2022

IARIW-TNBS Conference on "Measuring Income, Wealth and Well-being in Africa"

Paper Prepared for the IARIW-TNBS Conference, Arusha Tanzania April 14-15, 2022

Measuring Earnings Inequality in South Africa Using Household Survey and Administrative Tax Microdata

Andrew Kerr (University of Capetown, South Africa)

In this paper we use household survey and administrative tax microdata to examine the evolution of earnings inequality in South Africa over the period 1993-2017. South Africa is a country with extremely high levels of income inequality. Leibbrandt et al. (2010) used a variance decomposition to show that 85% of overall income inequality is caused by earnings inequality in the labour market, and that of this, one third is due to the large number of those not working, and two thirds is due to earnings differences between those in employment. Wittenberg (2017a) has investigated changes in earnings inequality over the post-Apartheid period, finding that, as measured by the Gini coefficient, earnings inequality increased in the 1990s and stabilised at a very high level from 2000 to 2011.

Research on earnings inequality in South Africa has focused mainly on household survey data, which has become ubiquitous since the 1993 Project for Statistics on Living standards and Development (PSLSD) conducted by the South African Labour and Development Research Unit (SALDRU) and the public release of household survey microdata from surveys conducted by Statistics South Africa (Stats SA). There has been no substantive description, however, of individual earnings inequality using the employee tax certificate (IRP5) data.

In this paper we undertake this work, as well as similar descriptions of earnings inequality using the Quarterly Labour Force Survey (QLFS) and the General Household Surveys (GHS), conducted by Statistics South Africa. We compare these two sets of estimates and use these to shed light on how inequality has changed over the post-Apartheid period. We do so using three main descriptive tools: Gini coefficients, the variance of log earnings and various percentile ratios. Wittenberg (2017a, 2017b) used percentile ratios to show that between 1994 and 2011 earnings inequality had decreased in the bottom half of the earnings distribution but increased in the top half. This motivates the continued use of percentile ratios since differing patterns in the top and bottom halves of the earnings distribution would not show up in the Gini or variance of log earnings.

The quality of the QLFS earnings data and the imputations performed by Statistics South Africa have been questioned by Kerr and Wittenberg (2019, 2021). One of the worrying aspects of the QLFS is the extreme changes in the Gini coefficient of earnings over a very short space of time.

We thus spend some time investigating the extent to which the inequality trends from the QLFS data are reliable.

The degree to which the quality of household survey data influences the conclusions about the extent of, changes in, and causes of inequality is important even in rich countries. For example, Hirsch and Schumacher (2004) documented that around 30% of the employed in the US Current Population Surveys (CPS) they investigated from 1973-2001 had imputed earnings. Subsequently, Lemieux (2006) and Autor and Acemoglu (2011) highlighted that in some of the CPS, earnings are imputed, and that in others they are not, making comparisons over time difficult. As a result, Lemieux (2006) elected to include only earners who answered the earnings questions, in order to exclude imputed values. He was able to do this because most CPSs include flags showing that an earnings value was imputed.

Our analysis shows that the earnings imputations undertaken in the QLFS surveys make the earnings data unreliable for estimating changes in earnings inequality. We can make this conclusion because the National Statistics Office conducts another household survey, the General Household Survey, which asks almost exactly the same earnings question, has almost exactly the same sample design and samples from the same primary sampling units (which are enumerator areas from the previous population census), but there is no earnings imputation. Despite the similarities this survey shows none of the oddities in earnings inequality trends since 2010, which we conclude is due to earnings imputation.

No work on earnings inequality has focused on the role of firms in generating inequality in earnings in South Africa thus far. An important question is whether inequality in earnings is the result of large average differences in earnings between firms, so that which firm a worker works for is very important. An alternative is that within all firms there is a high degree of inequality between well and poorly paid workers. Or there could be a situation in between. In this paper we provide a first look at the relative importance of within and between-firm inequality in contributing to the extremely high levels of inequality in South Africa, using matched firm and worker data provided by the South African Revenue Service (SARS), following methods suggested by Song et al. (2019).