

Impact of 2021 Census of Population and Housing on the Labour Force Survey (LFS) headline indicators

Authors: Tania Borg and Charlene Abela

1. Introduction

The Labour Force Survey (LFS) is a household survey which provides quarterly and annual statistical estimates of the labour status of persons aged 15 years and over in accordance with the Integrated European Social Statistics (IESS) framework regulation (EU) 2019/1700¹.

The LFS is conducted by selecting a random sample of 3,200 households per year, selected from a sampling frame of households which is maintained by the National Statistics Office (NSO). Prior 2023, the sample was selected through systematic random sampling, while from 2023 onwards, the sample was selected through a stratified random sampling. The LFS is a panel survey with a panel rotation of 2-(2)-2. Every quarter of the year, 800 households are chosen to participate in the first and second panel. Since panels are rotated, the households selected to participate in the first and second round are left out of the LFS for the two following quarters and then reintroduced to participate in the next two quarters. A household participating for the first time in one quarter is therefore contacted a further three times after their first participation. Since 2020, the mode of data collection for all households is computer assisted telephone interview (CATI). Data is collected on all members in the household, thus the probability of selection for individual household members is equal to the probability of selection for the household. Given that the LFS is conducted on a sample, its responses must be weighted and then calibrated with the latest estimates of the household population to produce representative estimates of the labour market.

Following the publication of the 2021 Census of Population and Housing, the demographic timeseries, including household population estimates, were revised to align the demographic series between the censuses (<https://nso.gov.mt/intercensal-population-revisions-2012-2021/>). As a result, a reweighting exercise was conducted to align results from the LFS to these demographic revisions thus mitigating any possible break in time series to the main LFS indicators. Reweighting was carried out on all quarterly datasets for the period Q1 2012 – Q3 2023. A comparative analysis was conducted to assess the impact of the revised weights on the headline indicators from the LFS, mainly employment and unemployment rates. This technical paper will firstly provide a brief overview of the methodology used to weight results from the LFS, including the major challenges. This will then be followed by a description of the methodology used to analyse the differences between the pre-revised and revised headline indicators. The results of the analysis are presented followed by discussion of these findings and conclusions.

¹ <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R1700>

2. Weighting methodology

2.1 Methodological challenges

The changing demographic composition of Malta has impacted and is expected to continue to impact the LFS results, particularly since the number of foreign residents in Malta has increased significantly over the last ten years. According to the 2021 Census of Population and Housing, foreign residents accounted for one fifth of Malta's total population, an increase of almost 20 percentage points foreign residents when compared to 2012 levels. Due to this increase in the share of foreigners in the total resident population, it is assumed that there is a larger proportion of foreigners in survey net samples.

Lower response rates among foreigners pose a significant challenge to data collection for all social surveys, including the LFS. Several factors may influence response rates among this population subgroup. Compared to Maltese residents, foreign residents in Malta are typically young adults (25 to 44 years old) and have a higher likelihood of being part of the labour market as evidenced by higher total employment rates among non-nationals as compared to nationals when estimated from the LFS. Apart from this, young adults may be engaged activities outside the home, such as participation in education and work. All these factors may reduce the chance of interviewers successfully contacting respondents in this population subgroup. Surveys conducted with non-nationals also face added challenges of language barriers and culture differences which hinder participation among this population group.

Since the current mode of data collection is CATI, it is heavily reliant on the availability of updated contact numbers in the sampling frame. Despite all efforts by NSO to obtain and maintain up-to-date contact information, reaching foreign residents remains challenging. When compared to Maltese nationals, non-nationals are more likely than Maltese nationals to change their contact numbers or place of residence therefore updating of contact information for this subgroup is especially challenging.

Increasing non-response remains a concern more so in a panel survey like the LFS as it directly impacts the number of respondents in subsequent panels and consequently the precision of estimates. Selective non-response, especially among foreign residents, may also generate bias in estimates when extrapolated to the target population. This, coupled with a larger grossing up factor, is contributing to increased variability in weights and hence any minor change among foreign residents who are part of the LFS sample is more likely to have a larger impact on the global estimates generated from the survey.

2.2 Current weighting methodology

In 2017, the NSO observed differences between LFS estimates and employment figures published by Jobsplus, the Public Employment Service (PES) agency. The divergence in the two sources was attributed to the large influx of non-nationals into the Maltese economy from 2014. This sudden increase in foreign nationals led to under-coverage of foreign residents in the sampling frame which, coupled with a higher rate of unit non-response among foreigners and under estimation in the population benchmarks used for weighting and calibration, impacted the LFS estimates. To address

these methodological shortcomings, the LFS weighting methodology was revisited with a view to adjust for the impact of sampling under-coverage and non-response by increasing the benchmarks at calibration stage and ensuring that the final estimates aligned more accurately with the target population. This revised weighting methodology has been used since 2017. Further information can be found in the published News Release available on https://nso.gov.mt/wp-content/uploads/News2018_153.pdf. The weighting methodology is carried out in two stages as described below:

Stage 1: Design weights

In the first wave, the design weights are computed based on the probability of selection of the household from the sampling frame. These are estimated using a post stratification weighting procedure based on district (LAU 1) and household size. For consecutive waves, the final cross-sectional weights of the preceding waves are taken as design weights based on the probability of selection.

Stage 2: Calibration

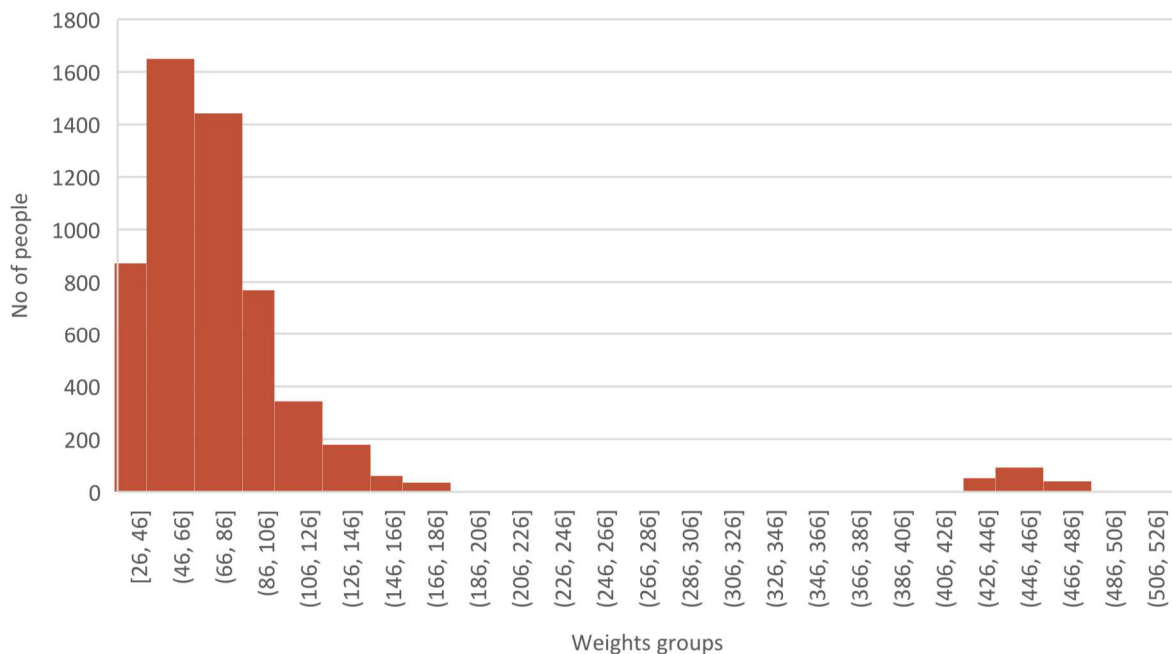
Calibration is a unified approach to calibrate survey estimates against auxiliary variables with the aim of reducing both the sampling error and non-response bias in a survey [1]. Calibration of the LFS is implemented using R-package ‘Sampling’ and the ‘calib’ function applying the logit method. Calibration is conducted on individuals and is based on the total population of persons living in private households. The following benchmarks are used for calibration:

- Geographical distribution (LAU 1)
- The 4 rotation panels
- Sex and 5-year age-groups (0-14,15-19,20-24...,75+)
- Citizenship (nationals, non-nationals and registered employed non-nationals)
- Number of households represented by the number of reference persons

Trimming is also applied to the final survey weights to minimise outliers. This is deemed necessary as large survey weights can lead to large sampling errors which in turn may produce fluctuations in survey estimates. While trimming survey weights may introduce a small bias into the estimates, it greatly reduces standard errors [2].

One must note that while the updated weighting methodology addressed the gaps in the estimates described previously, it led to an increase in the variability of the weights, which in turn increased the variance of the final estimates. Chart 1 presents the weights based on the post-Census household population estimates for Q4 2021. While most of the weights stood between 26 and 206, a minority of weights (particularly those for non-nationals) were between 406 and 526.

Chart 1: Histogram of quarterly weights using post-Census household population estimates: Q4 2021



3. Impact of reweighting based on post-Census demographic revisions

3.1 Preliminary results

To understand the potential impact of the post-Census demographic revisions on the LFS estimates, a preliminary analysis of the quarterly household population estimates was conducted to compare the revised population series to the pre-revision data. This reweighting exercise follows the publication of the 2021 Census of Population and Housing, which revealed an element of under-estimation in the foreign resident population which did not result in a significant change in the total population.

The analysis focused on the core employment and unemployment statistics, which are regarded as the LFS's headline indicators on the labour market. The criteria adopted in the LFS are consistent with international methodological guidelines set by the International Labour Organisation (ILO), allowing results to be compared to those from other EU Member States and countries. Based on these criteria, the LFS measures employment and unemployment according the following definitions:

Employment: All persons aged 15 to 89 who, during the reference week, reported to be in one of the following categories:

- working for at least 1 hour for pay or profit, including contributing family workers and paid casual work
- having a job or business but were temporarily not at work during the reference week but had an attachment to their job. This category includes:

- (a) persons not at work due to holidays, working time arrangements, sick leave, maternity or paternity leave;
- (b) persons in job-related training or formal education;
- (c) persons on parental leave, either receiving and/or being entitled to job-related income or benefits, or whose parental leave is expected to be 3 months or less;
- (d) seasonal workers during the off-season, where they continue to regularly perform tasks and duties for the job or business, excluding fulfilment of legal or administrative obligations;
- (e) persons temporarily not at work for other reasons where the expected duration of the absence is 3 months or less.

Employment rate: persons in employment (15-64 years) as a percentage of the population of working age (15-64 years).

Unemployment: All persons above 15 years of age who, during the reference week, reported to be:

- without work, and
- actively seeking work during the previous 4 weeks: i.e. had either carried out activities in the four-week period ending with the reference week to seek paid employment or self-employment or found a job to start within a period of at most 3 months from the end of the reference week. Examples of active job search include contacting JobsPlus, studying job advertisements or placing or updating CVs online, and
- currently available for work: i.e. available for paid employment or self-employment before the end of the 2 weeks following the reference week.

Unemployment rate: Unemployed persons (15 to 74 years) as a percentage of the labour force (15-74 years).

Estimates and the headline indicators derived from this survey were based on the revised household population estimates. Charts 2 and 3 below show the impact of the updated post-Census weights on the labour status indicators when compared to the pre revision estimates.

Chart 2: Total employment (15 years +) Q1 2012 – Q3 2023

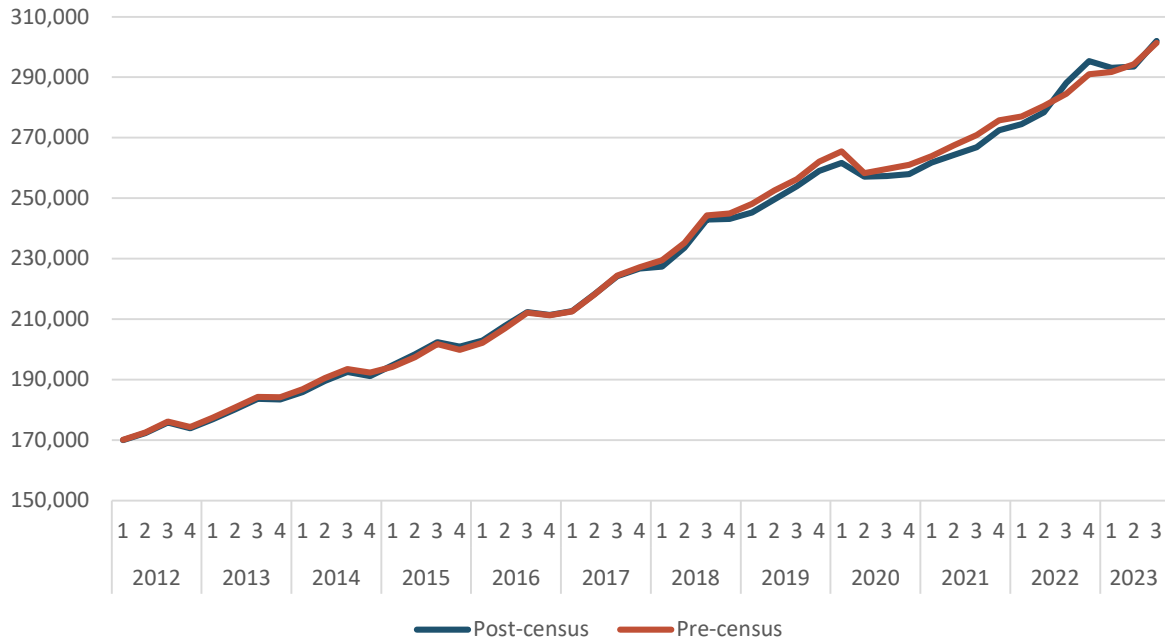
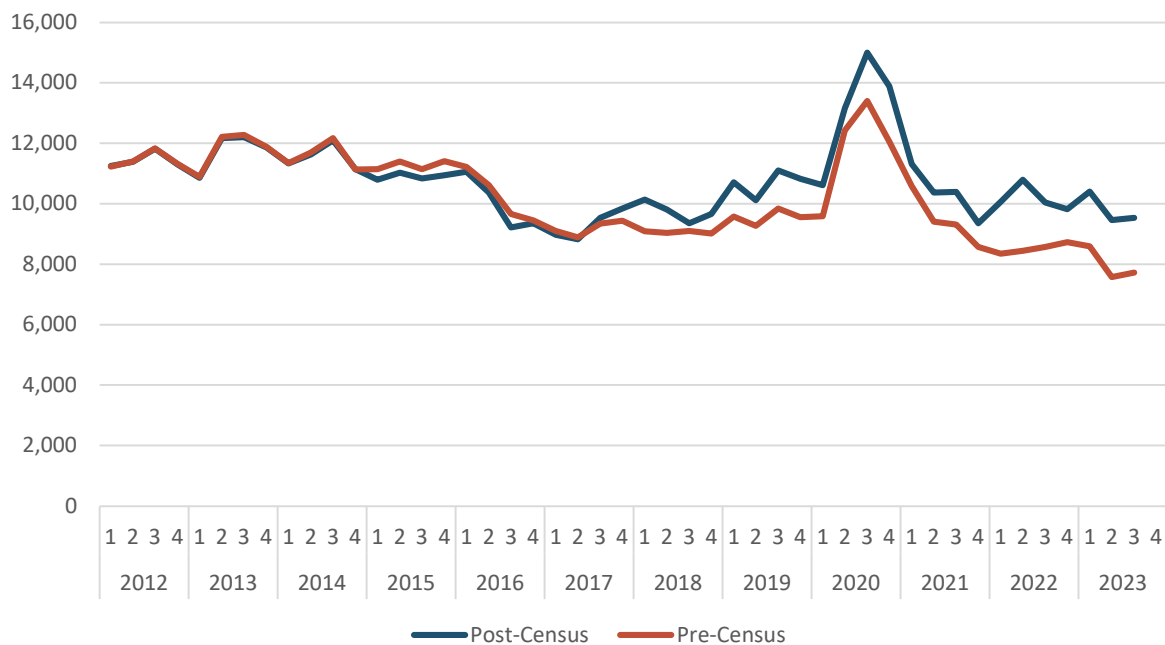


Chart 3: Total unemployment (15 years+) Q1 2012 – Q3 2023



Using the updated weights, estimates of total employment were slightly lower when compared to those estimated using pre-Census weights (Chart 2). On the other hand, Chart 3 shows an upward shift in unemployment from Q1 2018 onwards.

3.2 Comparative analysis of revised series

3.2.1 Methods – Design based approach

Charts 2 and 3 showed that the impact of the revised household population series on the LFS headline indicators became more pronounced from Q1 2018. To better understand the differences in this period, the variance and margin of error were calculated for the revised total employment and unemployment rates. Visual inspection of the rates before and after the revision was conducted to compare the point estimates and trend when considering the margin of error. The ultimate cluster method was used to calculate the variance which was conducted in R using the ‘Vardpoor’ package [3].

A design-based inference was used to investigate the difference between the parameter estimates using the pre-Census and post-Census population levels [4]. Prior to conducting the analysis, the time series of the difference between the two parameter estimates was tested for normality using the Shapiro-Wilk test prior analysis [2]. Because of the divergence from normality, the one-sample Wilcoxon signed-rank test, the non-parametric equivalent to the one-sample paired t-test, was used. The null and alternative hypothesis are described below:

H_0 : Median differences between the point estimates pre and post revision = 0

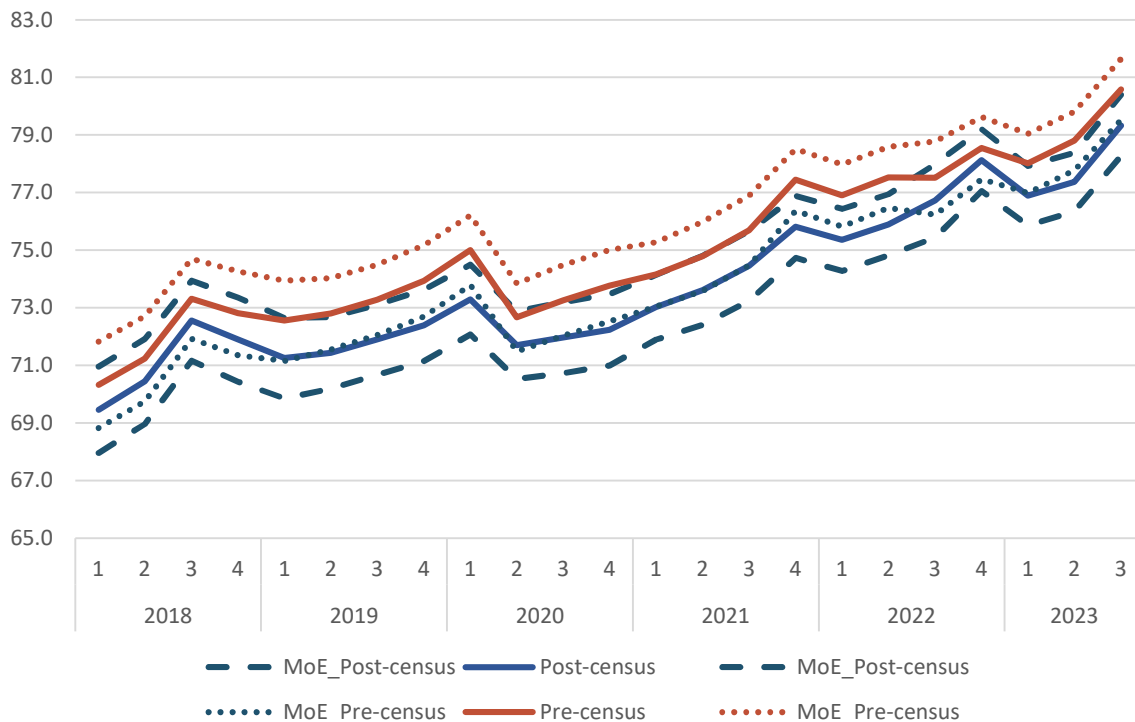
H_1 : Median differences between the point estimates pre and post revision \neq 0

To allow for comparison between the periods, analysis was conducted on estimates for the whole series, that is, 2012 to 2023, and also for the periods 2012 – 2017 and 2018 – 2023 separately.

3.2.2 Results

For the employment rate (Charts 2), from 2018 onwards, the revised employment rate appears to shift significantly from its previous trajectory, while keeping the same quarter on quarter trend. For some of the point estimates from 2019 onwards, the employment rate based on the revised weights falls outside the confidence interval of the employment rate based on the pre-revised weights (Chart 4). However, the significant difference is unclear, as the confidence intervals overlap; but sample estimates using pre-Census weights were not contained in the post-Census estimates confidence interval. As a result, further tests were necessary [5].

Chart 4: Employment rate and margin of error (MoE) pre and post revisions



The hypothesis testing to compare the difference between the pre-Census and post-Census estimates for the employment rate show that there was a statistically significant difference from 2012 to 2023 ($p < 0.05$). However, on dissecting the time series and analysing further the period 2012 to 2017, the hypothesis testing showed that the difference between pre and post-Census estimates for employment rates was not statistically significant ($p = 0.318$). This implies that the median difference in the estimates pre and post revision for this period did not differ significantly from zero and therefore no revision was required. However, for the period Q1 2018 to Q3 2023, the difference between the pre and post-Census for the total employment rate did differ significantly from zero ($p < 0.05$) (Table 2). This implies that from Q1 2018 the rates post revision did differ from the rates estimated from the LFS prior to the revision, hence indicating a level shift. This finding gives emphasis to the importance of updating employment LFS figures from Q1 2018 onwards to better reflect this level shift and the current demographic changes.

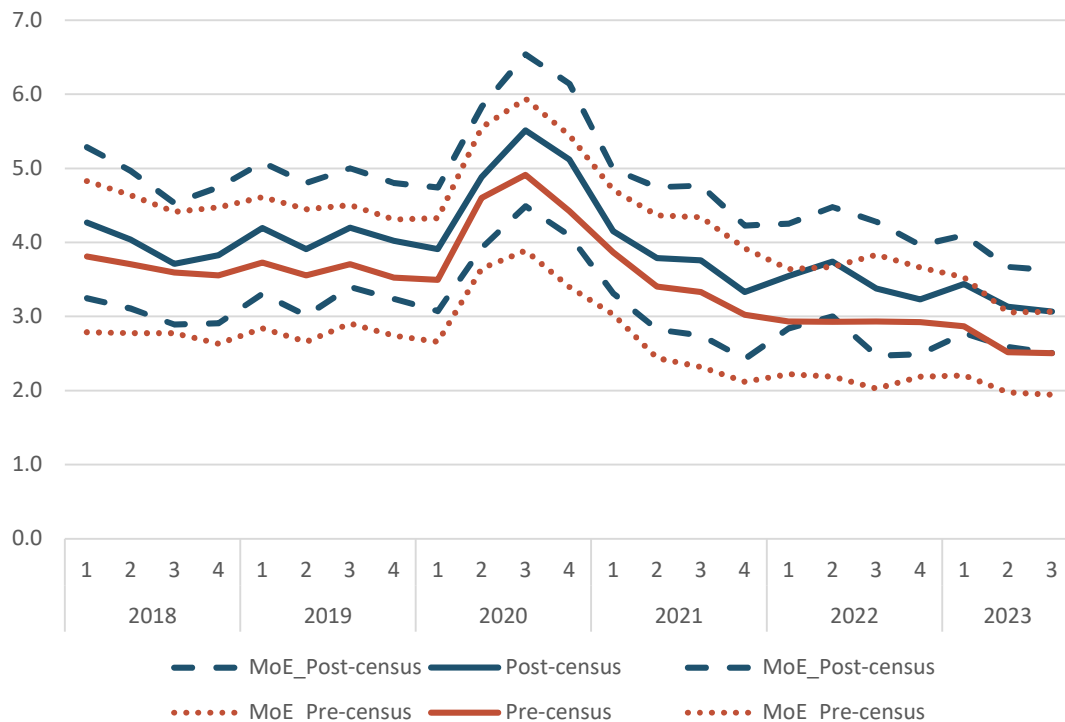
Table 2. One Sample Wilcoxon Signed Rank test comparing pre and post revisions for difference in employment rate

One-sample Wilcoxon Signed Rank Test

	Employment Rate		
	2012-2023	2012-2017	2018-2023
Total N	47	24	23
Test Statistic	661.000	42.500	276.000
Standard Error	68.907	17.514	32.839
Standardised Test Statistic	4.216	-0.999	4.202
Asymptotic Sig. (2-sided test)	0.000	0.318	0.000

Similar examination of the unemployment rate using the pre and post-Census weights was conducted. Chart 5 shows that all the data points are within the margin of error of the unemployment rate estimated with the pre-Census weights. This implies that despite the recalibration, from 2018, the overall trends in the unemployment rate have remained consistent (Chart 5).

Chart 5: Unemployment rate and margin of error (MoE) pre and post revisions



As was seen for the total employment rate, the hypothesis testing showed no significant difference in the unemployment rate for the period 2012 – 2017 ($p=0.112$) but a significant difference between unemployment rates for the 2018-2023 period ($p<0.05$) (Table 3). This further accentuates the level shift outcome for the time span Q1 2018 to Q3 2023, but not for the time span 2012 - 2017 (Table 3). This was even though, upon visual inspection, the revised point estimates in the period appeared to fall within the margin of errors of the pre-revised estimates. Therefore, as with the total employment rate, it was deemed necessary to revise the estimates for the unemployment rate from Q1 2018 onwards.

Table 3. One Sample Wilcoxon Signed Rank test comparing pre and post revisions for difference in unemployment rate

One-sample Wilcoxon Signed Rank Test

	Unemployment Rate		
	2012-2023	2012-2017	2018-2023
Total N	47	24	23
Test Statistic	564.000	19.500	276.000
Standard Error	60.805	12.278	32.684
Standardised Test Statistic	4.095	-1.588	4.222
Asymptotic Sig. (2-sided test)	0.000	0.112	<0.001

4. Conclusion

Disruptions in data continuity are plausible when population figures are revised if these are being used to calibrate survey estimates. Should a revision of a time series take place, data would be aligned to the new population levels and estimates would be break free. This approach is obviously ideal when dealing with labour market related indicators. The detection of structural breaks on the other hand ensures that policy recommendations are on target and result in more accurate forecasts.

The analysis presented here assessed the impact of reweighting the LFS employment and unemployment time series using revised post-Census household population estimates. Preliminary visual inspection of the rates indicated that the revised point estimates for unemployment were within the margin of error of the pre-revised estimates, while some of the revised point estimates for employment were not. Hypothesis testing showed that despite this, both indicators exhibited a difference between the pre and post revision estimates that was significantly different from zero for the period 2018 to 2023. No significant difference was noted for the time 2012 to 2017 for both labour status indicators. In view of these outcomes, LFS data will be revised from Q1 2018 until Q3 2023 to ensure a break free series for all LFS estimates.

References

- [1] Lundström, S. and Särndal, C. E. (1999). Calibration as a Standard Method for Treatment of Nonresponse. *Journal of Official Statistics*, 15: 305–327.
- [2] Cochran, W. (1977). *Sampling Techniques*, (3rd edition). New York: Wiley.
- [3] Hansen, M. H., Hurwitz, W. N., & Madow, W. G. (1953). *Sample survey methods and theory* (Vol. I). New-York: Wiley.
- [4] Van den Brakel, J., & Renssen, R. (1998). Design and Analysis of Experiments Embedded in Sample Surveys. *Journal of Official Statistics*, 14(3), 277-295.
- [5] Tools of the Trade: Sample Estimates compared using Confidence Intervals
https://www.health.pa.gov/topics/HealthStatistics/Statistical-Resources/UnderstandingHealthStats/Documents/Sample_Estimates_Compared_Using_Confidence_Intervals.pdf