The COVID-19 pandemic and polarisation of income distribution in South Africa

## **Supplementary file**

# Adeola Oyenubi

University of the Witwatersrand, South Africa

## **Contents**

|  |  |
| --- | --- |
|  | Page |
| Section S1 Supplementary tables  | 2 |
| Table S1 Summary statistics | 2 |
| Table S2 Polarisation indices(for respondents that reported monthly income) | 2 |
| Section S2 Technical definitions  | 3 |
| S2-1 The relative distribution | 3 |
| S2-2 Blinder-Oaxaca type decomposition of location and shape differences | 4 |

# Section S1 Supplementary tables

## **Table S1: Summary Statistics**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Mean | Std. Dev. | Min | Max |
| Income |  |  |  |  |
| February | 5747.03 | 8878.24 | 0 | 100000 |
| April | 4725.86 | 8530.03 | 0 | 100000 |
| Geo location |  |  |  |
| Trad | 0.15 | 0.36 | 0 | 1 |
| Urban | 0.81 | 0.39 | 0 | 1 |
| Farms | 0.04 | 0.20 | 0 | 1 |
| Employed |  |  |  |
| in April | 0.71 | 0.46 | 0 | 1 |
| in February | 0.09 | 0.29 | 0 | 1 |
| Tertiary Education | 0.40 | 0.49 | 0 | 1 |
| Self-reported poor health | 0.26 | 0.44 | 0 | 1 |
| Age in yrs. | 38.68 | 10.86 | 17 | 89 |
| African | 0.82 | 0.39 | 0 | 1 |
| Regular job C | 0.46 | 0.50 | 0 | 1 |
| Employment Contract A | 0.42 | 0.49 | 0 | 1 |
| Descent Job B | 0.09 | 0.28 | 0 | 1 |
| obs | 3047 |

A Employee has an employment contract

B This is a dummy that is 1 if the respondent is a Manager or professional and zero otherwise

C Omitted category includes casual work, self-employment and running a business

## **Table S2: Polarisation Indices (for respondents that reported monthly income)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Coef. | JackknifeStd. Err. | t | P>t | [95% Conf. | Interval] |
| MRP | 0.11 | 0.04 | 2.44 | 0.015 | 0.02 | 0.19 |
| LRP | 0.22 | 0.10 | 2.33 | 0.02 | 0.04 | 0.41 |
| URP | -0.01 | 0.03 | -0.54 | 0.592 | -0.06 | 0.04 |

## **Section S2 technical definitions**

## S2-1 The relative distribution

Let and represent the income distribution of the reference (in our case February) and comparison (in our case April) distributions. The relative distribution of to is defined as the distribution of a random variable

which is obtained from when transforming it with the cumulative distribution function of , . is a continuous random variable over the range , the realisations of , i.e., are referred to as ‘relative data’. Intuitively, can be interpreted as the set of positions that observations of the comparison population would have if they were located in the distribution of the reference population. The probability density function of can be obtained as a ratio of comparison density to the reference density at the relative data . The probability density function can be written as

 (1)

where and are the density functions of and respectively, and is the quantile function of . The relative density describes where individuals at various quantiles in the comparison (April) distribution are concentrated in terms of the quantiles of the reference (February) distribution. When the relative density is 1, this means that the two distributions have a similar density at the quantile of the reference population. In this case, has a uniform distribution. A relative density greater than 1 implies that the comparison distribution has more density than the reference at the quantile of the reference distribution. Similarly, a relative density that the less than 1 indicates the opposite.

The major advantage of the relative density method is that it allows for the decomposition of the relative distribution into changes in location (measured by the median in this study) of the densities being compared, and changes in shape (including differences in variance, asymmetry, and other distributional characteristics) that could be linked to factors such as polarisation. The decomposition can be written as

 (2)

where is the reference density function adjusted by an additive shift. This density has the shape of the reference distribution but the median of the comparison distribution. The adjustment is implemented with , which represents the difference between the medians of and . The first term of Equation 2, therefore, represents the location component, while the second term represents the shape component, i.e., the relative density net of the location effect. The shape component is useful for isolating movements (re-distribution) occurring between the reference and the comparison distributions. Thus, it is possible to determine whether there is polarisation of the income distribution (increases in mass at both tails), ‘downgrading’ (increases in the lower tail), ‘upgrading’ (increases in the upper tail), or convergence of income towards the median (decreases in both tails).

The information contained in the shape component of the decomposition can be quantified. The quantity is called the *median relative polarisation* (MRP) index. The index is normalised so that it varies between -1 and +1, with zero representing no shape effect (uniformly distributed shape component). Positive values represent greater polarisation (i.e., increase in the tail(s) of the comparison distribution relative to the reference distribution) and negative values represent less polarisation (i.e., convergence towards the median). The MRP can be written as

 (3)

Equation 3 shows that values further away from the median are given a greater weight, with weights increasing linearly with distance from the centre. The index can be interpreted in terms of a proportional shift of mass in the income distribution from central to less central values.

The polarisation index can be additively decomposed. This decomposition shows, if the polarisation represents movement of mass towards the lower or upper tail, the lower and upper relative polarisation indices (LRP and URP, respectively) are given by

 (4)

 (5)

with , like MRP, LRP and URP ranges from -1 to +1.

## S2-2 Blinder-Oaxaca type decomposition of location and shape differences

Clementi, Molini, and Schettino (2018) introduce a methodology for analysing the effects of covariates on the observed distributional changes due to both the location and shape shifts. Their approach integrates the relative distribution approach with the approach for decomposing wage differentials (Firpo et al., 2009), which makes use of Regression Influence Function (RIF) regressions. By this method, unconditional quantile regression is used to perform an aggregate decomposition of the location and shape differences. The decomposition separates the total difference at a quantile into a component that is due to differences in observable characteristics (endowment effect) and a component that is due to differences in returns (coefficient effect). This is implemented using the traditional Oaxaca-Blinder decomposition Blinder, 1973: Oaxaca, 1973), which can be written as

 (6)

where represent the total difference between the distributions being compared at quantile , represents the endowment effect, represents the coefficient effect and represents the effect due to interaction between and . Furthermore, once the RIF regressions of the quantile of comparison and reference distributions have been run, detailed decomposition into contributions attributable to each covariate can be performed. See Clementi, Molini, and Schettino (2018) for a more detailed presentation of this methodology.

In this paper, this approach is used to perform aggregate and detailed decomposition by gender groups for April. Specifically, this decomposition is used to identify factors that explain the differences in the location and shape differences between gender groups in the April income distribution. It is acknowledged that ideally the decomposition should be used to explain the differences across time (i.e., between February and April); however, this is not done for two reasons. First, the period between the two time points is very short (two months), so a priori one should not expect many covariates (except the ones that have to do with labour-market disengagement) to change dramatically. Second, the data do not contain separate information on covariates for the two time points. The respondents were interviewed between May and June and asked about their income and employment in February and April. In other words, we only have retrospective information about employment and income February. [[1]](#endnote-1)

Given this restriction, we instead use the decomposition methodology to explain the shape and location differences between gender groups in April. This is because existing narratives suggest that gender is an important factor when it comes to the effect of the pandemic on the labour market. Furthermore, gender is correlated with work characteristics, which are also correlated with variations in the effect of the pandemic.

1. Specifically with covariates at two time points similar analysis to the one being proposed can be done over time. This will allow the analysis to differentiate the ‘price effect’ over time from the ‘endowment effect’. This will be possible when more waves of the data become available. [↑](#endnote-ref-1)