



May 2018

## The sudden jump in the unemployment rate in 2015: Is there a break in the QLFS data?

*Jairo Arrow, retired statistician, Lüneburg, Germany*

*This article argues that there is a time-series break in StatsSA's Quarterly Labour Force Survey (QLFS) in the first quarter of 2015. The principal cause appears to be differences between the sampling frames before and after 2015, respectively based on Census 2001 and Census 2011. The break is signalled by a significant increase in the number of one-person households surveyed in the QLFS. It is critical for economists, policy analysts and policy-makers to be aware of this break in the data.*

### Introduction

As a methodologist of long experience with the statistical outputs of Statistics South Africa, I have retained a friendly interest in these figures, particularly in developments in its sample surveys. I was struck by the sharp spike in the unemployment rate from the Quarterly Labour Force Survey (QLFS) between the December quarter of 2014 and the March quarter of 2015, and by the greater volatility of the quarterly series thereafter. Seeking an explanation for the spike, I learnt that it coincided with the selection of a new QLFS sample from a redesigned 'master sample' (or sampling frame<sup>1</sup>) that was based on the 2011 Census. On the day the results of the March quarter of 2015 were released, the Statistician General, Mr PJ Lehohla, issued a 'Quality statement of QLFS'. In this he states:

---

<sup>1</sup> A sampling frame is a list of all units within a population that can be sampled, and may include individuals, households or institutions. In a statistical survey such as the QLFS, a sample of individuals is then drawn from this sampling frame – also called a 'master sample'. Further details on the design of the master sample can be found in a StatsSA technical report, *Quarterly Labour Force Survey Guide*, 2008.

The Quarterly Labour Force Survey (QLFS) of Q1:2015, using a redesigned master sample based on the 2011 Census, showed a large quarterly change in the number of unemployed persons (626 000) to approximately 5,5 million. This was accompanied by a decline of a similar magnitude in the other Not economically active [sic] (611 000), which resulted in an unemployment rate of 26.4%, 2.0 percentage points higher compared to the Q4: 2014 unemployment rate. ...

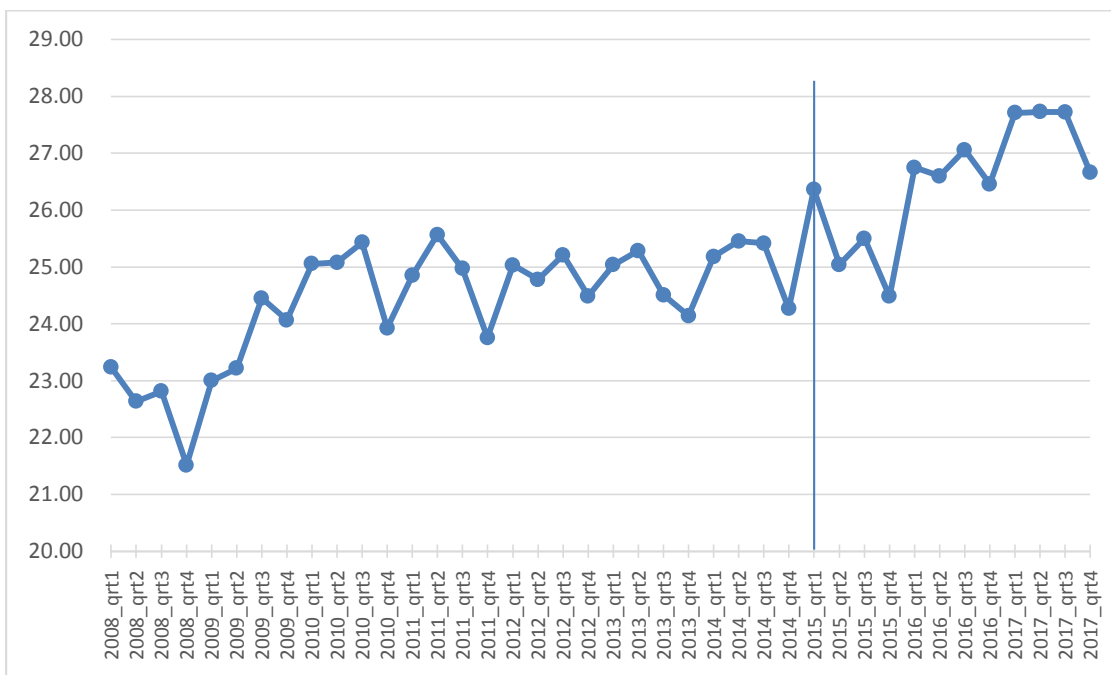
StatsSA initiated an evaluation project which included a parallel study in September-October 2015. ... Following this data confrontation, StatsSA is confident that the results produced during 2015 is accurate and that all quarters of 2015 are comparable. However, to better understand the dynamics of the Q4:2014 and Q1:2015 change another point of Q1:2016 is required. Such a comparison will provide the basis for determining whether a rebasing of the QLFS series will be required. A technical report, detailing the findings will be made available at the time of the QLFS Q1: 2016 release.

To my knowledge the promised technical report has not yet been released.

### Does the QLFS correctly reflect the unemployment trajectory?

Unemployment in South Africa has been on the increase since 2008, according to Stats SA. Data from the QLFS show that, while in 2008 the national unemployment rate stood on average at 22.5%, it had risen to 27.5% by 2017 – an increase of about five percentage points (see Figure 1). Between 2010 and 2014 the level of unemployment remained roughly constant at approximately 25%. However, significant increases, even jumps, started to be discernible from 2015 onwards.

**Figure 1. South African unemployment rate: QLFS 2008–2017**



Source: QLFS (all figures in this article)

For economists, policy analysts and policy-makers, it is critical to understand this sudden increase. In particular, does it reflect a fundamental, even structural change in the labour market? What would such a change in the structure of employment imply for fiscal, monetary and labour-market policy?

Not so fast, I would warn. Any sudden or unexplained jump in an important statistical number needs to be carefully scrutinised from a technical and statistical point of view before one rushes to interpret it as a substantive change in economic behaviour or structure. This may be the case here, as the analysis below suggests, using only StatsSA data.<sup>2</sup>

### **A change in the master sample**

The pre-2015 household surveys used a master sample, or sampling frame, that was designed using the results of the 2001 population and household census, while the post-2014 surveys use a redesigned sampling frame that is based on the results of the 2011 census. This is well-known amongst statisticians and StatsSA has openly made this known to users.

Indeed, it is standard practice amongst national statistical agencies in the world to periodically redesign the sampling frame when new data become available. This approach could result in a revision of certain statistics. The statistical agency would then clearly spell out why a revision was necessary, how it was implemented and any possible impact on related statistical results. Non-statisticians would then be assisted how to revise their time-series, while advanced users could make their own decisions but based on the new data.

Any issues with the census data that are used to generate the master sample would affect associated statistics such as the QLFS. There is, therefore, a need for further investigation on the cause of the break in the series.

I believe it is likely that a time-series break did in fact occur in the QLFS data with the introduction of the new master sample from the first quarter of 2015 as a direct consequence of the data used from 2011 census. A strong indication of such a break is found in (a) changes in the average size of households surveyed in the QLFS (and other StatsSA surveys) from 2015 onwards (see next section and Figure 2). In addition, there has been (b) a significant drop in the number of

---

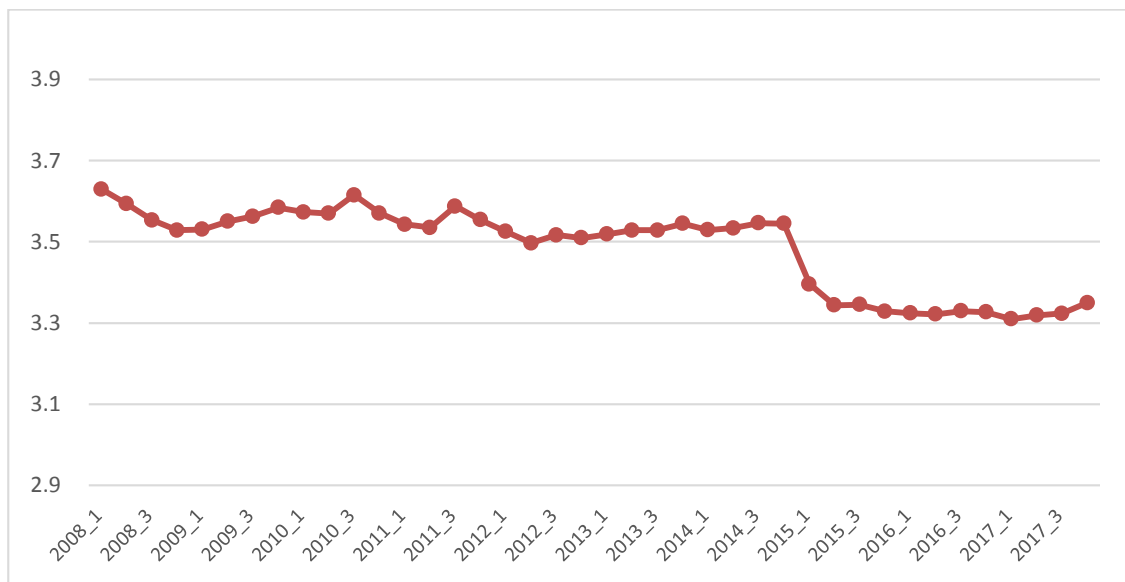
<sup>2</sup> In the following analysis I have used only Stats SA data which are publicly available on its website <http://interactive.statssa.gov.za:8282/webview/>. I have cross-checked my results with the published ones and they match, showing no difference on methods and sources.

households surveyed (Figure 3) and (c) a notable jump in the number of single-person households (Figure 4).

### Changes in the size of households surveyed

The average size of households is an important indicator of structural changes in the society. Typically, it changes gradually over time. But the results of the QLFS seem to indicate a sudden drop in the size of the average household, occurring at the same point in time in the first quarter of 2015 (as is apparent in Figure 2).

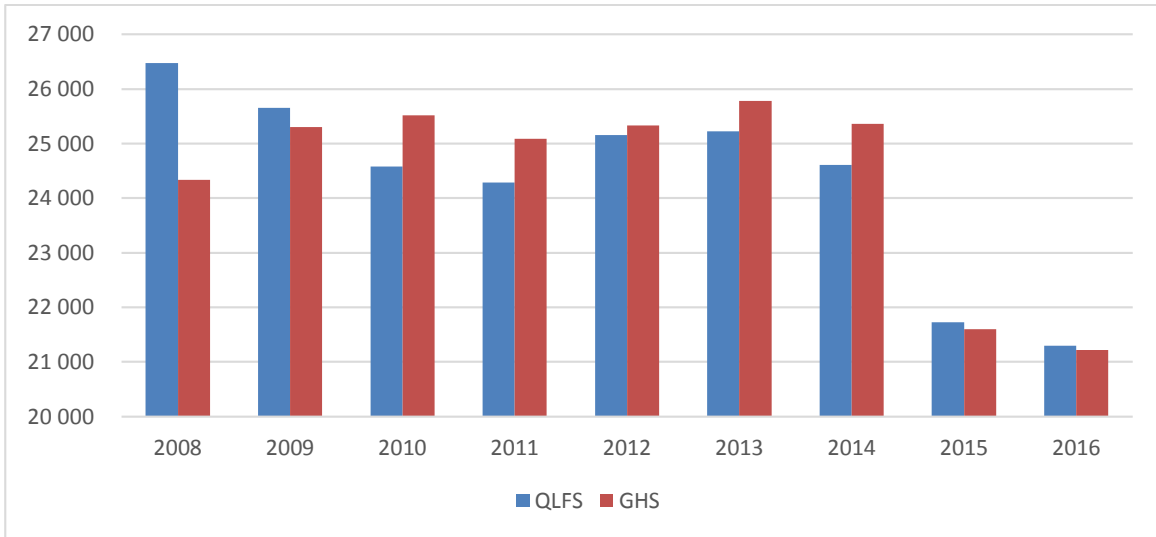
**Figure 2: Average household size in the QLFS: 2008–2017**



### Comparing the jump in the QLFS with the results in other surveys of StatsSA

I have tried to relate some of these findings to corresponding ones in another major household survey – the annual General Household Survey (GHS), which uses the same master sample as the QLFS. Figure 3 shows the number of households enumerated in each of the two surveys.

**Figure 3. Number of households sampled in the QLFS and GHS**



We see again that the change at the introduction of the new master sample in 2015. (To aid comparability, I have averaged the numbers of households in the QLFS over the four quarters of each year.) So, whatever seems to have caused a break in the QLFS series applies equally to the GHS as well.

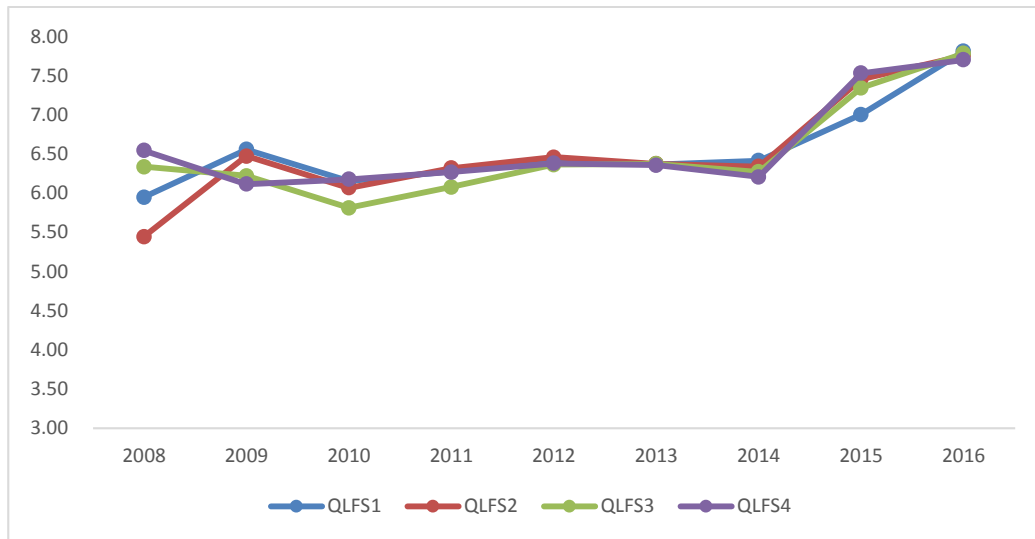
### Can one explain what happened?

I shared my findings with Mr Hussain Choudhry, the methodologist from Statistics Canada who designed both master samples for StatsSA. He commented:

... I find it interesting that the number of persons processed, and the average household size started the decline at the same time. It would seem to me that there is serious person nonresponse in the QLFS. In other words, all persons in the responding households do not get enumerated. The result would be smaller average household size, i.e. the estimated number of households by size would be biased [i.e. there would be an overestimation of the number of smaller-size households and an underestimation of the number of larger-size households].

To test Choudhry's observation, I plotted the percentage of persons reported as living in single person households for each year 2008 to 2016. The results are shown in Figure 4.

**Figure 4: Percentage of persons living in single households according to QLFS: 2008–2016**



As Choudhry surmised, in each quarter’s QLFS the recorded number of single-person households jumped in 2015 – suggesting that in many households not all members were recorded by the survey enumerators, thus resulting in the recording of many of these as ‘single-person’ households.

### Concluding remarks

It is evident that there may be an inconsistency between the QLFS results before 2015 and after 2015 – a break in the data series. This break coincides with the introduction of a new master sample for household surveys at Stats SA. From at least 2009 to 2014, the quarterly movements in the main labour force variables, particularly unemployment, appear plausible. However, the shift to the new master sample at the start of 2015 appears to coincide with a structural discontinuity in the level of the series, as reflected in a jump in the unemployment rate and other variables.

What is the possible impact of the jump in the number of recorded single-person households in the QLFS sample from 2015 on the estimated rate of unemployment? It appears that the unemployment rate for those persons is lower than for persons in multi-person households. For instance, in the first quarter of 2015 the rate for single-person households was 15.8% compared to 26.4% in multi-person households (own calculations from StatsSA data). Therefore, the effect of an increase in the number of single-person households has likely been to *understate* the increase of the unemployment rate in the post-2014 QLFS data series.

A tentative conclusion regarding the apparent break in the QLFS data at the start of 2015 is that there has been a change in the master sample which biases it towards smaller households, thus inducing a downward bias in the rate of unemployment. It is essential that analysts are acutely aware of such a break and its implications – and explicitly take it into account.

\* \* \*