SA-TIED

Southern Africa – Towards Inclusive Economic Development

Product market competition and the labour market

Evidence from South Africa

Francesco Amodio, Michele Di Maio, Yifan Li, and Patrizio Piraino

SA-TIED Working Paper #105 | March 2020

UNITED NATIONS UNIVERSITY
UNU-WIDER

national treasury
Department: National Treasury
REPUBLIC OF SOUTH AFRICA

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
IFPRI

& evaluation
Department: Planning, Monitoring and Evaluatio
REPUBLIC OF SOUTH AFRICA

the dti
Department: Trade and Industry
REPUBLIC OF SOUTH AFRICA

SARS

TIPS

About the programme

Southern Africa –Towards Inclusive Economic Development (SA-TIED)

SA-TIED is a unique collaboration between local and international research institutes and the government of South Africa. Its primary goal is to improve the interface between research and policy by producing cutting-edge research for inclusive growth and economic transformation in the southern African region. It is hoped that the SA-TIED programme will lead to greater institutional and individual capacities, improve database management and data analysis, and provide research outputs that assist in the formulation of evidence-based economic policy.

The collaboration is between the United Nations University World Institute for Development Economics Research (UNU-WIDER), the National Treasury of South Africa, the International Food Policy Research Institute (IFPRI), the Department of Monitoring, Planning, and Evaluation, the Department of Trade and Industry, South African Revenue Services, Trade and Industrial Policy Strategies, and other universities and institutes. It is funded by the National Treasury of South Africa, the Department of Trade and Industry of South Africa, the Delegation of the European Union to South Africa, IFPRI, and UNU-WIDER through the Institute's contributions from Finland, Sweden, and the United Kingdom to its research programme.

Corresponding author: ppiraino@nd.edu

WIDER Working Paper 2020/39

Product market competition and the labour market

Evidence from South Africa

Francesco Amodio,[1] Michele Di Maio,[2] Yifan Li,[3] and Patrizio Piraino[4]

March 2020

Abstract: We study the relationship between product market competition and labour market outcomes in South Africa. We combine firm-level data from tax records with individual-level data from the labour force survey. We estimate markups across sectors, and derive a measure of employment concentration in high-markup sectors across South African district municipalities. We then test whether individual labour market outcomes differ systematically in those district municipalities where employment is more concentrated in high-markup sectors. We find that higher employment concentration in high-markup sectors is associated with higher unemployment and lower likelihood of transitions from unemployment to employment. This is differentially more the case for non-White, lower educated, and young individuals. The relationship remains strong when conditioning on a rich set of individual- and district-level covariates, including employment concentration per se.

Key words: product market competition, tax data, unemployment

JEL classification: D24, J64

Acknowledgements: We are thankful to Jan Eeckhout and all participants at the SA-TIED meeting in Pretoria (November 2018). Errors remain our own. We gratefully acknowledge funding from UNU-WIDER, SSHRC Canada, and the CEPR/DFID PEDL initiative.

---

[1] McGill University, Montreal, Canada and CIREQ, Montreal, Canada; [2] University of Naples Parthenope, Naples, Italy; [3] McGill University, Montreal, Canada; [4] University of Notre Dame, Notre Dame, Indiana, USA; corresponding author: ppiraino@nd.edu

This study has been prepared within the UNU-WIDER project 'Southern Africa—Towards Inclusive Economic Development (SA-TIED)'.

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9256-796-5

https://doi.org/10.35188/UNU-WIDER/2020/796-5

Typescript prepared by Gary Smith.

The United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

1 Introduction

The relationship between competition in the product market and labour market outcomes is gaining increasing attention in the academic literature, as well as in the policy debate. Evidence from the USA shows that market concentration correlates negatively with the labour share across industries (Autor et al. 2017; Barkai 2017). In addition, lower competition leads to higher prices and lower demand. The demand for labour decreases as a result, and so do real wages even if the aggregate labour supply is perfectly elastic (De Loecker et al. 2018). Finally, the degree of adjustment of variable inputs—including labour—is lower when markups are higher. These arguments altogether suggest that lack of competition in the product market is not only detrimental to consumer welfare through prices, but has implications for labour demand, wages, and labour market dynamism.

We investigate empirically the validity of these claims using data from South Africa, a country characterized by persistently high unemployment. Since 1997, the unemployment rate in South Africa has always been higher than 20 percent, with peaks of more than 30 percent. While declining, unemployment remained high even between 2001 and 2008, when the economy was growing at a rate of more than 2 percent per year (World Bank 2017). The available evidence suggests that this is not the result of temporary shocks, but rather due to a number of structural factors (Banerjee et al. 2008). The high and persistent unemployment rate is a puzzle that has yet to be fully explained (IMF 2011).

This paper studies the relationship between market power (lack of competition) in the product market and labour market outcomes in South Africa. For this purpose, we combine firm-level data from the South African Revenue Service and National Treasury (SARS-NT) from 2009 to 2014 with Quarterly Labour Force Survey (QLFS) data for the same years. We estimate markups across sectors using the methodology of De Loecker and Warzynski (2012). We combine this information with baseline employment shares in the QLFS data to derive a measure of employment concentration in high-markup sectors across district municipalities. The combination of these two measures allows us to answer the following questions:

1. Do labour market outcomes differ systematically in those districts where employment is more concentrated in high-markup sectors?

2. Do unemployment probabilities and patterns of transition from unemployment to employment differ in districts where employment is more concentrated in less competitive sectors?

Our analysis delivers three sets of results. First, we find that unemployment is higher in those district municipalities where employment is more concentrated in high-markup sectors. Second, we find that the likelihood of transitioning out of unemployment and into employment is also lower in those districts. These results are robust to conditioning on a wide set of covariates at the individual and district municipality levels. Importantly, conditioning on the extent of employment concentration per se does not affect the results (Azar et al. 2017; Benmelech et al. 2018). Third, we find the negative relationship between unemployment and product market competition to be stronger for non-White, lower educated, and young individuals.

Our analysis and results have clear policy relevance. The presence of a strong relationship between market power in the product market, on the one hand, and labour market rigidities and unemployment, on the other, suggests that competition and concentration play a role in shaping income inequality, and that the reasons for the high unemployment in South Africa must not be restricted to the usual culprits—skill mismatch or the strength of unions—but must include the lack of competition in the product market. Market concentration also affects consumer welfare through prices, raising the need for policy intervention in this domain. This implies that the policies designed to increase competition in the domestic market would benefit consumers, and, more importantly, contribute to the increase of labour demand and the dynamism in the labour market.

The remainder of this paper is organized as follows. The next section describes the data we use in our empirical analysis. Section 3 illustrates the empirical strategy, Section 4 presents the results, and Section 5 concludes.

2 Data and measurement

For the purpose of our analysis, we combine two main sources of data from South Africa, one at the individual level and one at the firm level.

2.1 Individual-level data

We use the first wave of the 2008 QLFS to derive employment shares by sector at the district municipality level. The data contain information on the industry of each employed individual. This allows us to match the three-digit industry category in the QLFS data to its corresponding International Standard Industrial Classification Rev. 4 (ISIC4) sector classification, so that we can derive weighted measures of local labour market employment compositions at that level.[1] These can then be directly linked to the markups estimated on the SARS-NT data. We derive unemployment stocks and flows using QLFS data from 2009 to 2014. We draw this information from the Post-Apartheid Labour Market Series (PALMS), which is a stacked cross-sectional dataset created by DataFirst at the University of Cape Town. The QLFS surveys are regarded as the most reliable source of labour market data in South Africa. We will compute the measures needed for our analysis at the district municipality level. While the QLFS samples are not designed to be representative at this level, our empirical analysis should not be invalidated to the extent that the degree to which data are unrepresentative is uncorrelated with our variables of interest (see also Magruder 2012). In addition, our benchmark specifications control extensively for both observed and unobserved local-level heterogeneity at the district municipality level.

The QLFS also has a rotative panel component tracking observations by dwelling identifier. We use this information to construct individual-level panel data. We identify individuals using information on gender and age in the same dwelling across waves. This is to minimize the risk that respondents found in the same dwelling may in fact be different household members or unrelated individuals who moved in between waves.

Table 1 reports the summary statistics for the variables we use at the individual level, namely unemployment status, an indicator variable for the transition from unemployment to employment between two consecutive quarters, and a range of individual-level characteristics that we use as control variables in the regression analysis.

[1] We define our benchmark sectors by two-digit Standard Industrial Classification (SIC) sector classifications, which are extracted from the first two digits in the three-digit industry categories used in QLFS. As the SIC classification is derived from the ISIC Rev. 4, we manually match the ISIC industry categories to the two-digit SIC categories based on Statistics South Africa's 'Classifications Standard Industrial Classification of all Economic Activities (SIC) Seventh Edition'.

Table 1: Summary statistics: individuals

	Mean	Std dev.	Min.	Max.	Obs.
Unemployed	0.13	0.33	0.00	1.00	1,170,150
Not in the labour force	0.40	0.49	0.00	1.00	1170150
Discouraged worker	0.06	0.24	0.00	1.00	1,170,150
Unemployment–employment transition	0.05	0.22	0.00	1.00	222,783
Age	35.10	14.01	15.00	64.00	1,170,150
High-educated	0.34	0.47	0.00	1.00	1,170,150
Female	0.55	0.50	0.00	1.00	1,170,150
African	0.78	0.41	0.00	1.00	1,170,150
Coloured	0.12	0.32	0.00	1.00	1,170,150
Married	0.37	0.48	0.00	1.00	1,170,150
Widow/widower	0.05	0.21	0.00	1.00	1,170,150
Divorced or separated	0.03	0.17	0.00	1.00	1,170,150
Household size	2.77	1.49	1.00	23.00	1,170,150

Notes: the table reports summary statistics for the variables used throughout the empirical analysis. We focus on working-age individuals (aged 15–64) in the PALMS dataset from 2008 (Wave 23) to 2014 (Wave 50); observations are weighted by survey sample weight. This table reports individual employment status and demographic characteristics. The definition of being unemployed includes only individuals who are actively searching for work. The discouraged workers are those who are not searching but are willing and able to work. Individuals who are not willing and/or not able to work are defined as not in the labour force. The implied unemployment rate over the period considered is 0.24. High-educated people are defined as people having 12 or more years of schooling. Restricting the sample to all individuals that are unemployed at time $t-1$, Unemployment–employment transition is a dummy variable which takes value 1 if the individual is unemployed at $t-1$ and becomes employed at time t, and 0 otherwise. Individual demographic characteristics includes age, education, gender, race, marital status, and household size.
Source: authors' own construction based on QLFS data.

2.2 Firm-level data

Our second source of information is the South African National Treasury's firm-level dataset. This is an unbalanced panel dataset created by merging several sources of South African administrative tax data. We use the newly available firm-level data that the Treasury has prepared in collaboration with SARS-NT. This firm-level data contains information on inputs, outputs, employment, and prices from tax records. In fact, we can either directly observe or derive the value of: sales, capital stock, labour stock (from the weighted sum of IRP5 forms) and other variable inputs, investments, and total variable costs of production. These data are available from 2009 to 2014. We also observe the two-digit ISIC4 sector the firm belongs to. This information is all that is needed to estimate markups as in De Loecker and Warzynski (2012). Table A1 in the Appendix shows the number of firms per two-digit sector that are used in the estimation of markups.[2]

2.3 Sector-level markups

Our empirical analysis starts with the calculation of markups at the sectoral level. Defined as the ratio of price over marginal cost, markups can be estimated using different methods employed in the empirical industrial organization literature. As detailed data on price and marginal cost are usually unavailable, markup estimation depends on the granularity of the available data and the choice of assumptions. On one hand, the 'demand-based' approach requires assumptions on the shape of the demand function and market structure, so that marginal cost and markups are estimated with the associated demand elasticity and firms' optimal pricing behaviour, as in Bresnahan (1982) and Berry et al. (1995). On the other hand, the 'production-based' approach proposed by De Loecker and Warzynski (2012) builds upon the insight

[2] Note that the industry classification used in the SARS-NT firm panel is the two-digit ISIC4. The ISIC4 industry categories are matched to the two-digit SIC categories we defined in the QLFS before the estimation of markups.

of Hall (1988) on the relationship between price and marginal cost. Under the assumption that firms minimize total production costs, markups can be derived as the product of the input revenue share and output elasticity of any chosen variable input, where the output elasticity is inferred from the estimation of the production function.

In our analysis, we adopt the methodology of De Loecker and Warzynski (2012), which has been widely used in a number of recent contributions to the literature (De Loecker et al. 2018, 2016). We formally illustrate this approach in the Appendix. In a nutshell, this method leverages a cost-minimization framework to show how markup–the ratio between output price and marginal cost—is equal to the product of the inverted revenue share of variable inputs and their output elasticity. If markets are perfectly competitive, the revenue share of each input is equal to its output elasticity, and markup is equal to 1. If the firm has any degree of market power in the product market, the revenue share of each input is lower than its output elasticity, and markup is higher than 1.

The advantage of this approach is that it requires no assumptions on the demand structure. Intuitively, it is measured as technology-adjusted cost share. As the input cost shares are often available in the firm-level accounting data, we can estimate markup combining this input cost share with an estimate of the output elasticity. The output elasticities of variable inputs are essentially the parameters of the production function. We can thus assume and estimate a Cobb–Douglas production function using again standard techniques such as ordinary least squares (OLS), Olley and Pakes' (1996) and Levinsohn and Petrin's (2003). As a first step, we obtain estimates of markup μ_s of sector s as defined at the two-digit SIC level.

We adopt the 'production-based' method in our markup estimation for several reasons. First, we want to look at the broad impact of employment concentration in high-markup sectors on labour market outcomes. It requires comparison across all industries in the whole economy, rather than focusing on a specific industry. Therefore it is reasonable to impose as few assumptions as we can on the market structure in the markup estimation. Second, SARS-NT data cover all firms' balance sheet information in South Africa, from which the input share is directly calculated.

2.4 Employment concentration in high-markup sectors

As a second step, we combine information on employment shares across sectors in each district municipality from the QLFS with the estimated markups from SARS-NT to obtain a measure of employment concentration in high-markup sectors at the district municipality level. Let L_d be the number of employed individuals in district municipality d in year 2008 obtained from QLFS. Let L_{sd} be the number of individuals employed in sector s in the same district municipality d in the same baseline year. We let employment concentration in high-markup sectors in the district municipality be equal to

$$m_d = \sum_s \mu_s \frac{L_{sd}}{L_d} \tag{1}$$

where μ_s is the previously derived measure of markup at the sector level. Notice that m_d is higher if a larger share of workers in district municipality d is employed in sectors with higher estimated markups.

Notice that this measure of employment concentration in less competitive sectors is time-invariant and calculated according to baseline employment shares in 2008. This is because we want to rule out reverse causality as a possible source of bias, meaning the possibility that changes in labour market conditions affect the employment distribution across sectors.

Table 2 shows the summary statistics for the district municipality level variables that we use in our analysis, including the main variable of interest m_d. Other covariates measured at baseline are: the total population, the share of high-educated population,[3] and the share of African, Coloured, and White population in the district municipality.

Table 2: Summary statistics: district municipality level

	Mean	Std dev.	Min.	Max.	Obs.
Labour concentration in high-markup sectors	1.88	0.43	0.95	2.88	53
Population (millions)	0.55	0.55	0.05	2.41	53
Share of high-educated population	0.29	0.10	0.12	0.57	53
Share of Coloured population	0.15	0.26	0.00	0.92	53
Share of White population	0.10	0.08	0.00	0.29	53
Share of urban population	0.57	0.33	0.00	1.00	53

Notes: the table reports baseline district municipality controls in 2008 (wave 23). The district municipality levels used are those in the QLFS.
Source: authors' own construction based on QLFS data.

3 Empirical strategy

Our empirical analysis combines quarterly QLFS data on labour market outcomes from 2008 to 2014 with our estimated measures of employment concentration in high-markup sectors at the district municipality level to implement the following regression specification:

$$Y_{idt} = \delta_t + \tau_p + \beta m_d + \mathbf{X}'_{it}\gamma + \mathbf{Z}'_{dt}\sigma + u_{idt} \tag{2}$$

where Y_{idt} is the labour market outcome of interest for individual i living in district municipality d at time t. In the following, we consider two main outcomes. First, we consider unemployment status. The dependent variable is a dummy that takes value 1 if the individual is unemployed as narrowly defined (i.e. excluding discouraged workers). Individuals who are not willing and not able to work are defined as not in the labour force. Here, our main interest is the probability of being unemployed, so the zeros also include individuals who are discouraged workers or out of the labour force.[4] Second, we consider the probability of transitioning out of unemployment into employment. To this end, we exploit the longitudinal dimension of the QLFS dataset to track individuals across waves.[5] In this case, we restrict the sample to all individuals who are unemployed at time $t-1$. The outcome variable is a dummy variable that takes value 1 if the individual is unemployed at time $t-1$ and becomes employed at time t, and 0 otherwise.

m_d is the measure of employment concentration in high-markup sectors at the district municipality level as defined in Equation 1. Our coefficient of interest is β, which captures whether any systematic relationship exists between the level of employment concentration in high-markup sectors and labour market outcomes at the district municipality level.

$\mathbf{X}_{idt}$ consists of a vector of time-varying variables including age, age squared, if the individual has higher education, marital status (single, married, widow, divorced), and household size. $\mathbf{X}_{idt}$ also consists of a vector of time-invariant individual controls including gender and the population group (African,

[3] High-educated people are defined as individuals having 12 or more years of schooling.

[4] We also consider samples excluding individuals that are not in the labour force for robustness in Tables A3 and A4 in the Appendix.

[5] As the QLFS panel only tracks individuals by dwelling identifier, we construct our individual panel data for the same individual if he/she meets the following three conditions across waves: (1) same dwelling identifier; (2) same gender; (3) same age or one year older. This allows us to be reasonably sure it is the same individual linked through time, rather than respondents found in the same dwelling who may be different household members or unrelated individuals who moved in between waves.

Coloured, Indian/Asian, White). $\mathbf{Z}_{dt}$ is a vector of district municipality-level characteristics, such as total population, shares of different population groups, share of higher-educated individuals, and share of urban population—all measured at baseline. We also include as control a Herfindahl-type index of employment concentration at the district municipality level (Azar et al. 2017; Benmelech et al. 2018). This is because we want β to reflect the extent to which employment is concentrated in less competitive sectors, instead of capturing variation in employment concentration itself. The index is defined as

$$HHI_{dt} = \sum_{s} \left(\frac{L_{sdt}}{L_{dt}} \right)^2$$

where L_{dt} is the number of employed individuals in district municipality d at time t, and L_{sdt} is the number of individuals employed in sector s in the same district municipality d at the same time.

In our regressions, we also include a set of fixed effects. δ_t indicates quarter (wave) fixed effects which net out overall time trends. τ_p is the full set of province fixed effects to account for all time-invariant differences across broader geographical areas. We also consider a specification in which we include the full set of interactions between the province fixed effects and the quarter (wave) fixed effects. The province-wave fixed effects sweep out province-level shocks such as technological shocks and changes in policies or institutions that affect competition. u_{idt} captures any residual determinants of the outcome of interest. We cluster standard errors at the district municipality level to account for any correlation between residuals that belong to observations from different individuals and years but from the same district municipality.

4 Results

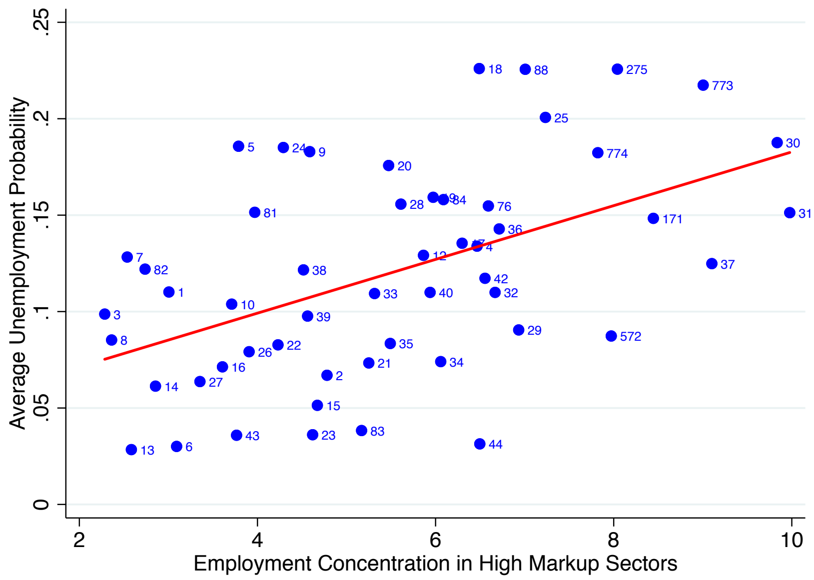
We start by plotting the probability of unemployment at the district municipality level against our measure of employment concentration in high-markup sectors across district municipalities. Figure 1 plots the unemployment probabilities (averaged over the period 2008–14) against m_d. The data indicate a strong positive correlation between the two measures, showing that probability of unemployment is higher in district municipalities with higher concentration in less competitive sectors.

Similarly, Figure 2 plots the transition probability from unemployment to employment (averaged over the period 2008–14) against employment concentration in high-markup sectors across district municipalities. The data indicate a negative correlation between the two measures. This suggests that in district municipalities where employment is more concentrated in less competitive sectors, the percentage of the unemployed who transition to employment in the following quarter is smaller.

Table 3 shows the first set of regression results. The labour market outcome of interest is the individual-level unemployment status—that is, the dependent variable is a dummy taking value 1 if the individual is unemployed and 0 otherwise. In column 1, we report the estimated coefficient for the individual probability of being unemployed from regression model 2 when we only include m_d—our measures of employment concentration in high-markup sectors at the district municipality level—and the set of time and province fixed effects. The estimated coefficient of interest is positive and significant at the 1 per cent level. Its magnitude is such that moving from the district with the lowest level of concentration in less competitive sectors ($m_d = 0.95$) to the one with the highest ($m_d = 2.88$) is associated with an increase in the probability of being unemployed of 0.58 percentage points, or a 4.4 per cent increase over the mean.[6]

[6] The results are similar when we exclude individuals that are not in the labour force from our sample, as shown in Table A3 in the Appendix: the magnitude is such that moving from the district with the lowest level of concentration in less compet-

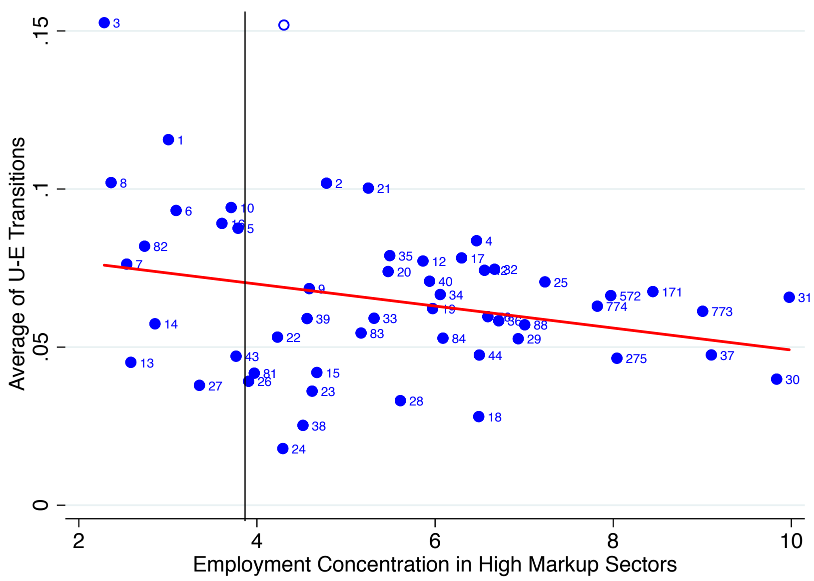
Figure 1: Unemployment and employment concentration in high-markup sectors



Notes: the figure plots unemployment probabilities averaged over the period 2008–14 against employment concentration in high-markup sectors across district municipalities, as calculated using the employment distribution in 2008.

Source: authors' construction based on data from the QLFS.

Figure 2: Unemployment–employment transitions and employment concentration in high-markup sectors



Notes: the figure plots the probability of the transition from unemployment to employment averaged over the period 2008–14 against employment concentration in high-markup sectors across district municipalities, as calculated using the employment distribution in 2008.

Source: authors' construction based on data from the QLFS.

---

itive sectors ($m_d = 0.95$) to the one with the highest ($m_d = 2.88$) is associated with an increase in the probability of being unemployed of 0.77 percentage points, or a 6 percent increase over the mean.

Table 3: Unemployment and employment concentration in high-markup sectors

	Unemployed			
	(1)	(2)	(3)	(4)
Employment concentration	0.003***	0.003**	0.004**	0.004**
in high-markup sectors	(0.001)	(0.001)	(0.002)	(0.002)
Age		0.020***	0.020***	0.020***
		(0.001)	(0.001)	(0.001)
Age (squared)		–0.000***	–0.000***	–0.000***
		(0.000)	(0.000)	(0.000)
High education		0.023**	0.022**	0.022**
		(0.010)	(0.010)	(0.010)
Gender (female)		–0.007*	–0.007*	–0.007*
		(0.004)	(0.004)	(0.004)
Coloured		–0.024**	–0.031***	–0.031***
		(0.009)	(0.006)	(0.006)
Asian/Indian		–0.059***	–0.061***	–0.061***
		(0.018)	(0.021)	(0.021)
White		–0.106***	–0.113***	–0.112***
		(0.006)	(0.007)	(0.007)
Widow/widower		0.020***	0.022***	0.022***
		(0.002)	(0.003)	(0.003)
Divorced or separated		0.021***	0.022***	0.022***
		(0.006)	(0.005)	(0.005)
Never married		0.070***	0.069***	0.069***
		(0.005)	(0.005)	(0.005)
Household size		0.008***	0.008***	0.008***
		(0.001)	(0.001)	(0.001)
Population district municipality (baseline)			–0.000	–0.000
			(0.000)	(0.000)
Share high-educated population (baseline)			0.048	0.043
			(0.143)	(0.145)
Share Black population (baseline)			0.007	0.007
			(0.122)	(0.123)
Share Coloured population (baseline)			–0.008	–0.009
			(0.124)	(0.125)
Share urban population (baseline)			0.066	0.067
			(0.041)	(0.041)
ln(HHI)			–0.029***	–0.029***
			(0.009)	(0.010)
Wave FE	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	No
Province × year FE	No	No	No	Yes
Observations	1,170,150	1,170,150	1,170,150	1,170,150
R^2	0.012	0.058	0.062	0.063

Notes: *p < 0.1; ** p < 0.05; *** p < 0.01. Standard errors in parentheses. Dependent variable is a dummy equal to 1 if an individual is unemployed and 0 otherwise. The measure for employment concentration in high-markup sectors is calculated based on the trimmed sample of OLS markups excluding the 1st and 99th percentiles within each sector. Standard errors clustered at the district municipality level.

Source: authors' construction based on data from the QLFS.

The estimated coefficient remains positive and highly significant when we include the set of individual characteristics (column 2) and the time-varying district municipality characteristics (column 3). Looking at the signs of some of the control variables provides some insights into the South African labour market. The results indicate that the probability of being unemployed is smaller for White individuals (compared to all other groups) and for those who are married. At the same time, unemployment is more prevalent among African Black individuals (compared to all other groups) and those with a large household. Finally, in column 4, we include the full set of province–year fixed effects. This allows us to net out any differential trend in unemployment across provinces. The results do not change: individuals living in district municipalities with higher employment concentration in high-markup sectors are more likely to be unemployed.[7]

Table 4 shows the regression results when we have as the dependent variable the probability of transitioning from unemployment to employment—that is, a dummy that takes value 1 if the individual is unemployed at time $t-1$ and becomes employed at time t, and 0 otherwise. The sample is restricted to only individuals that are unemployed at time $t-1$. The estimated coefficient for m_d—our measures of employment concentration in high-markup sectors at the district municipality level—is always negative and is significant (close to significant at the level of 11.4 in column 4) in the most demanding specifications (i.e. when we include province fixed effects and the full set of interactions between province fixed effects and time fixed effects). This indicates that unemployed individuals living in district municipalities with higher employment concentration in high-markup sectors are less likely to transition into employment in the next quarter. Unemployment–employment transition probabilities are lower for females, for single individuals, and those with larger households, and for those living in district municipalities with a larger population, while they are higher for older people (coefficient for age is positive, but that for age squared is negative) and for those with high education.

4.1 Individual-level heterogeneity

We also look at heterogeneity in the relationship between employment concentration in high-markup sectors and labour market outcomes along a set of individual characteristics. To begin, we test whether the effect of m_d on unemployment is heterogeneous across race groups and varies with the gender, age, and education level of the individual. Results are reported in Table 5. As shown in column 1, the positive association between living in a district characterized by high employment concentration in high-markup sectors and being unemployed is much smaller for White individuals. While there is no differential effect by gender, the education level makes a difference: the impact of m_d is smaller for high-educated workers (column 3). Moreover, the association between living in a district municipality with high employment concentration in high-markup sectors and unemployment probability is stronger for young individuals (column 4).[8] As a robustness check, Table A4 in the Appendix shows the results corresponding to Table 5 on the sample excluding individuals that are not in the labour force. If anything, the magnitudes of our main coefficients increase, consistent with our main interpretation.

Next, we look at the possible heterogeneous effect of m_d on unemployment–employment transitions by race, gender, age, and education level of the individual. Results are reported in Table 6. As shown in column 3, being highly educated increases the probability of exiting unemployment in district municipalities where the employment concentration in high-markup sectors is higher. Again, there is no differential effect by gender (column 2).[9]

---

[7] The magnitude and significance remain similar for our main variable of interest when we exclude individuals that are not in the labour force from our regression, as shown in Table A3 in the Appendix. Note that the coefficient for high education turns to significantly negative, while that for being female turns to significantly positive, indicating that the probability of being unemployed is larger for women and for low-education individuals.

[8] We define young individuals as people who are between 15 and 24 (inclusive) years of age (see International Labor Organization: www.ilo.org/ilostat-files/Documents/description_LFPR_EN.pdf).

[9] Note that Tables 4 and 6 are unchanged whether or not we include individuals who are not in the labour force, since the unemployment–employment transition variables remain the same.

Table 4: Unemployment–employment transitions and employment concentration in high-markup sectors

	Unemployment–employment transition			
	(1)	(2)	(3)	(4)
Employment concentration	–0.001**	–0.001**	–0.001*	–0.001
in high-markup sectors	(0.001)	(0.001)	(0.001)	(0.001)
Age		0.008***	0.008***	0.008***
		(0.000)	(0.000)	(0.000)
Age (squared)		–0.000***	–0.000***	–0.000***
		(0.000)	(0.000)	(0.000)
High education		0.003*	0.003*	0.003*
		(0.001)	(0.001)	(0.001)
Gender (female)		–0.022***	–0.022***	–0.022***
		(0.001)	(0.001)	(0.001)
Coloured		0.006	0.001	0.000
		(0.004)	(0.003)	(0.003)
Asian/Indian		0.009*	0.003	0.002
		(0.005)	(0.004)	(0.004)
White		–0.001	–0.004	–0.004
		(0.007)	(0.007)	(0.007)
Widow/widower		0.005*	0.006*	0.006*
		(0.003)	(0.003)	(0.003)
Divorced or separated		–0.003	–0.002	–0.002
		(0.003)	(0.003)	(0.003)
Never married		–0.006***	–0.005***	–0.005***
		(0.002)	(0.002)	(0.002)
Household size		–0.003***	–0.003***	–0.003***
		(0.000)	(0.000)	(0.000)
Population district municipality (baseline)			–0.000**	–0.000**
			(0.000)	(0.000)
Share high-educated population (baseline)			–0.026	–0.028
			(0.075)	(0.076)
Share Black population (baseline)			–0.101	–0.102
			(0.062)	(0.062)
Share Coloured population (baseline)			–0.058	–0.058
			(0.056)	(0.057)
Share urban population (baseline)			–0.002	–0.002
			(0.018)	(0.019)
ln(HHI)			0.017***	0.018***
			(0.004)	(0.004)
Wave FE	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	No
Province × year FE	No	No	No	Yes
Observations	222,783	222,783	222,783	222,783
R^2	0.005	0.013	0.014	0.015

Notes: **p* < 0.1; ** p < 0.05; *** *p* < 0.01. Standard errors in parentheses. The sample is restricted to all individuals who are unemployed in the previous wave. The dependent variable is a dummy equal to 1 if the individual transitions from unemployment to employment, and 0 otherwise. The measure for employment concentration in high-markup sectors is calculated based on the trimmed sample of OLS markups excluding the 1st and 99th percentiles within each sector. Standard errors clustered at the district municipality level.

Source: authors' construction based on data from the QLFS.

Table 5: Unemployment and employment concentration in high-markup sectors: individual-level heterogeneity

	Unemployed			
	(1)	(2)	(3)	(4)
Employment concentration in high-markup sectors	0.004**	0.004**	0.005***	0.003**
	(0.002)	(0.002)	(0.002)	(0.002)
× Coloured	–0.001			
	(0.001)			
× Asian/Indian	–0.004*			
	(0.002)			
× White	–0.003***			
	(0.001)			
× Female		0.001		
		(0.001)		
× High education			–0.004***	
			(0.001)	
× Young				0.002***
				(0.001)
Individual controls	Yes	Yes	Yes	Yes
District mun. controls	Yes	Yes	Yes	Yes
Wave FE	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	No
Province × year FE	No	No	No	Yes
Observations	1,170,150	1,170,150	1,170,150	1,170,150
R^2	0.063	0.063	0.064	0.046

Notes: *$p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard errors in parentheses. The dependent variable is a dummy equal to 1 if the individual is unemployed and 0 otherwise. The measure for employment concentration in high-markup sectors is calculated based on the trimmed sample of OLS markups excluding the 1st and 99th percentiles within each sector. Standard errors clustered at the district municipality level.

Source: authors' construction based on data from the QLFS.

Table 6: Unemployment–employment transition and employment concentration in high-markup sectors: individual-level heterogeneity

	Unemployment–employment transition			
	(1)	(2)	(3)	(4)
Employment concentration	–0.001	–0.002*	–0.002*	–0.002*
in high-markup sectors	(0.001)	(0.001)	(0.001)	(0.001)
× Coloured	–0.002**			
	(0.001)			
× Asian/Indian	–0.001			
	(0.001)			
× White	0.002			
	(0.001)			
× Female		0.000**		
		(0.000)		
× High education			0.001***	
			(0.000)	
× Young				0.001**
				(0.000)
Individual controls	Yes	Yes	Yes	Yes
District mun. controls	Yes	Yes	Yes	Yes
Wave FE	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	No
Province × year FE	No	No	No	Yes
Observations	222,783	222,783	222,783	222,783
R^2	0.015	0.015	0.015	0.013

Notes: *$p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard errors in parentheses. The sample is restricted to all individuals who are unemployed in the previous wave. The dependent variable is a dummy equal to 1 if the individual transitions from unemployment to employment, and 0 otherwise. The measure for employment concentration in high-markup sectors is calculated based on the trimmed sample of OLS markups excluding the 1st and 99th percentiles within each sector. Standard errors clustered at the district municipality level.

Source: authors' construction based on data from the QLFS.

5 Conclusions

This paper investigates whether individual labour market outcomes differ systematically depending on the level of competition in the product market. To this end, we combine individual-level data from the South African labour force survey with firm-level data from tax records. We derive a measure of employment concentration in high-markup sectors across South African district municipalities, and investigate its relationship with individual unemployment status and the likelihood of transitioning into employment.

We find that higher employment concentration in less competitive sectors is associated with higher unemployment and lower likelihood of transitions from unemployment to employment. This is differentially more the case for non-White, poorly educated, and young individuals. The relationship remains strong when conditioning on a rich set of individual- and district-level covariates, including employment concentration per se. Our findings suggest that lack of competition in the product market has implications for the labour market unemployment stock and flows, opening the way to further studies on the role that (lack of) competition can have in explaining the high and persistent unemployment rate in South Africa.

References

Autor, D., D. Dorn, L.F. Katz, C. Patterson, and J.V. Reenen (2017). ‘The Fall of the Labor Share and the Rise of Superstar Firms’. Working Paper 23396. Cambridge, MA: NBER.

Azar, J., I. Marinescu, and M.I. Steinbaum (2017). ‘Labor Market Concentration’. Working Paper 24147. Cambridge, MA: NBER.

Banerjee, A., S. Galiani, J. Levinsohn, Z. McLaren, and I. Woolard (2008). ‘Why Has Unemployment Risen in the New South Africa?’ *The Economics of Transition*, 16(4): 715–40.

Barkai, S. (2017). ‘Declining Labor and Capital Shares’. Mimeo.

Benmelech, E., N.K. Bergman, and H. Kim (2018). ‘Strong Employers and Weak Employees: How Does Employer Concentration Affect Wages?’ Working Paper 18-15. Washington, DC: Center for Economic Studies, US Census Bureau.

Berry, S., J. Levinsohn, and A. Pakes (1995). ‘Automobile Prices in Market Equilibrium’. *Econometrica*, 63(4): 841–90.

Bresnahan, T. (1982). ‘The Oligopoly Solution Concept Is Identified’. *Economics Letters*, 10(1–2): 87–92.

De Loecker, J., J. Eeckhout, and G. Unger (2018). ‘The Rise of Market Power and the Macroeconomic Implications’. Working Paper 23687. Cambridge, MA: NBER.

De Loecker, J., P.K. Goldberg, A.K. Khandelwal, and N. Pavcnik (2016). ‘Prices, Markups, and Trade Reform’. *Econometrica*, 84: 445–510.

De Loecker, J., and F. Warzynski (2012). ‘Markups and Firm-Level Export Status’. *American Economic Review*, 102(6): 2437–71.

Hall, R. (1988). ‘The Relation Between Price and Marginal Cost in U.S. Industry’. *Journal of Political Economy*, 96(5): 921–47.

IMF (2011). ‘South Africa’s Unemployment Puzzle’. Technical Report. Washington, DC: International Monetary Fund.

Levinsohn, J., and A. Petrin (2003). ‘Estimating Production Functions Using Inputs to Control for Unobservables’. *Review of Economic Studies*, 70(2): 317–41.

Magruder, J.R. (2012). ‘High Unemployment Yet Few Small Firms: The Role of Centralized Bargaining in South Africa’. *American Economic Journal: Applied Economics*, 4(3): 138–66.

Olley, G.S., and A. Pakes (1996). ‘The Dynamics of Productivity in the Telecommunications Equipment Industry’. *Econometrica*, 64(6): 1263–97.

World Bank (2017). ‘World Bank Data: South Africa’. Technical Report. Washington, DC: World Bank.

Appendix

Additional tables

Table A1: Sample firm distribution by broad industry

Industry	Description	2009	2010	2011	2012	2013	2014
11	Agriculture, hunting, and related services	4,249	6,062	7,159	7,558	7,508	6,998
12	Forestry, logging, and related service	314	469	495	513	484	459
13	Fishing, operation of fish hatcheries	229	324	338	357	336	318
21	Mining of coal and lignite	3,342	5,623	5,875	5,849	5,540	5,098
23	Mining of gold and uranium ore	65	121	126	127	116	114
24	Mining of metal ores, except gold and uranium ore	1,104	1,846	1,932	1,907	1,791	1,741
25	Other mining and quarrying	970	1,641	1,702	1,688	1,596	1,463
30	Manufacture of food products, beverages	1,218	2,061	2,151	2,121	2,008	1,841
31	Manufacture of textiles, clothing	3,198	5,309	5,531	5,523	5,312	4,961
32	Manufacture of wood and products	2,558	3,889	4,078	4,145	4,062	3,893
33	Manufacture of coke, refined petroleum	6,129	9,617	10,165	10,266	9,897	9,451
34	Manufacture of other non-metallic mineral products	562	811	914	917	894	844
35	Manufacture of basic metals, fabricate	10,389	16,890	17,707	17,801	17,123	16,263
36	Manufacture of electrical machinery	1,100	1,948	1,982	1,960	1,833	1,673
37	Manufacture of radios, televisions	4,998	8,114	8,096	8,074	7,706	7,252
38	Manufacture of transport equipment	821	1,332	1,400	1,416	1,351	1,265
39	Manufacture of furniture; manufacturing	1,307	1,996	2,084	2,136	2,088	2,039
42	Collection, purification, and distribution	4,370	6,780	6,936	6,831	6,491	6,000
50	Construction	2,579	4,450	4,630	4,657	4,511	4,298
61	Wholesale and commission trade	302	392	425	441	426	396
62	Retail trade, except motor vehicles	1,857	2,821	2,927	2,931	2,828	2,685
63	Sale, maintenance, and repair of motor vehicles	116	185	189	187	172	162
64	Hotels and restaurants	2,876	4,444	4,637	4,645	4,511	4,250
71	Land transport; transport via pipeline	136	210	223	223	228	229
75	Post and telecommunications	140	215	223	228	216	199
81	Financial intermediation	1,035	2,065	2,075	2,093	1,979	1,854
82	Insurance and pension funding	22,567	36,219	38,453	39,019	37,330	34,937
85	Renting of machinery and equipment	3,362	5,071	5,445	5,528	5,371	5,192
86	Computer and related activities	645	951	1,065	1,121	1,094	1,059
87	Research and development	1,058	1,615	1,716	1,737	1,693	1,557
88	Other business activities	7,129	10,925	11,465	11,548	10,977	10,107
91	Public administration and defence act	279	653	697	704	698	674
92	Education	919	1,250	1,410	1,457	1,408	1,304
93	Health and social work	2,149	5,076	5,597	5,782	5,565	5,273
94	Other community, social, and personal	248	424	453	473	447	417
95	Activities of membership organization	27	52	50	58	54	53
96	Recreational, cultural, and sporting	38	86	105	106	102	98
99	Other service activities	1,053	1,845	1,940	1,960	1,885	1,774

Notes: the table reports the industry distributions for the firms used in the production estimation of the SARS-NT data from 2009 to 2014. The IO code is based on the BEA industry classification at the sector level. Industry classifications come from the SIC used in the QLFS.

Source: authors' compilation.

Table A2: Estimated output elasticities by sector

		OLS			OP		
Industry	Description	Labour	Capital	RTS	Labour	Capital	RTS
11	Agriculture, hunting, and related services	0.54	0.22	0.76	0.43	0.33	0.76
21	Mining of coal and lignite	0.97	0.07	1.04	0.79	0.23	1.02
23	Mining of gold and uranium ore	0.64	0.08	0.72	0.55	0.01	0.56
24	Mining of metal ores, except gold and uranium ore	0.64	0.14	0.78	0.50	0.29	0.79
25	Other mining and quarrying	0.65	0.11	0.76	0.62	0.26	0.88
30	Manufacture of food products, beverages	0.77	0.11	0.88	0.63	0.07	0.70
31	Manufacture of textiles, clothing	0.74	0.13	0.87	0.59	0.10	0.69
32	Manufacture of wood and products	0.68	0.14	0.82	0.56	0.22	0.78
33	Manufacture of coke, refined petroleum	0.78	0.09	0.87	0.66	0.07	0.73
34	Manufacture of other non-metallic mineral products	0.81	0.07	0.88	0.71	0.18	0.89
35	Manufacture of basic metals, fabricate	0.84	0.08	0.92	0.75	0.13	0.88
36	Manufacture of electrical machinery	0.71	0.07	0.78	0.63	0.23	0.86
37	Manufacture of radios, televisions	0.81	0.09	0.90	0.74	0.04	0.78
38	Manufacture of transport equipment	0.58	0.12	0.70	0.51	0.17	0.68
39	Manufacture of furniture; manufacturing	0.69	0.08	0.77	0.60	0.07	0.67
42	Collection, purification, and distribution	0.76	0.11	0.87	0.40	0.44	0.84
50	Construction	0.81	0.05	0.86	0.74	0.02	0.76
61	Wholesale and commission trade	0.79	0.11	0.9	0.69	0.05	0.74
62	Retail trade, except of motor vehicle	0.61	0.11	0.72	0.55	0.08	0.63
63	Sale, maintenance, and repair of motor vehicles	0.97	0.02	0.99	1.15	0.02	1.17
64	Hotels and restaurants	0.71	0.13	0.84	0.62	0.11	0.73
71	Land transport; transport via pipeline	0.61	0.13	0.74	0.56	0.67	1.23
75	Post and telecommunications	0.61	0.26	0.87	0.79	0.23	1.02
81	Financial intermediation	0.63	0.08	0.71	0.36	0.31	0.67
82	Insurance and pension funding	0.67	0.07	0.74	0.58	0.06	0.64
85	Renting of machinery and equipment	0.58	0.1	0.68	0.60	0.12	0.72
86	Computers and related activities	0.55	0.07	0.62	0.46	0.11	0.57

Notes: the table reports the estimated output elasticities for labour, capital, and returns to scale (RTS) from the production estimation of the SARS-NT data from 2009 to 2014. The first panel shows the results from OLS regressions. The second panel shows the results from the Olley–Pakes production function estimation with correction of simultaneous bias. The IO code is based on the BEA industry classification at the sector level. Industry classifications come from the SIC used in the QLFS.
Source: authors' compilation.

Table A3: Unemployment and employment concentration in high-markup sectors (excluding not in labour force)

	Unemployed			
	(1)	(2)	(3)	(4)
Employment concentration	0.004**	0.005***	0.007***	0.007***
in high-markup sectors	(0.002)	(0.002)	(0.002)	(0.002)
Age		–0.021***	–0.021***	–0.021***
		(0.002)	(0.002)	(0.002)
Age (squared)		0.000***	0.000***	0.000***
		(0.000)	(0.000)	(0.000)
High education		–0.042***	–0.043***	–0.043***
		(0.013)	(0.012)	(0.012)
Gender (female)		0.023***	0.023***	0.023***
		(0.006)	(0.006)	(0.006)
Coloured		–0.048***	–0.049***	–0.049***
		(0.012)	(0.009)	(0.009)
Asian/Indian		–0.075***	–0.068**	–0.069**
		(0.025)	(0.026)	(0.026)
White		–0.120***	–0.124***	–0.123***
		(0.007)	(0.008)	(0.008)
Widow/widower		0.001	0.003	0.003
		(0.004)	(0.005)	(0.005)
Divorced or separated		0.031***	0.033***	0.032***
		(0.008)	(0.007)	(0.007)
Never married		0.076***	0.075***	0.075***
		(0.006)	(0.006)	(0.006)
Household size		0.017***	0.017***	0.017***
		(0.002)	(0.002)	(0.002)
Population district municipality (baseline)			–0.000	–0.000
			(0.000)	(0.000)
Share high-educated population (baseline)			0.185	0.174
			(0.263)	(0.267)
Share Black population (baseline)			0.120	0.120
			(0.207)	(0.208)
Share Coloured population (baseline)			0.047	0.045
			(0.217)	(0.219)
Share urban population (baseline)			0.068	0.070
			(0.067)	(0.067)
ln(HHI)			–0.058***	–0.057***
			(0.017)	(0.018)
Wave FE	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	No
Province × year FE	No	No	No	Yes
Observations	697,434	697,434	697,434	697,434
R^2	0.010	0.106	0.111	0.112

Notes: **p* < 0.1; ** p < 0.05; *** *p* < 0.01. Standard errors in parentheses. The dependent variable is a dummy equal to 1 if an individual in the labour force is unemployed and 0 otherwise. The measure for employment concentration in high-markup sectors is calculated based on the trimmed sample of OLS markups excluding the 1st and 99th percentiles within each sector. Standard errors clustered at the district municipality level.

Source: authors' compilation.

Table A4: Unemployment and employment concentration in high-markup sectors (excluding not in labour force): individual-level heterogeneity

	Unemployed			
	(1)	(2)	(3)	(4)
Employment concentration	0.007***	0.006**	0.009***	0.006**
in high-markup sectors	(0.002)	(0.002)	(0.003)	(0.002)
× Coloured	–0.000			
	(0.002)			
× Asian/Indian	–0.003			
	(0.002)			
× White	–0.004***			
	(0.001)			
× Female		0.001		
		(0.001)		
× High education			–0.006***	
			(0.002)	
× Young				0.006***
				(0.001)
Individual controls	Yes	Yes	Yes	Yes
District mun. controls	Yes	Yes	Yes	Yes
Wave FE	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	No
Province × year FE	No	No	No	Yes
Observations	697,434	697,434	697,434	697,434
R^2	0.113	0.112	0.114	0.101

Notes: *p < 0.1; ** p < 0.05; *** p < 0.01. Standard errors in parentheses. The dependent variable is a dummy equal to 1 if the individual is unemployed and 0 otherwise. The measure for employment concentration in high-markup sectors is calculated based on the trimmed sample of OLS markups excluding the 1st and 99th percentiles within each sector. Standard errors clustered at the district municipality level.

Source: authors' compilation.

Markup estimation

The estimation of markup follows the same method as in De Loecker and Warzynski (2012). Start with a production technology:

$$Q_{it}(\mathbf{V}_{it};K_{it};\Omega_{it}) = F_{it}(\mathbf{V}_{it};K_{it};)\Omega_{it}$$

The associated Lagrangian function (with one composite input) is:

$$L(V_{it};K_{it};\Omega_{it}) = P_{it}^{V}V_{it} + r_{it}K_{it} - \lambda_{it}(Q_{it}(.) - Q_{it})$$

First-order condition with respect to the variable input V gives

$$\frac{\partial Q_{it}(.)}{\partial V_{it}}\frac{V_{it}}{Q_{it}} \equiv \theta_{it}^{V} = \frac{1}{\lambda_{it}}\frac{P_{it}^{V}V_{it}}{Q_{it}}$$

where θ_{it}^{V} is the output elasticity of the variable input V and the Lagrangian multiplier λ_{it} is a measure of marginal cost. Rearranging the terms we have that the markup μ_{it} is defined as price over marginal cost:

$$\mu_{it} \equiv \frac{P}{\lambda} = \theta_{it}^{V}\frac{P_{it}Q_{it}}{P_{it}^{V}V_{it}}$$

where the output elasticity θ_{it}^{V} is estimated from the production function estimation and $\frac{P_{it}Q_{it}}{P_{it}^{V}V_{it}}$ is the inverse of input revenue share of the variable input V, which could be directly calculated from firm-level accounting data.