudge above. Between 2007-2008 and again between 2008-2009, part time employment shows a similar pattern with a decrease between 0.8-1.7 percentage points between the years. Another decrease happens between 2010-2011 or between 1.3-2.3 percentage point decrease. Since then, the number of part time employees in Iceland has been increasing (except for between 2012-2013 when the change was non significant) with the changes between 2014-2015 being 1.8-2.9 and 2.7-3.6 between 2015-2016, in percentage points. From 2003 the annual increase of part time employees in Iceland has never been higher than between the years 2014-2015 and 2015-2016.

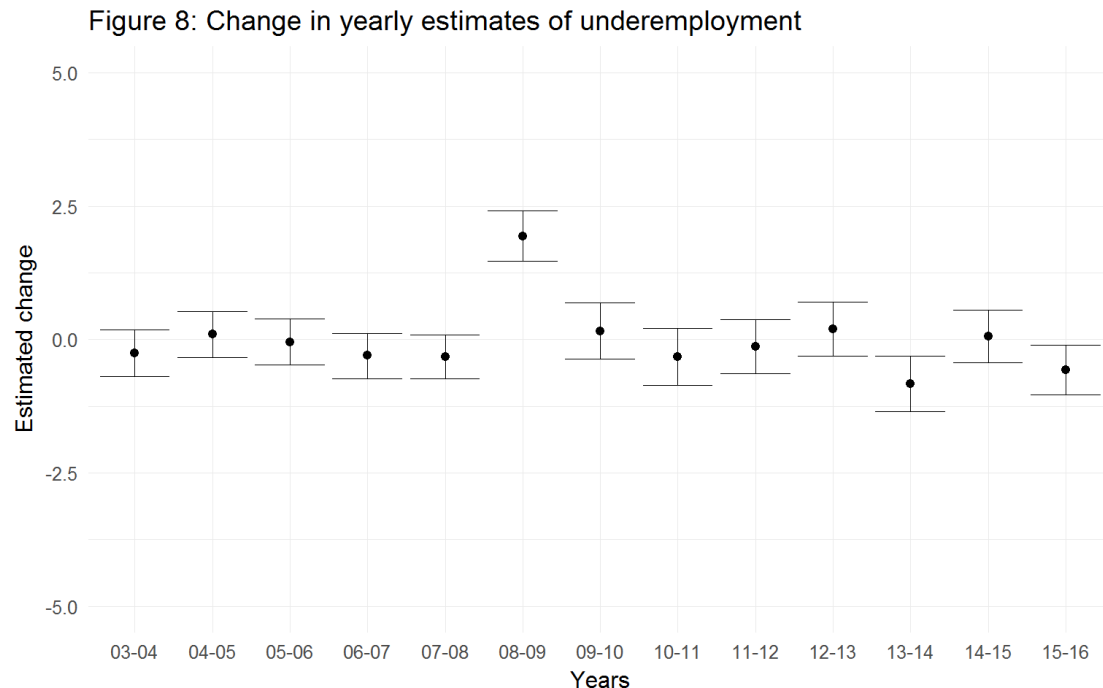


Figure 7 shows the bootstrapped changes in yearly estimates of part underemployment with the corresponding 95% confidence intervals. The figure shows first and foremost that the annual estimates of underemployment in Iceland are fairly stable from 2003 to 2016. Only in two cases are there significant differences in annual estimation of underemployment between two years: Between 2008-2009 underemployment rose significantly by 1.5-2.4 percentage points while between 2013-2014 underemployment decreased by between 0.3-1.4 percentage points.

Discussion

Taken together, these results show a clear picture of the annual changes that have happened in the Icelandic labour force from the year 2003-2016. Additionally, they give the user a chance to examine the data in an accurate way and draw valid conclusions from the presentation of confidence intervals for annual changes. The usability of the figures is enhanced because of the sound methodology that was used in estimating variance and computing the confidence intervals.

Because of the small size of the Icelandic population and (thankfully!) a low rate of unemployment, analysis of the Icelandic labour market can be a bit tricky. By incorporating more information (e.g. by creating the New-ILO variable) and by using specific methods to estimate the variance in the data, it is possible to do more fruitful analysis with the data and provide users with more helpful information. Statistics Iceland has used Small Area Estimation methods in its analysis which also provides possibilities to make more accurate information from small subgroups in the sample.

What is lacking is to incorporate the longitudinal aspect of the ISLFS in the variance estimation method and it is still unknown to what extent that affects the estimation. Some

advice on how to proceed with this would be greatly appreciated i.e. how add the longitudinal character of the data collection within a variance estimation framework using bootstrap and/or how to estimate the effect of the longitudinal aspects of a survey on variance estimation.

This way of presenting the results is just one way of many for which to present uncertainty in published figures. If this is the best possible way - or even satisfactory is unknown. We are very interested in hearing about how uncertainty in statistical data can be effectively communicated to users of official statistics. Any practical know-how and, especially empirical data that has been collected on this issue would be much appreciated.