The Gender Wage Gap: Insights From Administrative Tax Data

Michelle Pleace, Matthew Clance, Nicky Nicholls

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
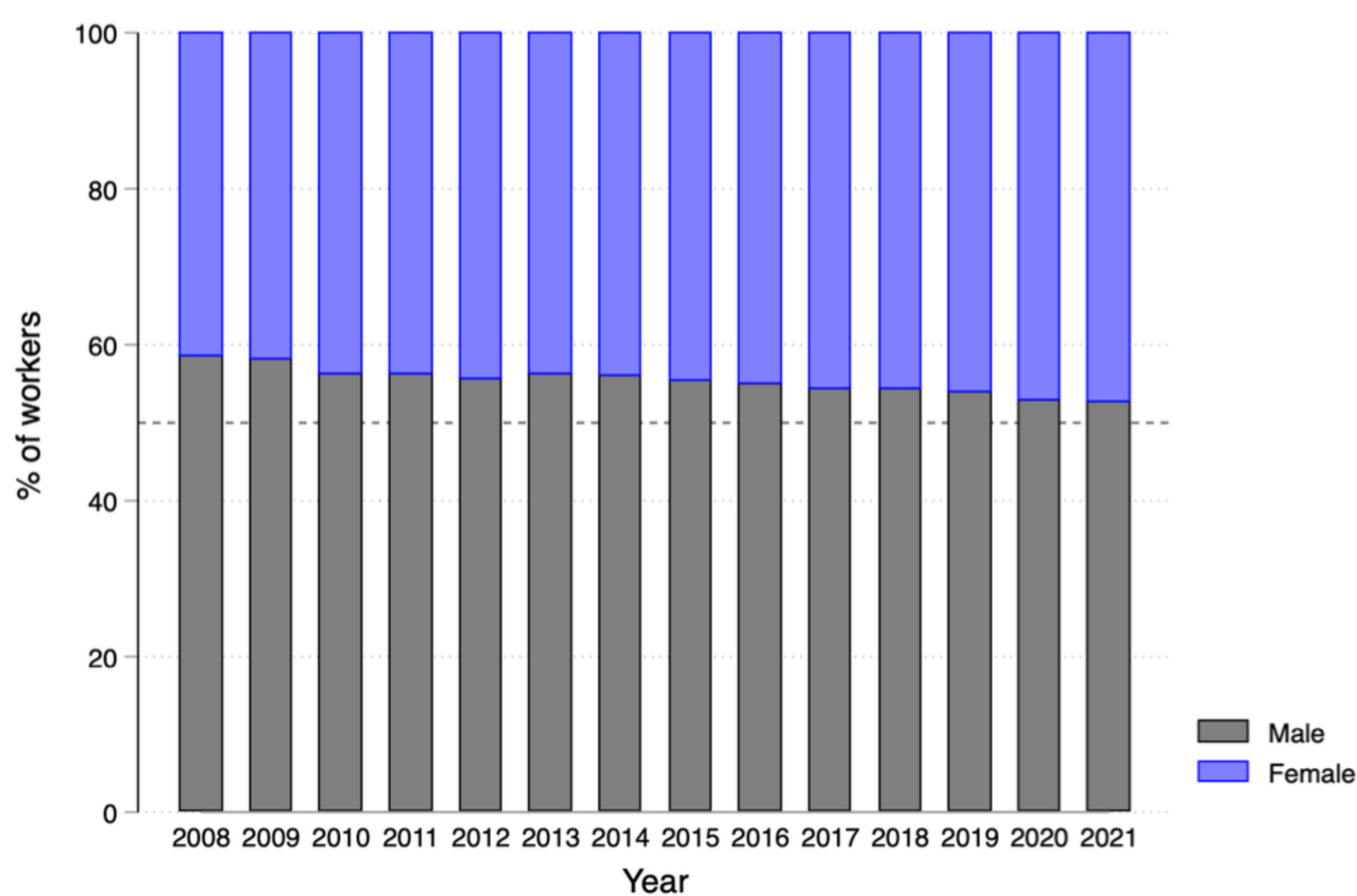YUNIBESITHI YA PRETORIA

1. Objectives

- Obtain an estimate of the gender wage gap for South Africa's formal economy (2008-2021).
- Determine the male income premium at the mean of the income distribution and at the conditional quantiles of the income distribution.
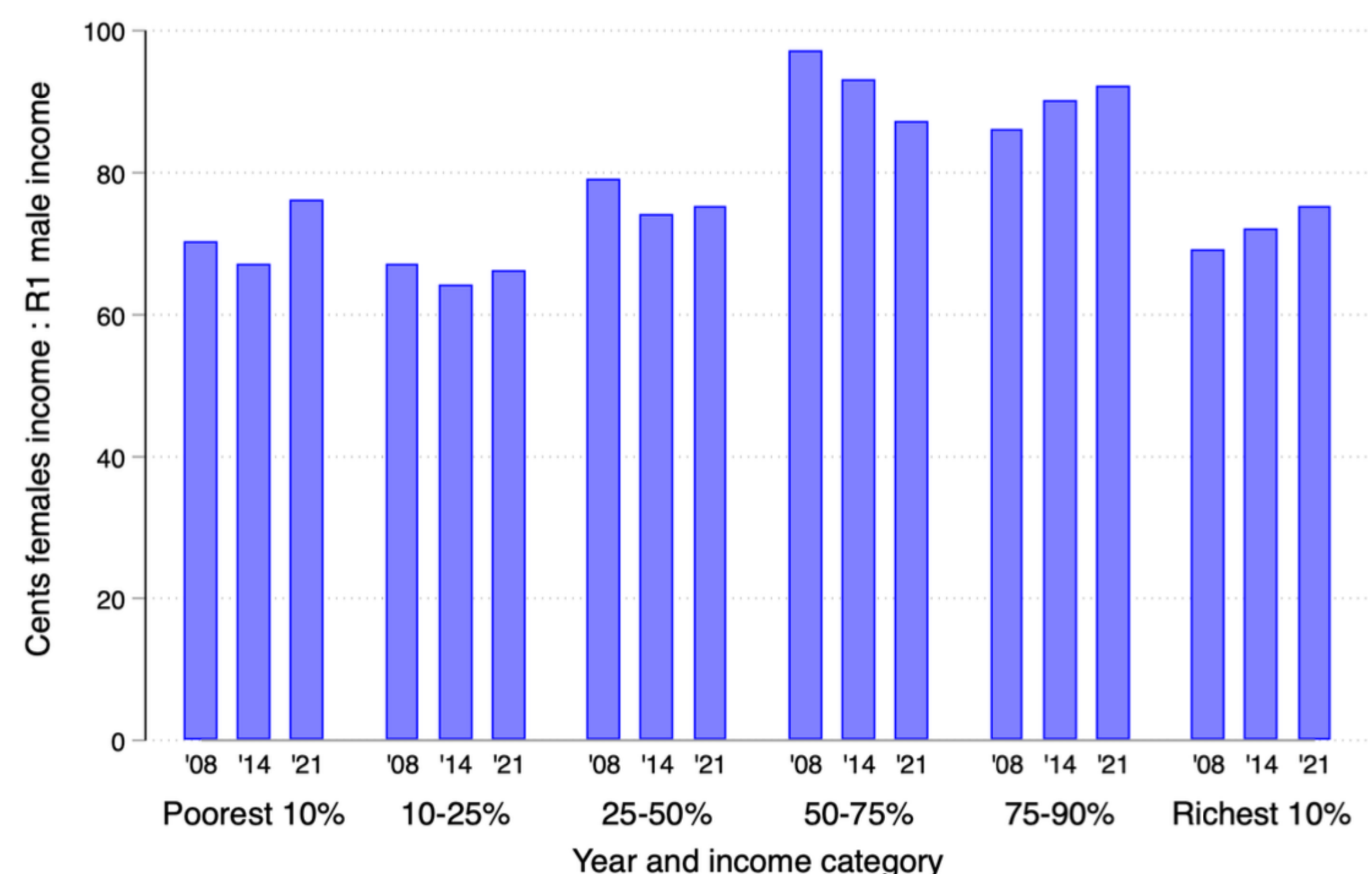
2. Data

- Our study uses the South African Revenue Service administrative tax data for 2008-2021 (National Treasury and UNU-WIDER, 2022). The dataset is constructed from anonymised individual tax records which disclose actual income earned.
- We have data relating to labour income, age, industry that an individual works in and province that an individual lives in.

3. Gender Employment Gap

- The demand side of the labour market shows that formal employment numbers are equalising over time.
- Females may be more active in the labour force due to households needing multiple incomes, a change in social norms, and female-headed households needing income to support their families.

Percent of employed people by gender (2008-2021)

4. Gender Pay Gap by Income Categories

- Gender wage gap persists across the income distribution, where the tails have the lowest levels of parity.
- At the 'Poorest 10%' category, females saw the highest increase in earnings relative to male labour income from 2014 to 2021, possibly due to the implementation of the National Minimum Wage in 2019.
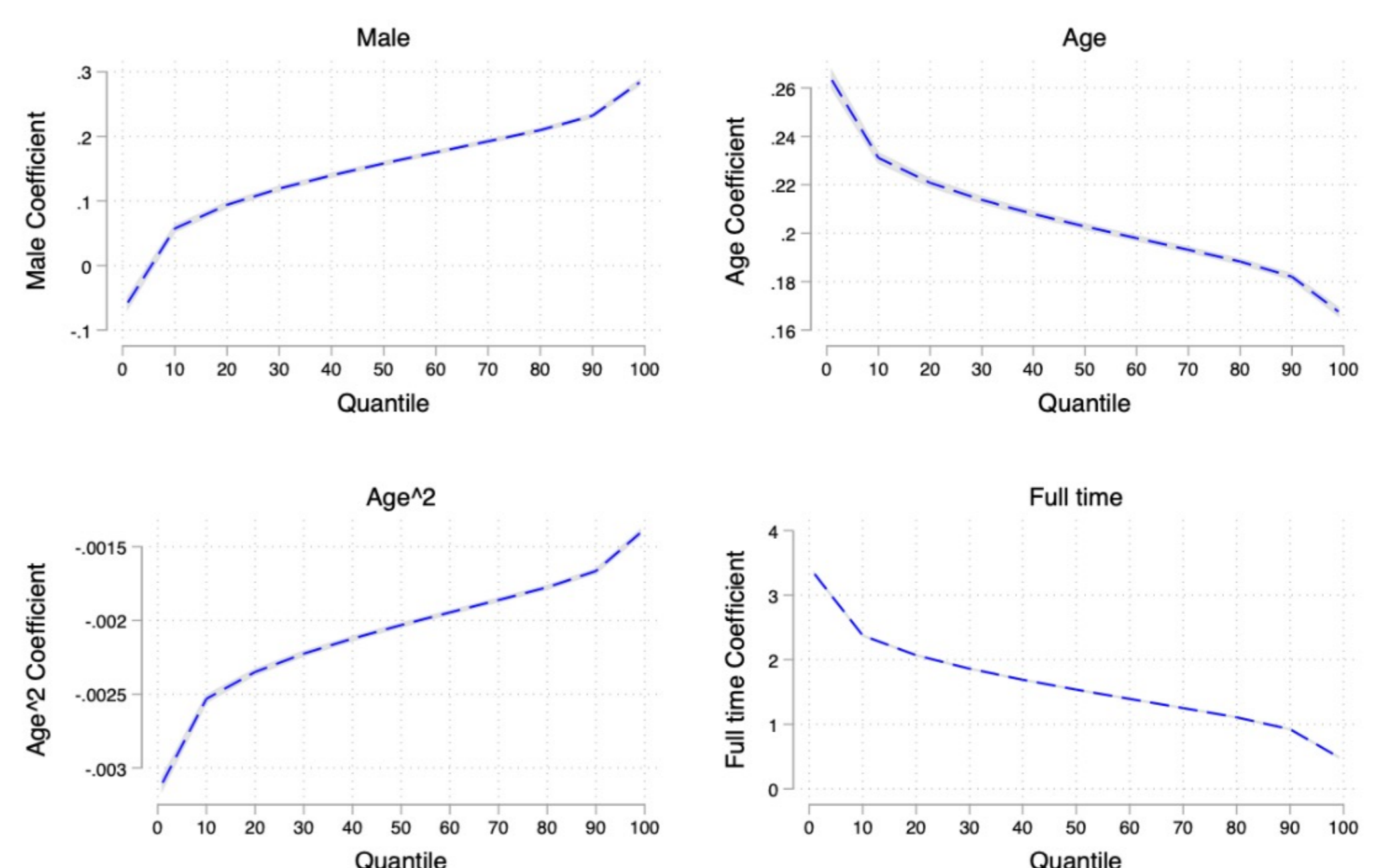
Female cents earned per R1 male income by gender and income category (2008-2021)

5. Gender Pay Gap at the mean

- Males have an income premium of 35.6% (column 7).
- The male income premium is smaller in a high-skill sector (23.4%) relative to the other-skill sector (35.6%).
- The high-skill male income premium can reflect the occupational segregation gap: in the health care industry women are predominantly nurses, while men are doctors.

Linear regression estimates: the impact of gender on real wage earned

	Dependent variable: Natural log of real income						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Male	0.305***	0.329***	0.293***	0.298***	0.248***	0.301***	0.305***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
High skill		0.434***	0.415***	0.410***		0.451***	0.447***
		(0.00)	(0.00)	(0.00)		(0.00)	(0.00)
Age			0.181***	0.179***	0.134***	0.180***	0.179***
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age^2			-0.002***	-0.002***	-0.001***	-0.002***	-0.002***
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Full time			1.781***	1.781***	1.577***	1.781***	1.781***
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Male · High skill						-0.093***	-0.094***
						(0.00)	(0.00)
Predictive margins: real wage difference by gender and skill level							
Male						0.292***	0.296***
						(0.00)	(0.00)
Male · Other skill						0.301***	0.305***
						(0.00)	(0.00)
Male · High skill						0.208***	0.210***
						(0.00)	(0.00)
Province included	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year included	No	No	No	Yes	Yes	No	Yes
Industry included	No	No	No	No	Yes	No	No
Obs	60776758	60776758	60776758	60776758	60776681	60776758	60776758
Adj. R^2	0.034	0.039	0.403	0.408	0.503	0.403	0.408

Standard errors clustered at the individual level are in parentheses.

Male is coded one for male individuals and zero otherwise. *High skill* refers to the industries that are considered to require a high skill level to work in as defined by Chen et al. (2022), these industries are: 1) 'Legal Activities'; 2) 'Scientific Research and Development'; 3) 'Architectural and Engineering activities, Technical testing and Analysis'; 4) 'Education Services'; 5) 'Human Health Activities'; 6) 'Accounting, Bookkeeping, Auditing activities and Tax consultancy'; 7) 'Activities of head offices and Management Consultancy Activities'; 8) 'Commercial Art and Photography'; 9) 'Advertising and Market Research'. *Age* indicates the age of the individual, and *Age^2* is the quadratic function of the variable *age*. *Full time* returns one when the individual is a full-time employee and 0 otherwise. *Province* indicates where the individual lives. *Year* controls for the year. *Industry* controls for the industry that an individual works in using a 5-digit SIC code.

* p<0.10, ** p<0.05, *** p<0.01

6. Gender Pay Gap in the high-skill sector at conditional quantiles

- The gender wage differential is highest at the right tail of the conditional income distribution. In 2021, men occupied 68.4% of management positions (Stats SA, 2021) where top executive women earn 72c for every R1 of male income (PWC, 2020).

Quantile regressions: Influence of factors over the conditional income distribution for high-skill population

7. Conclusion

- The number of employed women in the formal sector increased.
- The gender wage gap is most pronounced at the tails of the income distribution, where policy interventions should be focused.

8. References

National Treasury and UNU-WIDER (2022). 'IRP5 Worker-Level Data 2008–2021 [dataset]. Version 5.0 (beta)'. Pretoria: South African Revenue Service [producer of the original data], 2022. Pretoria: National Treasury and UNU-WIDER [producer and distributor of the harmonized dataset], 2022.

Price Waterhouse Cooper. (2020). Practices and Remuneration Trends Report. (August).

Statistics South Africa. (2021). *Quarterly Labour Force Survey: Quarter 2 2021* (tech. rep. August).