

# Combining Data Sources to Measure Inequality in Developing Countries: The Example of South Africa

Murray Leibbrandt  
Wider Senior Research Fellow  
Director, African Centre of Excellence for Inequality Research

UNU-WIDER – DISD – DESA Workshop on Inequality

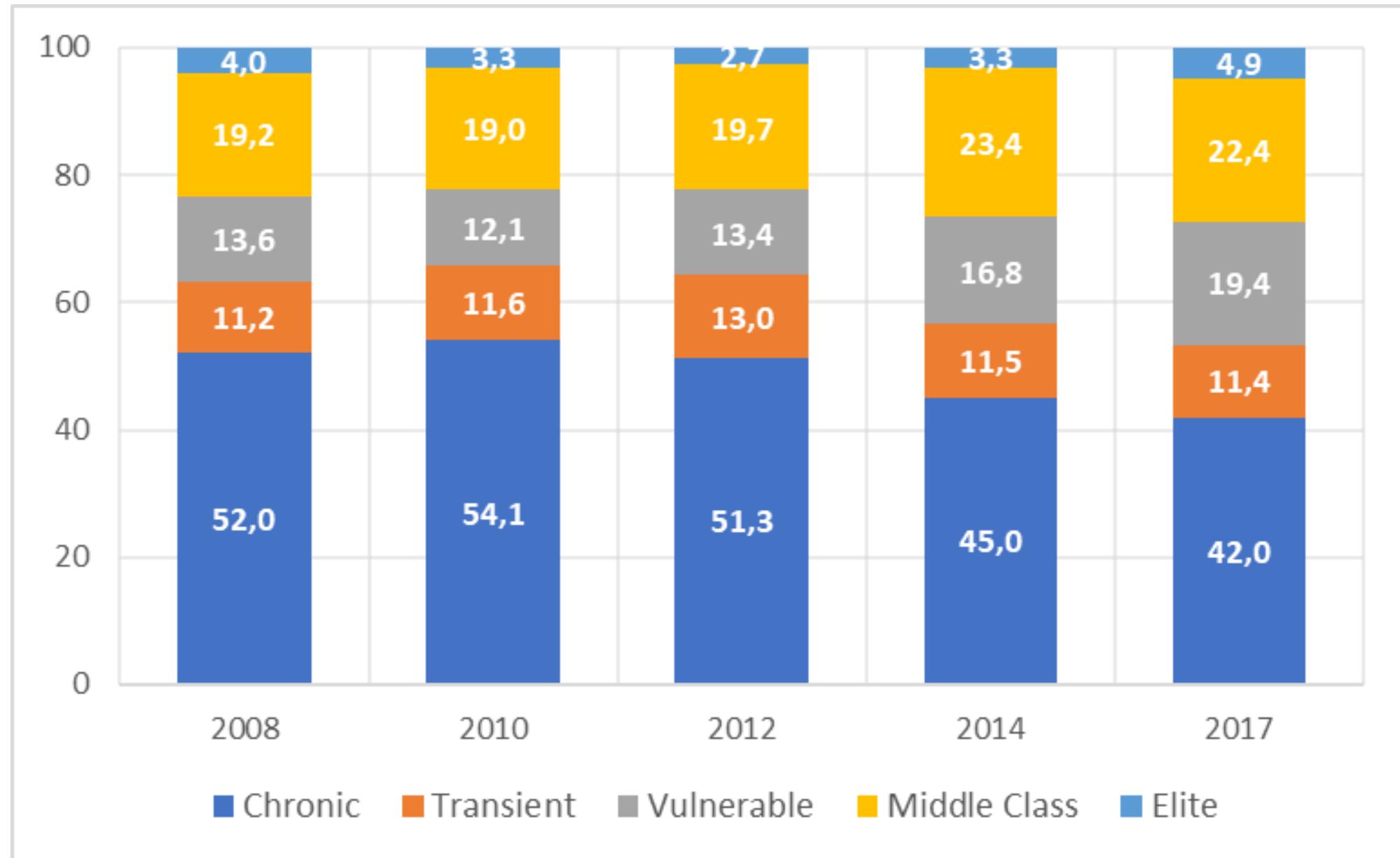
United Nations Headquarters

6 May 2019

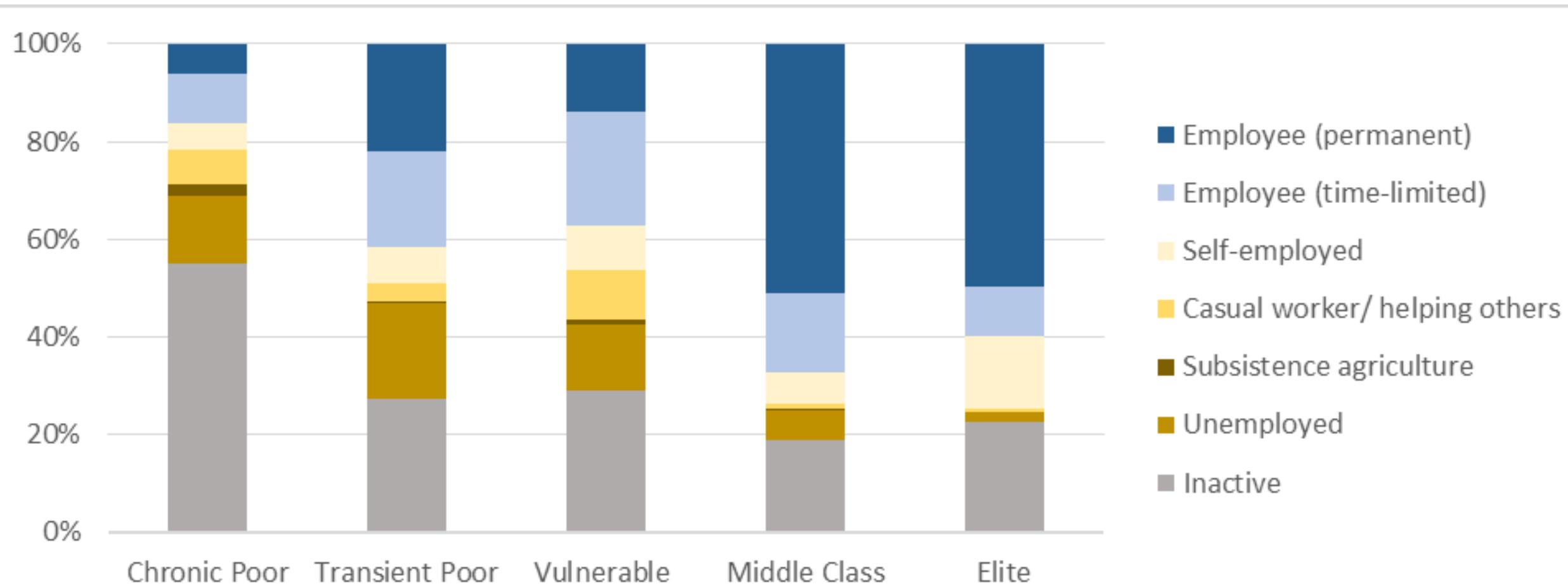
# Why is South Africa an Interesting Case Study?

- Very high inequality and a legacy that policy is trying to overcome its inequality
- An intermediate case in terms of data:
  - Good survey data
  - Fairly decent tax data and administrative data systems
  - Is integrated into the world inequality measurement exercises and policy evaluation exercises
  - But is in an intermediate position both in terms of the data and the use of the data
- Starting with the Giants Project and pushing on to current Wider projects, **this research is imbedded in a partnership between the local NSO, government ministries and researchers**

# Socio-economic class sizes, 2008 – 2017



# Economic activity of the household head in 2017



# Data Check!

Data across the distribution is crucial (Bottom, Middle and Top)  
Society as a whole produces poverty and everything else!

## Bottom

- Agric (Not SA but elsewhere),
- Informal labour market earnings, Home production (all crucial for gendered perspective)
- Different wealth assets, poverty dynamics, multidimensionality (Not a voyeuristic exercise)

## Middle

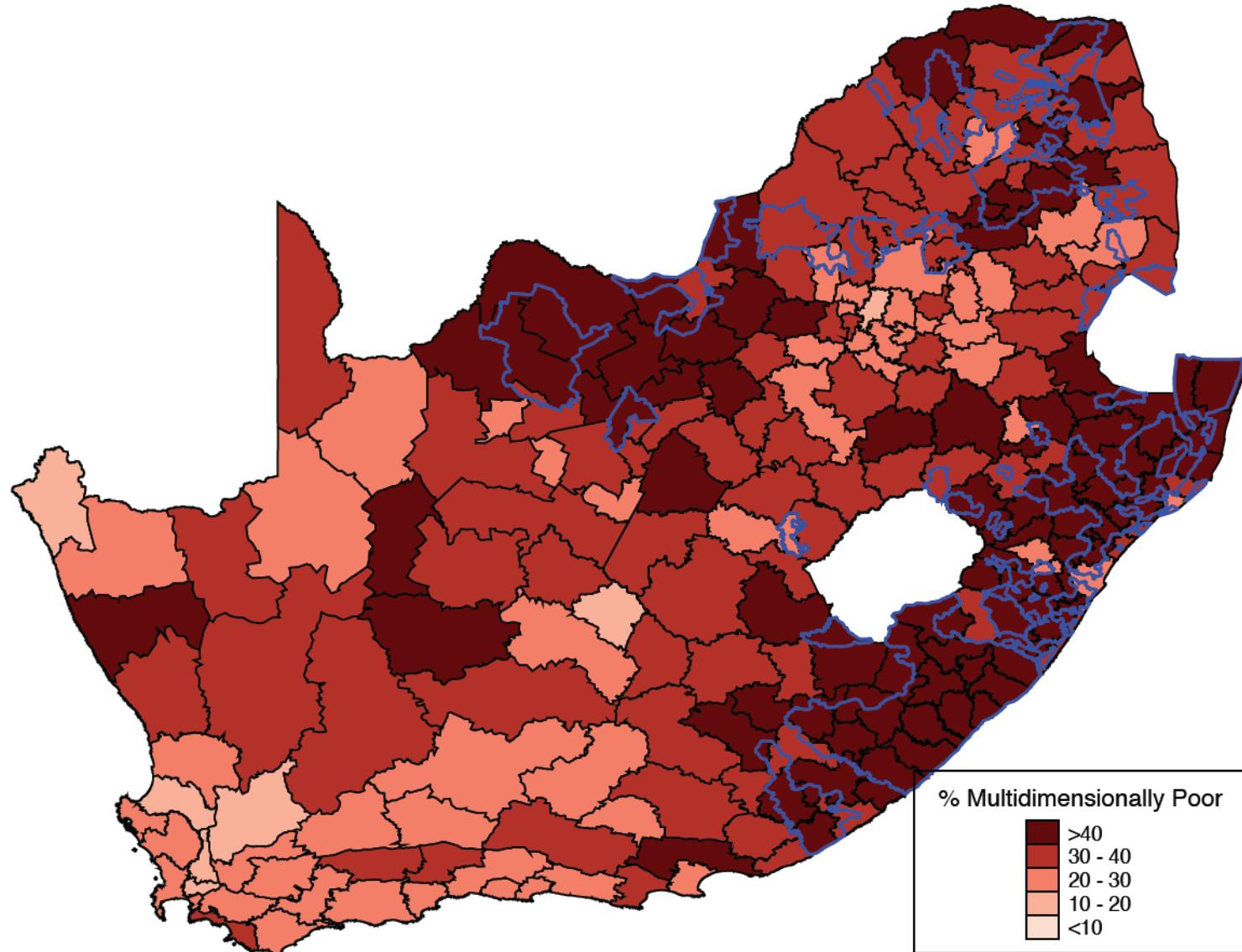
- Labour market dynamics – firm data, tax data, demand for labour, trade-non-trade relationships, inclusivity of growth
- Middle Class issues (panel data) – Panel data Inclusivity of Growth from the supply side
- Rich multidimensionality Multidimensional interactions.

## Top

- Taxes and Survey data, is crucial at the top end. High inequality often means that taxable income only comes from the top 20% or so

# The Bottom End

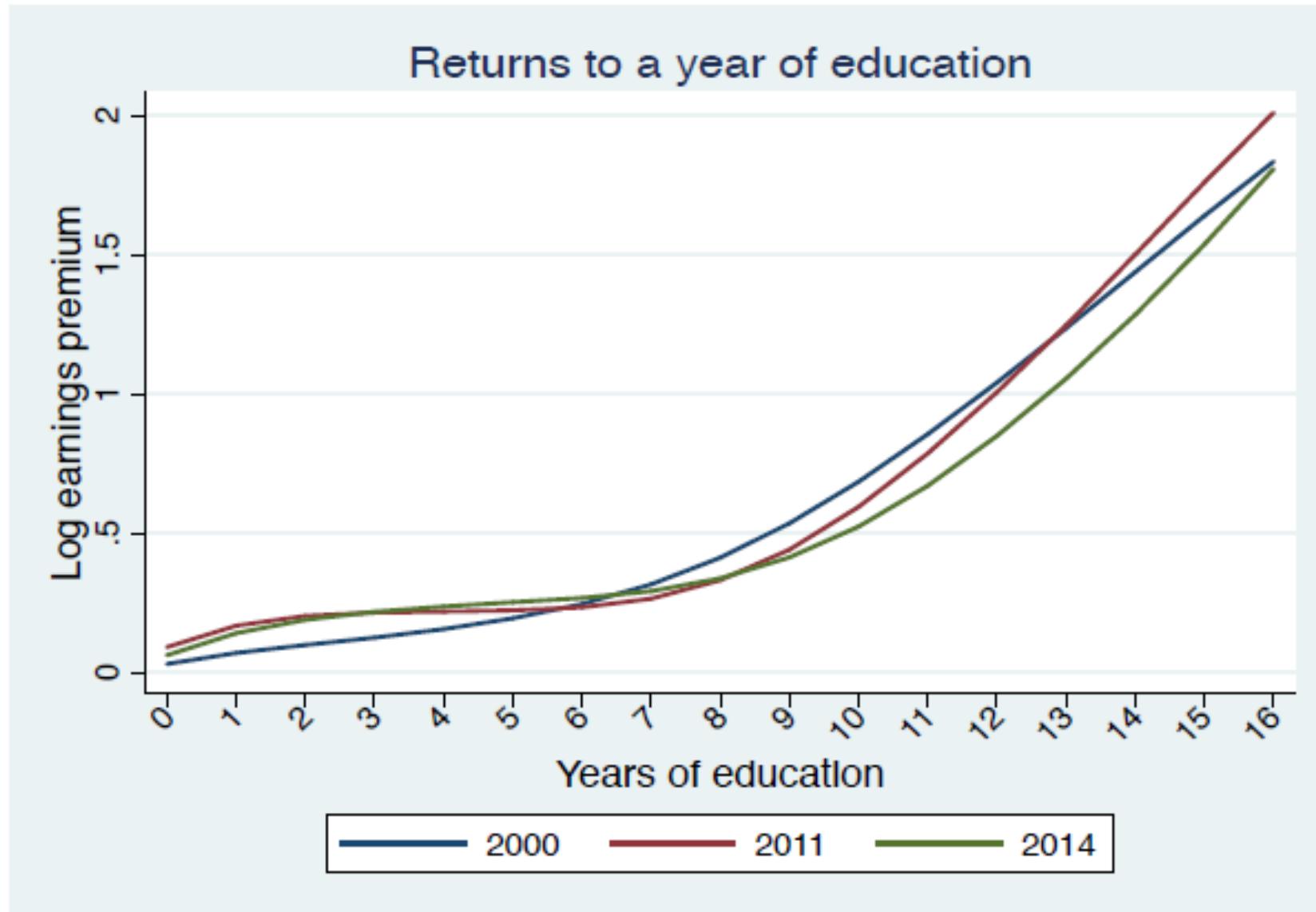
# Incidence of multidimensional poverty amongst youth in South Africa, by municipality, 2011 - including former homeland boundaries



# Non-tax Administrative Data

- Multidimensional with a focus on the interactions. Examples from NIDS.
- Quality of services comes from the admin data

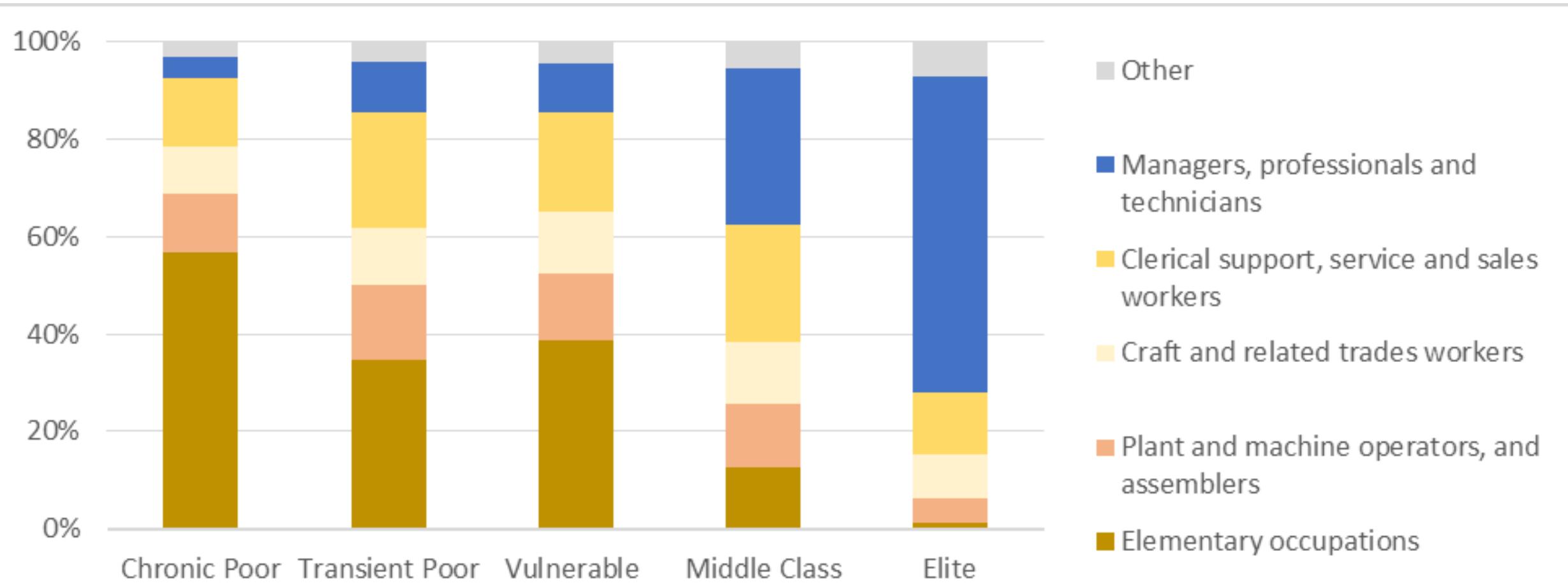
Figure 7: Returns to an additional year of education relative to no schooling



Source: Own calculations using PALMS V3.1 dataset.

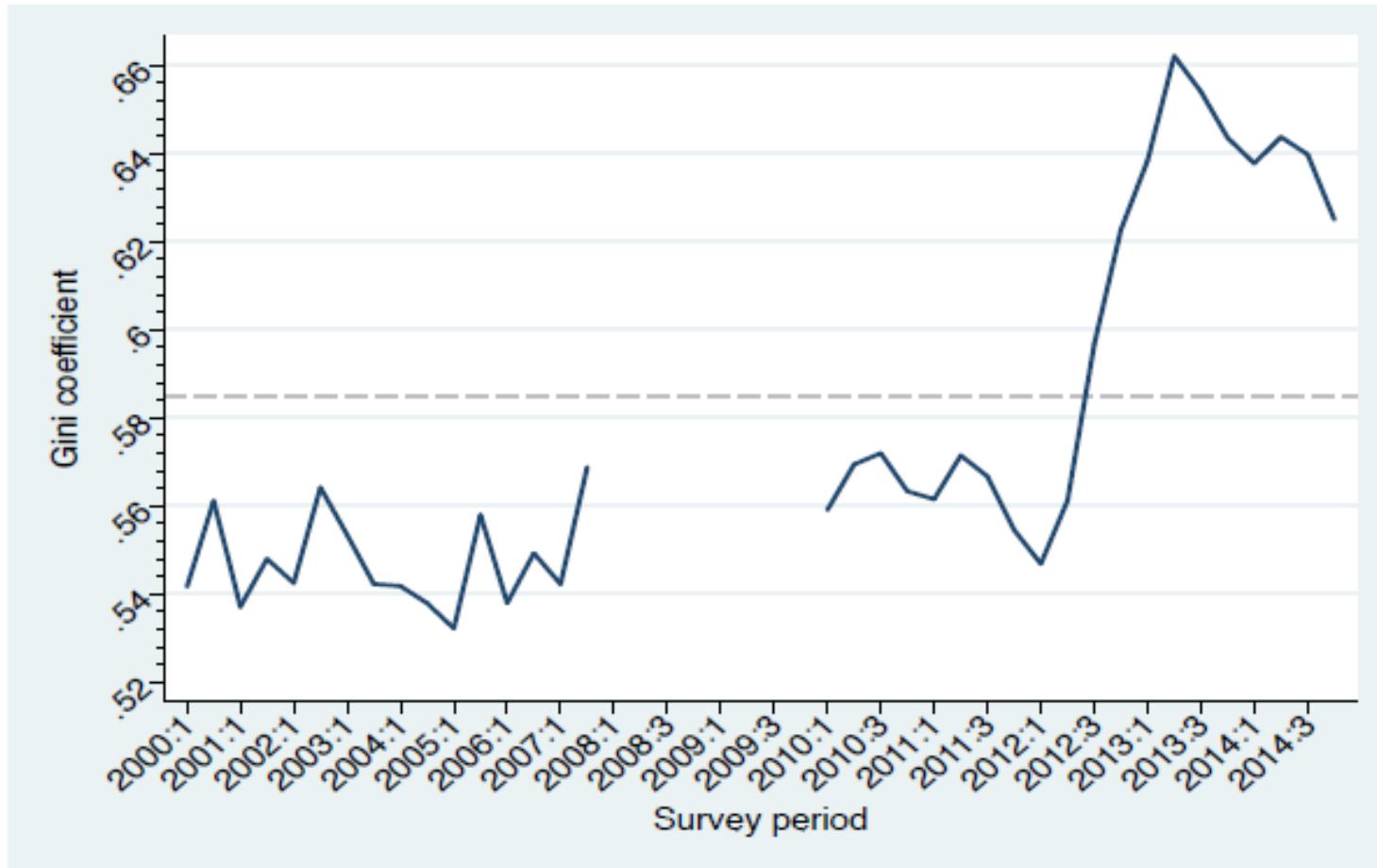
Back to the labour market and the middle parts of the distribution

# Occupation of the household head (employees only)



Be serious about interrogating the quality of the data  
**within** each data source as the merging between

Figure 3: The Gini coefficient of earnings over time



Source: Own calculations using PALMS V3.1 dataset.

# How do we understand changes in the demand for labour?

- Company Tax Data
- Firm size and exports/firm size and responsiveness to ETI
- Matched Employer-employee data
- Behaviour of firms (very limited). Limited variables
- But also the harder to measure self-employment, casualised employment, subsistence farming
- Firm Surveys?

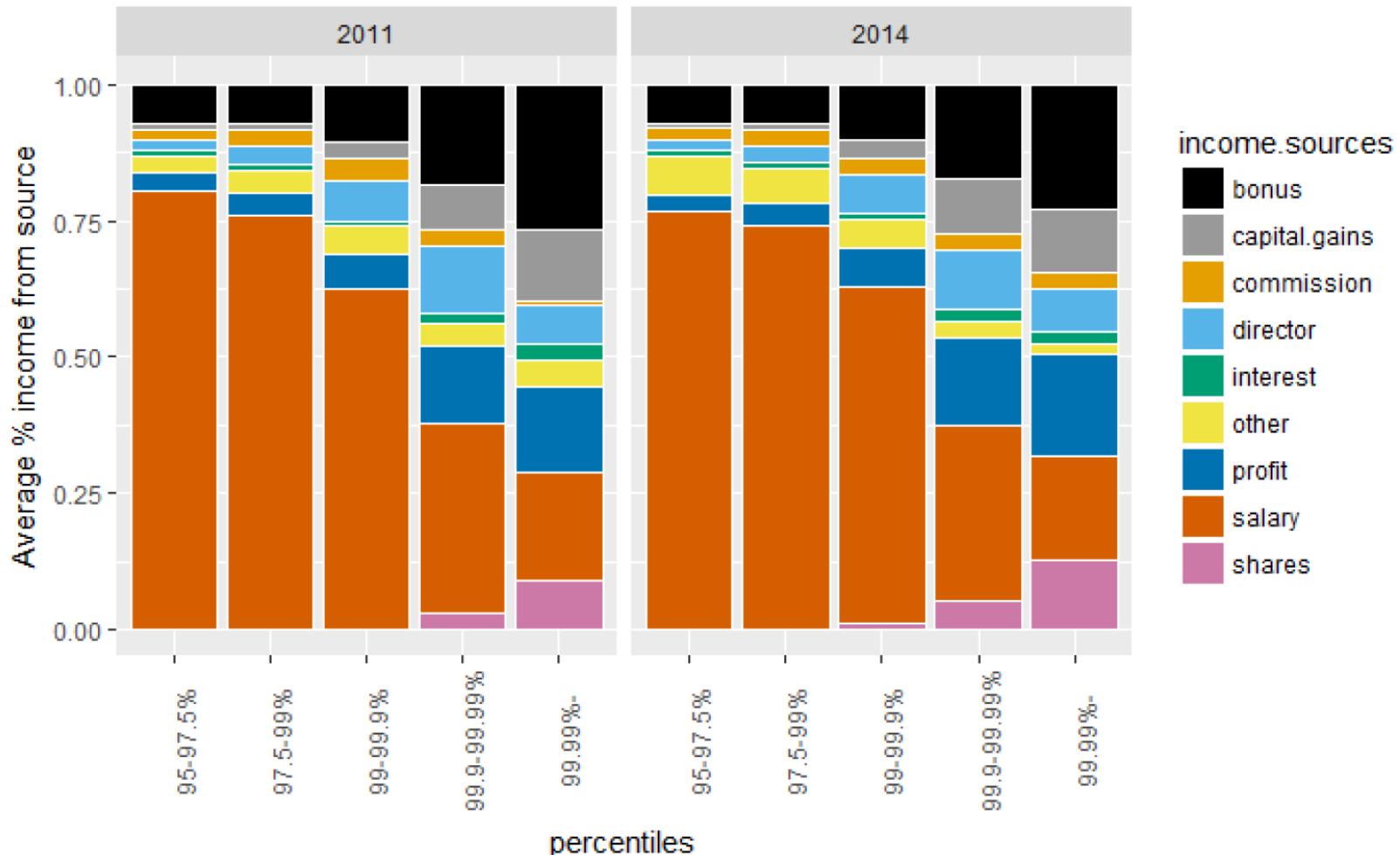
# The Top End

	Employees	Self-employed	Everyone
QLFS data	0.567	0.684	0.592
Adjusted QLFS data	0.599	0.716	0.620

Table 4: Gini coefficients for earnings adjusting for underreporting

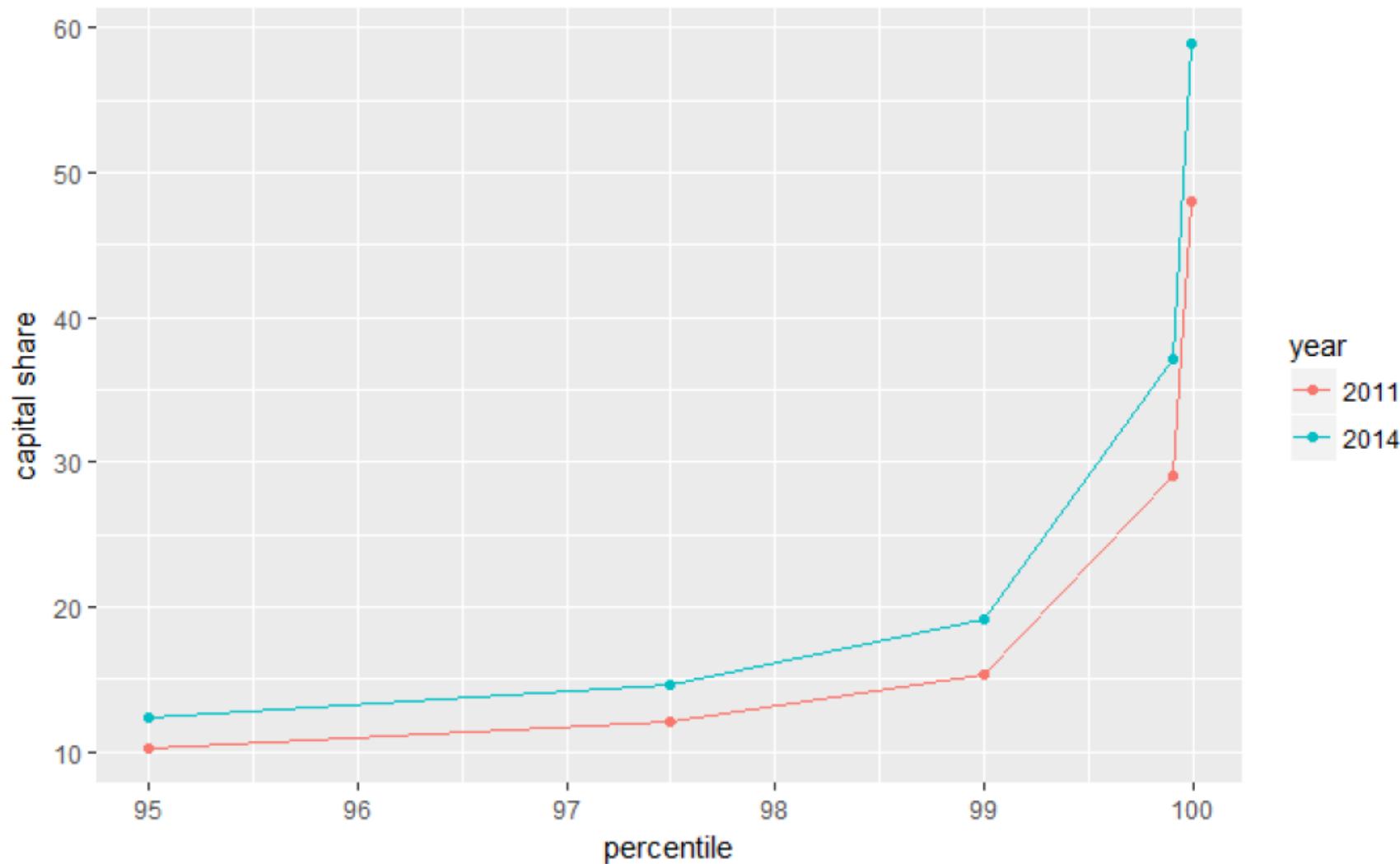
# There is additional information in Tax Data

Figure 7: Sources of income by percentile and year



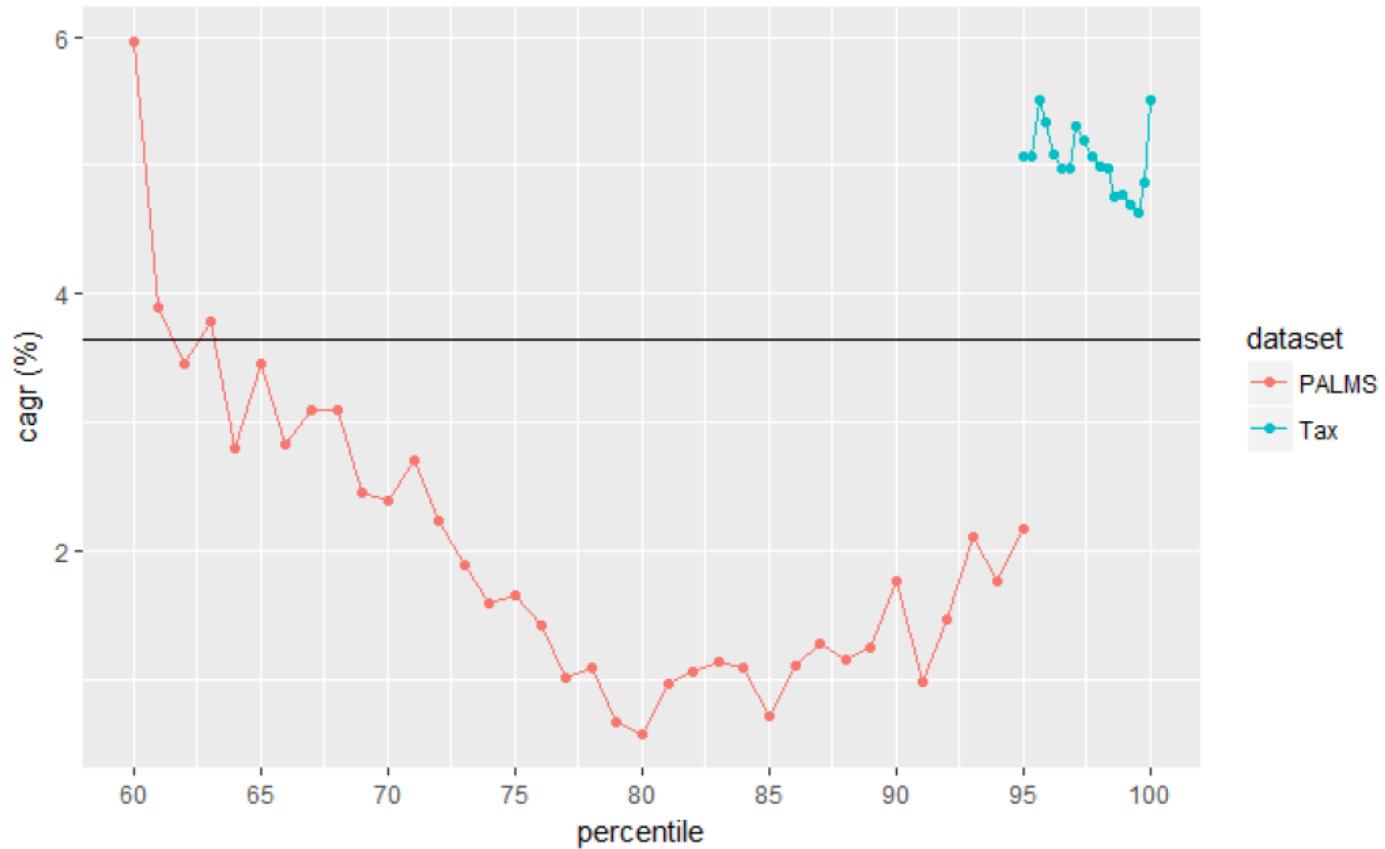
Source: South African Revenue Service (2015), Authors' calculations.

**Figure 9: Share of income from capital**



Source: South African Revenue Service (2015), Authors' calculations.

**Figure 6: Growth incidence curve 2003-2015 from combined survey and tax data**

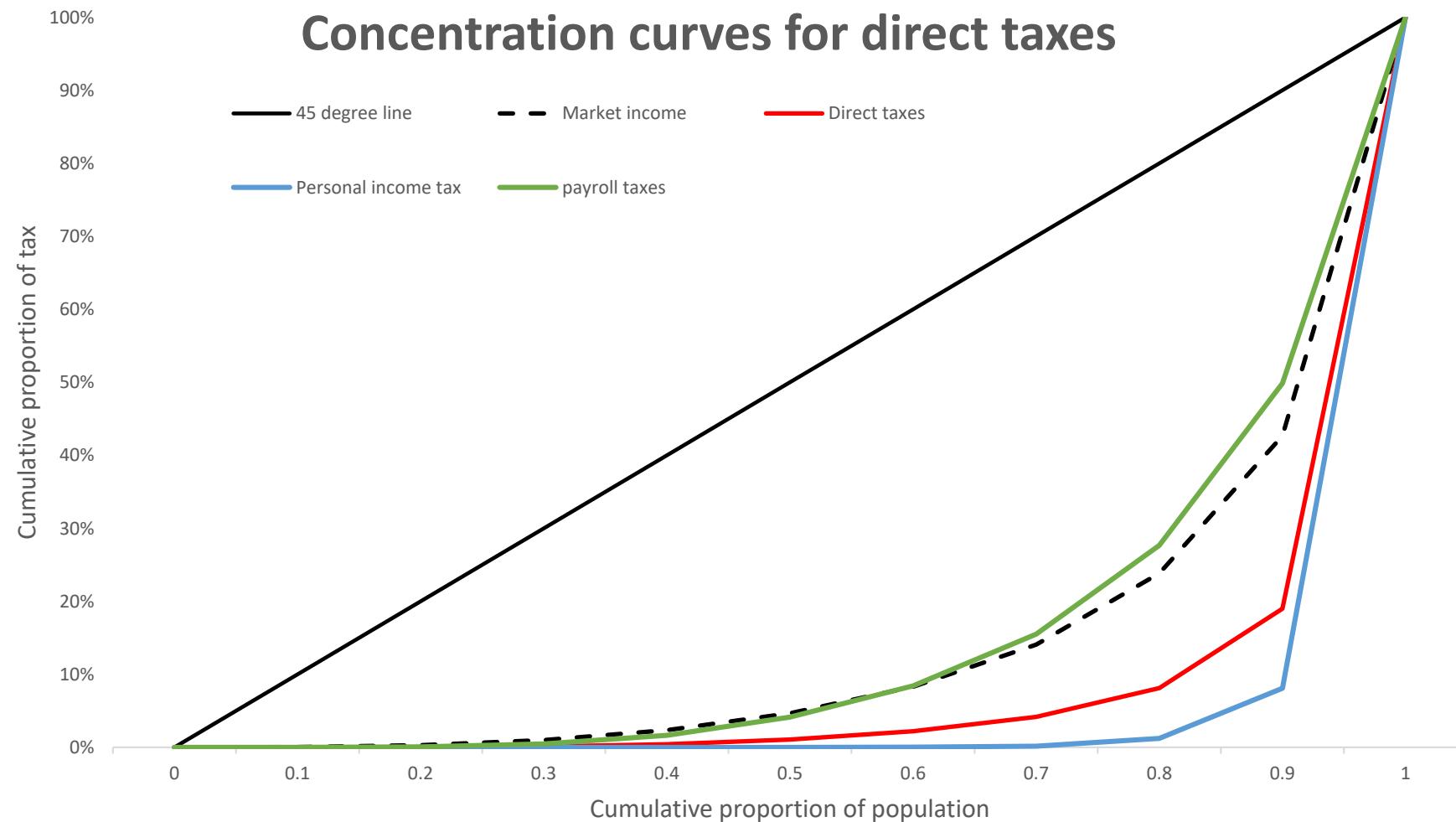


*Notes:* Horizontal line indicates GNI growth over the period.

*Source:* Tax data from South African Revenue Service (2017), survey data from Post-Apartheid Labour Market Series (2017). Authors' calculations.

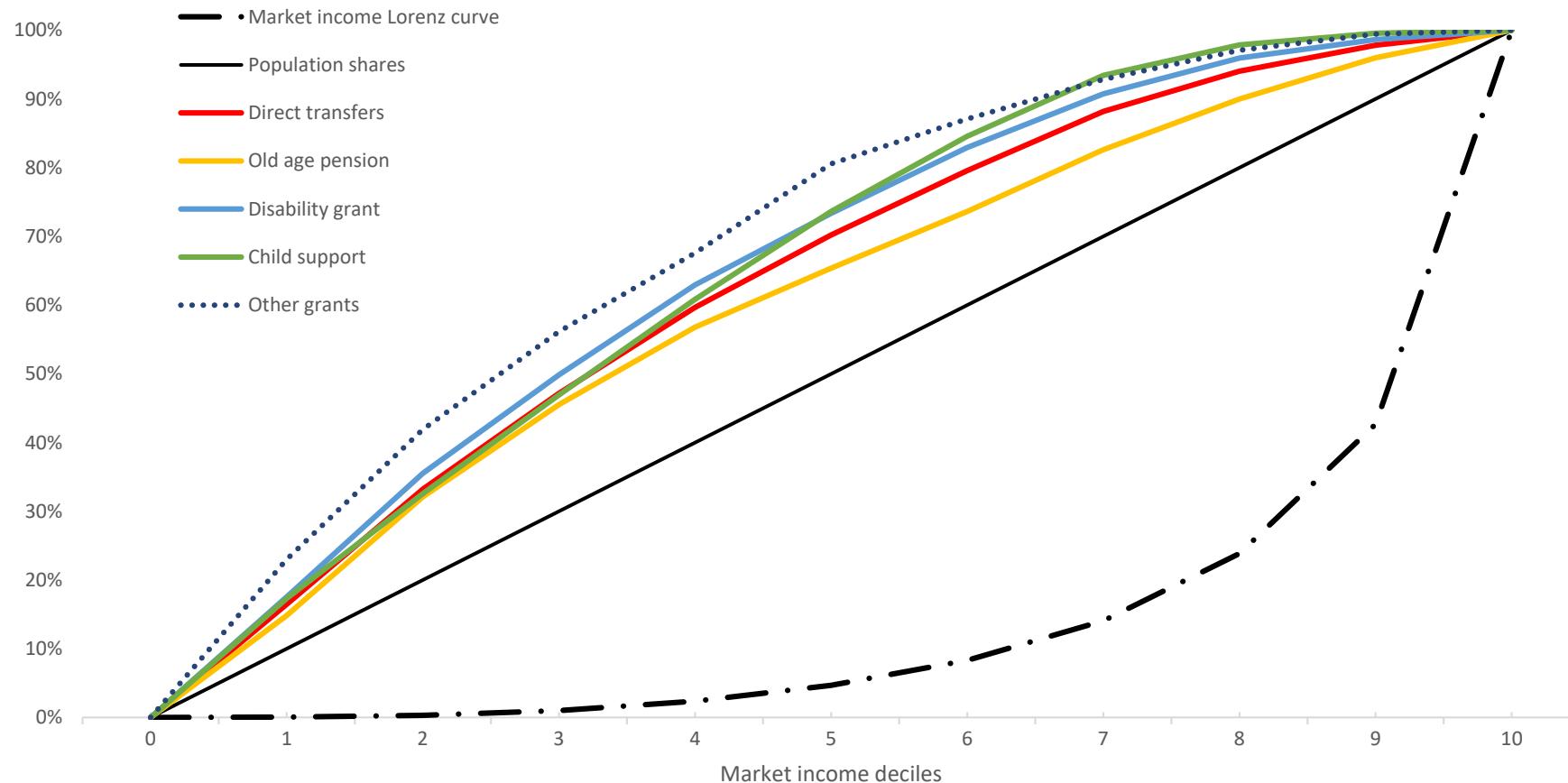
# Towards Policy: Assessing the impact of taxes and transfers on poverty and inequality in SA

# TAXES



# Social Expenditures

Concentration curves for social cash grants in South Africa, by decile



# Concluding Points

- There are extremely high returns to combining data in terms of better understanding and better policy making
- Even without tax data it is worth thinking hard about how to use administrative data well and in combination with household surveys and the census
- Admin data essential to take us to quality of services
- But, all data have their flaws
- Engagement with the NSOs is crucial
- Have to imbed these activities in local research-policy making processes

## Premium of earnings in tax data over QLFS earnings

