Doctoral Thesis

University of Trento

School of Social Sciences

Doctoral School in Economics and Management / Sociology and Social Research / Local Development and Global Dynamics

Social capital and the labour market: Essays on trust, inequality and employment

Sara Tonini

November 2017

Acknowledgements

To Charlotte.

It always seems impossible until it's done

- N. Mandela -

Contents

Chapter 1

Introduction

According to the 2017 World Economic Forum, the factors that pose a serious risk to today's global economy are rising inequality and the polarization of societies, which in turn threat the social cohesion[1]. In particular, This doctoral dissertation contributes to the understanding of these major current challenges, by investigating the extend of unequal access to opportunity in education and in the labour market in the former communist countries; the potential of diversity in the South African multicultural society in terms of employment; the formation of interpersonal trust at the individual level in Germany.

Chapter 2 provides estimates for inequality of opportunity in post-communist countries in terms of individual labour income, employment status, quality of jobs, and educational attainment. Inequality might not always be considered undesirable. In fact, differential achievements reflecting different level of effort and/or freely made choices are deemed fair. Equity in accessing opportunities, rather than equality in the distribution of outcomes, has thus become the goal of policy-makers and the focus of a lively debate about inequality and social welfare among researchers. Us-

---

[1]The Global Risks Report 2017, 12th Edition published by the World Economic Forum available online at: http://reports.weforum.org/global-risks-2017/

ing data from the third round of the Life in Transition Survey (2016), we show that inequality of opportunity for earning an income is higher in the transition region than in Western European countries. Parental background is the key circumstance to explain unequal access to opportunity, both in the labour market and in education. As for employment opportunities, access to good-quality jobs is more unequal than access to any kind of job, on average. In terms of the evolution of inequality of opportunity over time, we show that individuals who started their education right after the fall of the Berlin wall are confronted with higher level of inequity in accessing tertiary education. This chapter is the result of a joint project with Michelle Brock, from the European Bank for Reconstruction and Development and Vito Peragine, from the University of Bari.

Chapter 3 investigates how within-black ethnic diversity affects labour market outcomes of the black South Africans in post-Apartheid South Africa. Results suggest that ethnic diversity has a positive effect on employment rate among the black South Africans, and it mostly affects ethnic groups with relatively larger population size and people who are less educated. To address the endogeneity of ethnic composition, we exploit the historical origins of the location of blacks' homelands and argue that districts more equidistant to multiple homelands are more diverse. We then propose a model of a coordination game to explain these findings, implying that as inter-ethnic communication requires more skills and efforts than intra-ethnic connection, ethnically diverse districts motivate people to invest more in social skills in response to higher rate of inter-ethnic interaction. The acquisition of these social skills can make them better equipped for the labour market. We show that our mechanism is related to, yet distinct in important aspects from current models on inter-group interaction. This chapter has been produced together with Peng Zhang, from the University of Cambridge.

In chapter 4, we aim at understanding the formation of trust at the individual

level, given the impact that it has been recognized to have on economic development. Theoretical work highlights the role of the transmission of values such as trust from parents to their children. Attempts to empirically measure the strength of this transmission relied so far on the cross-sectional regression of the trust of children on the contemporaneous trust of their parents. We introduce a new identification strategy which hinges on a panel of parents and their children drawn from the German Socio-Economic Panel. Our results show that a half to two thirds of the observed variability of trust is pure noise irrelevant to the transmission process; this noise strongly biases the parameter estimates of the OLS regression of children's trust on parents' trust. However an instrumental variable procedure straightforwardly emerges from the analysis; the dynamics of the component of trust relevant to the transmission process shed light on the structural interpretation of the parameters of this regression; the strength of the flow of trust that parents pass to their children as well as of the sibling correlations due to other factors are easily summarized by the conventional R^2 of a latent equation. In our sample, approximately one fourth of the variability of children's trust is inherited from their parents while two thirds are attributable to the residual sibling correlation. This chapter is a joint contribution together with Corrado Giulietti, from Southampton University and Enrico Rettore, from University of Trento.

Throughout this work, the concepts of transmission and persistence, both across generations and more broadly of historical patterns, has an important role to play. In particular, the analysis of inequality of opportunity (chapter 2) is primarily concerned with the estimation of the degree to which family background, together with other predetermined personal characteristics determine a person s educational and labour market outcomes. Similarly, when studying the formation of interpersonal trust at the individual level (chapter 3), we are interested in understanding (and estimating) how much of this personal trait (or value) is inherited from parents. In investigating the effect of ethnic diversity on labour market outcomes in South

Africa (chapter 4), we could not overlook the persistent pattern of cultural medley resulted from centuries of historical events.

Chapter 2

On Fairness: Evidence from Post-Communist Countries

2.1   Introduction

In recent years, inequality of opportunity has attracted considerable interest from researchers and policy makers[1]. Evidence from existing surveys shows that people consider income inequalities arising from exogenous circumstances less acceptable than those resulting from individual choices and effort. In other words, what seems to matter for a just society is the access to opportunities, rather than the distribution of outcomes. Alesina and La Ferrara (2005) argue that the origin of inequality, more than its level, affects social attitudes towards redistributive policies. Moreover, recent studies investigating the relation between inequality and aggregate economic performances find that unequal access to opportunities, more than income inequality, has a negative impact on economic growth (Marrero and Rodríguez, 2013; Ferreira et al., 2014; Murphy and Topel, 2016).

---

[1]See Roemer and Trannoy (2015); Ramos et al. (2016); Ferreira and Peragine (2016) for a comprehensive review of the literature on inequality of opportunity, both from a theoretical and a methodological viewpoint.

The issue of inequality, especially of its fairness, is particularly relevant in Eastern Europe and Central Asia. In fact, after experiencing decades of communism, with its egalitarianism, the countries of the region have undergone major economic and social changes. As they transitioned from a planned to a market economy, those countries faced processes of privatization and trade liberalization, and the dynamics of wage setting started to follow the law of demand and supply (Mitra and Yemtsov, 2007; Milanovic, 1999; Rose and Viju, 2014). Highly qualified workers could enjoy significant earning premiums when employed in highly skilled occupations. Additionally, de-industrialization (Ivaschenko, 2002) and reforms in the social and tax systems (Flemming and Micklewright, 2000) have been serious challenges for the transitional governments. The result of this restructuring was an increase in income inequality (Milanovic, 1998).

Despite the abundance of studies investigating the degree of total inequality, its roots and its socio-economic repercussion in the post-communist countries[2], estimates of inequality of opportunity for the region are scarce and often limited to few countries that joined the European Union, for which data is available. Above all, the empirical literature lacks comparable estimates for the whole Eastern Europe and Central Asia region. We contribute to the literature by filling this gap, and provide estimates for inequality of opportunity in 29 post-communist countries of the region and 5 comparator countries in Western Europe. To calculate how unequal is the access to opportunities in the different countries of our sample, we use data from the third round of the Life in Transition Survey (2016). Importantly, we do not focus exclusively on inequality in earning an income, rather we estimate inequality of opportunity for getting a job, distinguishing between any kind of job and a "decent" one, and for attaining tertiary education. The objective is twofold. Firstly, unequal access to income might be, at least partially, the result of inequality

---

[2]See Perugini and Pompei (2015) for a comprehensive review of the changes in inequality in Central and Eastern European countries, during and after the transition from socialism

in accessing opportunities at different stages of an individual's life (unequal access to good-quality job or, earlier, to education). Furthermore, when focusing on income, unemployed and inactive individuals are generally disregarded. By analyzing inequality of opportunity for getting higher education and a job, we propose an alternative approach to the imputation of income (Checchi et al., 2015) for taking into account non-working individuals.

The chapter is structured as follows. In Section 2 we provide a conceptual framework for our analysis and discuss the methodology used. We describe the data employed in Section 3, and present the results for the different outcomes in Section 4. Finally, we draw some conclusions in Section 5.

2.2   Conceptual Framework and Methodology

Conceptually, the study of (in)equality of opportunity requires the distinction between the effects on individual outcomes of factors for which the subject is not held responsible - called circumstances - and of effort (or any other factor over which the individual has control) (Roemer, 1993, 1998). While circumstances are generally measured by aspects of an individual's childhood and family environment at birth[3], his/her effort is empirically proxied by years of schooling or productivity at work. Rooted in a Rawlsian philosophical tradition, inequality of opportunity theories support the idea that only inequality arising from different levels of effort can be considered ethically acceptable. In other words, inequality due to differential access to opportunity depending on circumstances is considered unfair. Therefore, the aim of egalitarian policies is to compensate individuals who face inequities (*compensation principle*), without impeding outcomes to vary in response to effort (*reward*

---

[3]In their recent contribution, Hufe, Peichl, Roemer, and Ungerer (Hufe et al.) argue that all achievements and decisions taken by the child before the age of consent can be considered circumstances.

principle).

In accordance with this conceptual framework, empirical studies have developed two main approaches to measuring inequality of opportunity. In the *ex-ante* approach, individuals with the same set of circumstances are grouped together to constitute different types. Inequality of opportunity is then calculated as the inequality between types. The differences in outcomes after circumstances have been taken into account are assumed to be due to effort. Clearly, the residual category of "effort" includes all of the factors which affect individual outcomes other than the specific set of circumstances considered (i.e. luck, ability, innate talent, other unobservable circumstances)[4]. By contrast, the *ex-post* approach focuses on the inequality among subjects who have exerted the same level of effort. Since it is plausible to assume that circumstances themselves affect the level of effort, Roemer (1993) suggested to rank an individual in the effort distribution of his/her own type. The population is thus divided into tranches, which group together individuals at the same percentile of the type-specific distribution. Given the considerable difficulties in measuring effort, empirical contributions have normally rely on the *ex-ante* perspective to estimate inequality of opportunity (Bourguignon et al., 2007; Checchi et al., 2010; Lefranc et al., 2008, 2009; Marrero and Rodríguez, 2012; Brunori et al., 2013). In line with the literature, we follow the same approach to analyze inequality of opportunity in the transition region.

Formally, consider an outcome y and a vector of circumstance variables, C. Effort is measured by a scalar variable, e. Let suppose that all determinants of y, including luck, can be classified into either the vector C or the scalar index e, such that:

$$y = g(C, e) \qquad (2.1)$$

---

[4]Due to data limitations, the set of circumstances included in any study is only a subset of all possible exogenous factors. Ferreira and Gignoux (2011) and Niehues and Peichl (2014) show how to calculate the lower and upper bounds of inequality of opportunity, respectively.

The population can be partitioned into a set of groups (types, T) that are completely homogeneous in terms of circumstances. Within each subgroup k, individuals differ only in their level of effort. Given the outcome distribution for type k, $F_k(y)$, and its population share, q_k, the overall distribution of the outcome for the population as a whole is $F(y) = \sum q_k F_k(y)$.

In this context, the inequality due to circumstances would be eliminated when all types faced the same opportunity set (or opportunity sets with the same value). Inequality of opportunity can thus be estimated by computing an inequality measure $I(.)$ over the counterfactual distribution where each individual's outcome is replaced by the value of his/her opportunity set, v_i:

$$I(\tilde{y}), \tilde{y} = v_i \qquad (2.2)$$

The opportunity set faced by subjects belonging to a given type k is extracted from the outcome distribution of individuals in the same type. In empirical analysis, the value of the opportunity is very often summarized by its mean, μ_k, so that $v_i = \mu_k$. In other words, estimating inequality of opportunity amounts to measuring the inequality in the counterfactual (smoothed) distribution. The specific index $I(.)$ employed varies across studies and depends on the nature of the outcome of interest (continuous versus binary variable). The most used indexes are the Gini, the mean logarithmic deviation and Theil for continuous variables, while a dissimilarity index is generally utilized when the outcome is dichotomous.

2.2.1 Decomposition of the inequality of opportunity measure

To investigate the relative importance of each circumstance in explaining unequal access to opportunities, a Shapley decomposition has been employed. The Shapley

value has been firstly used in game theory with the aim to provide a rule for dividing a given surplus among members of a coalition and has been extended to inequality analyses by Shorrocks (2013). By applying the Shapley decomposition, we aim at identifying how much the measure of inequality of opportunities would change when we add a circumstance to different pre-existing sets of circumstances.

Formally, the change in the inequality index when circumstance k is added to a subset Z of circumstances is given by:

$$\Delta I_k = \sum_{Z \subset K} \frac{|m|!(\kappa - |m| - 1)!}{\kappa!} [I(Z \cup k) - I(Z)] \tag{2.3}$$

where K indicates the whole set of κ circumstances, and z is a subset of K that includes z circumstances variables except k. $I(Z)$ is the inequality measure for the subset Z and $I(Z \cup k)$ is the index obtained after adding circumstance k to the subset Z.

Let $I(\kappa)$ be the inequality index for the set of κ circumstances. The contribution of circumstance κ to $I(\kappa)$ is defined by:

$$C_k = \frac{\Delta I_k}{I(\kappa)} \tag{2.4}$$

where $\sum_{i \in K} C_i = 1$.

2.3   Data

To estimate inequality of opportunity in the region we use data from the third wave of the Life in Transition Survey (LiTS), conducted jointly by the European Bank for Reconstruction and Development and the World Bank. The survey has been carried out in 2016 in 34 countries, mainly in the Eastern Bloc and Central Asia,

interviewing about 51,000 households[5]. Importantly, the LiTS contained information on individual earnings, in addition to labour market outcomes and educational attainments. Moreover, the survey included a module with detailed questions on the respondent's socio-economic background, which are useful to measure circumstances in empirical applications. In particular, there is information on the educational levels and sector of occupation of the parents and the respondent's place of birth (rural/urban).

In calculating the inequality of opportunity measure, we follow the literature (Marrero and Rodríguez, 2012; Checchi et al., 2015; Brunori et al., 2013) and consider gender, birthplace, ethnic and family background as circumstances affecting individual outcomes irrespective of one's responsibility. More in details, place of birth identifies whether the individual is born in a rural or in a urban area. Ethnicity is summarized in a binary variable which takes value 1 if the respondent claims to be part of a minority group in the country where s/he lives, and 0 otherwise. Clearly, being part of a minority may represent very different concepts, depending on how discriminated the group is. For example, Roma people in Czech Republic do not face the same challenges as Russian people in Kyrgyzstan. As for the family background, we use the parental educational attainments, measured on a four-point scale, and whether at least one of the parents has been a member of the Communist Party. For the estimation of inequality of opportunity for education, we also use the number of books at home during childhood to capture more broadly the parental human capital. The categorical variable is coded in five categories: a) 0 to 10 books; b) 11 to 25 books; c) 26 to 100 books; d) 101 to 200 books; e) more than 200 books. We do not include the father's and mother's sector of occupation since

---

[5]The previous waves of the survey have already been used to analyze inequality of opportunity in the region. In particular, Abras et al. (2013) estimated the degree of inequality of opportunity in labor market outcomes for a number of countries in Eastern Europe and Central Asia, using data from the Life in Transition Surveys conducted in 2006. Based on the second round of the Life in Transition Survey (2010), Brock et al. (2013) estimated the inequality of opportunity with regard to household assets.

it might suffer from important recall bias (a substantial proportion of answers is missing) and it may not be very informative of the type of job (and tasks).

We apply different sample selection rules, according to the specific outcome under study. Details will be given in each of the dedicated subsection of Section 4. However, we anticipate that only working age individuals (18-64 years old) are considered when focusing on labour market outcomes (earnings and employment status), while we do not impose any restriction for educational outcomes.

2.4 Results

2.4.1 Inequality of opportunity for earnings

Following a number of empirical contributions on inequality of opportunity, we begin by focusing on the acquisition of individual labor income. The main reasons for choosing individual earnings over household income are twofold: one has to do with the quality of the available data, and the other is a more conceptual motive. Firstly, answers to the question on household income are not always reliable: in a number of cases, the reported monthly household income is lower than the monthly individual earnings. Conceptually, household (per capita) income conveys a bunch of other elements unrelated to the individuals' effort, luck and opportunities in the acquisition of their labour income, such as assortative mating, fertility decisions and non-labor income sources (Bourguignon et al., 2007).

The non-response issue is common to the questions on both individual and household income and the relative bias cannot be disregarded. As reported in Table 2.1 (Column 4), the non-response rate varies considerably among countries. In order to investigate the potential differences between the group of the respondents and the group of the non-respondents in terms of individual labour income, we estimate the

predicted earnings based on a number of observable characteristics[6]. Despite being selective, non-response does not lead, in general, to economically worrisome differences in the predicted income distribution for the two groups (Table 2.2 and Table 2.3). Thus, we are confident that the estimates based on the sub-sample of the respondents are representative of the whole population of interest[7]. Despite some significant differences, the comparison between the inequality of the income distribution from our data with official sources displays a positive correlation[8] (Figure 2.1).

The analysis uses individual self-reported earnings over the past 12 months, which may come from formal or informal, permanent or seasonal employment[9]. We do not include self-employed and employers in our estimates, since their self-reported labor income is more susceptible of measurement error. The universe examined consists of all employees aged 18-64.

In general, inequality of opportunity for earning an income[10] is higher in the formerly planned economies than in the Western countries in our sample (i.e. Germany, Italy, Cyprus) and varies considerably across the region (Table 2.5, columns (a)). This result confirms the findings of previous contributions, which have esti-

---

[6]More in details, we regress the logarithm of earnings on the following set of variables, which are relevant in explaining differences in the acquisition of income: birthplace (whether rural or urban), gender, age (and its squared values), self-reported health status (binary variable), own education (binary variable for having attained tertiary education), parental education, parental membership to the Communist Party, ethnicity (whether part of a minority group), and marital status. Then, we predict the (log)earnings of the individuals who declare to be working but choose to not report any labour income

[7]An alternative approach to the problem, which has been followed by Checchi et al. (2015), is to include the non-respondents in the analysis by imputing their income. The estimates for the inequality of opportunity for earning an income and all main results in this section are robust to this alternative specification. Nevertheless, given that the imputation methodology assumes that data are missing at random and in some countries the proportion of the missing information is high, we prefer to rely on the results obtained from the non-missing data only.

[8]Official estimates refer to the latest observation of the Gini index on net income available for each country in the Standardized World Income Inequality Database (Solt, 2016). Note that for some countries, for example Azerbaijan, the latest recorded observation refers to 2010 or earlier.

[9]The values of the labour income variable have been winsorized at the 0.5% level at the top. Due to data unavailability, Albania has been removed from the sample

[10]The value of inequality of opportunity for income is equal to the Gini index calculated on the distribution of the OLS regression's predicted log-income.

mated inequality of opportunity in Europe (Marrero and Rodríguez, 2012; Checchi et al., 2015). In a global perspective, unfairness is much lower than in other emerging economies (i.e. Brazil and India) and in the US. It has to be underlined that estimates of inequality of opportunity are not immediately comparable across studies, though. In fact, methodologies (parametric versus non-parametric), outcomes (individual earnings versus individual disposable income) and circumstances can differ substantially[11]. On average, inequity in acquiring a labour income is lower in the 15 countries that are part of the European Union than in the remaining 18 countries (0.105 versus 0.12), despite the relatively high estimates for Estonia, Greece, Hungary, Latvia and Bulgaria, where inequality of opportunity is above the regional mean value.

In order to be informative of the importance of (un)fairness in accessing opportunities, the absolute level of inequality of opportunity has to be related to the total level of (income) inequality. Firstly, it has to be noted that the relationship between total inequality and inequality of opportunity is strong and positive: countries with high levels of income inequality also have high levels of inequality of opportunity (Figure 2.2). In order to understand how much of the total income inequality can be explained by circumstances we may want to use the relative measure of the inequality of opportunity, i.e. the ratio of (the absolute level of) inequality of opportunity to the Gini coefficient calculated on the distribution of individual earnings. Figure 2.3 shows that in the great majority of countries (75% of the sample), a third or more of the total income inequality is due to individual circumstances. On average, the relative importance of circumstances in explaining total inequality is higher in the EU countries than in the rest of the formerly planned economies (37% versus 34.7%).

---

[11]Many studies use the mean logarithmic deviation, or Theil 0, instead of the Gini index, as an inequality measure. In our sample, the average inequality of opportunity estimated using the Theil 0 equals 0.024, compared with 0.04, 0.082 and 0.22 for the US, India and Brazil, respectively.

The relative contributions that specific circumstances make to overall inequality of opportunity also vary greatly across regions and countries[12]. Figure 2.4 shows the role played by parental background (i.e. parental educational attainments and membership to the Communist Party), gender and other circumstances (i.e. the status of the birthplace and ethnicity) in explaining the "unfair" part of inequality. In both EU and non-EU members states, a large percentage of inequality of opportunity can be traced back to an individual's parental background: nearly 50% in the EU countries and about 40% in the remaining countries, on average. Despite being different from the analyses on intergenerational income mobility, the relative importance of the parental background in explaining inequality of opportunity speaks in favor of some degree of intergenerational persistence. Gender is the second most important factor, explaining between a quarter and half of overall inequality of opportunity in most countries, with a stronger role in EU states. Place of birth and ethnicity contribute only residually to explaining unequal access to income opportunities.

The estimates presented in this section suffer from a number of limitations. As we have discussed at the beginning of the section, non-response can be a serious problem and potential errors in the reported earnings have also to be taken into account (especially in the case of self-employed individuals). In order to strengthen our results, we estimate the inequality of opportunity for having a job as well as for getting a decent one (for more details and the definition of "decent" job, see the following section). This way, we can include in the sample all (working age) individuals, regardless of the availability of information on their income. Even when available and reliable, data on 12-month individual earning raises another problem. In fact, for the estimation to be more precise, the outcome of relevance should be the individual *permanent* income (earnings). In periods during which the

---

[12]We calculate the relative contribution of each of the circumstances by employing the Shapley value decomposition (Shorrocks, 1982; Sastre and Trannoy, 2002).

"job churn" (short-term employment and a number of career changes in a person's life) and the labour market flows are considerable, the observation of earnings at a single given point in time can be very misleading. The lack of panel data does not allow us to construct a more robust income (earnings) variable. Using a binary variable that indicates whether the individual has a job (at the same single given point in time) does not help solve the issue either. To bolster our analysis, we additionally estimate the inequality of opportunity for education, in particular for having some tertiary education. For adults, education can be considered stable over time. By including estimates for education, we also aim at tacking another potential problem: the difference across countries in the labour market participation and in the unemployment rate, which may raise concerns about the comparability of the estimates across countries.

2.4.2 Inequality of opportunity for employment

The empirical literature on the measurement of inequality of opportunity has mainly focused on opportunities for the acquisition of income, but it may be useful to adapt the same framework to the space of employment and education. In fact, in order to better understand the drivers of unequal opportunities to earn income, it is helpful to explore inequality of opportunity at key junctures along the career path. These include getting a job, the "quality" of the job, and obtaining higher education. At each stage, pre-determined circumstances may affect the available opportunities and thus the choices made by individuals.

We consider two definitions of opportunity in the labor market: having a job, and having a *decent* job. The universe examined consists of all adults aged 18-64 in the labour force, i.e. individuals who are working - both as employees and self-employed or employers - or unemployed available for work, and actively looking for a job.[13]

---

[13]More precisely, we include in the labour force all individuals 18-64 years old who: a) work

On average, in our sample, the unemployment rate based on a reference period of a week is 16.5%, with big differences between countries. This figure is higher than the ILO estimated average unemployment rate both for Central Europe and the Baltic region (9.1% in 2014), and for Europe and Central Asia region (9% in 2014) (Figure 2.5).

Another important aspect of the labour market conditions faced by workers is the quality of jobs. The definition of a *decent* job derives from the ILO standards on the quality of employment, taking into account the information that are available in the survey. In particular, the quality of the job is determined on the basis of the contract arrangement and the working time. An individual of our reference population has a decent job if he works under a written contract (has a registered enterprise, in the case of self-employed) and he has enough but not excessive working time (more than 20 hours per week, but no more than the median number of hours worked during a standard week in the country[14]). According to our definition, among those having a job, on average only about 50% have a decent job. Among the countries with a percentage of decent jobs lower than 50%, there are none of the EU member states (Greece being an exception).

Being interested in measuring inequality of opportunity in the labor market, we focus on the inequality deriving uniquely from circumstances, for which the individual cannot be held responsible. Since our outcomes (i.e. having a job and having a decent job) are binary variables, the between groups inequality is summarized into a dissimilarity index, D-index, and is estimated by running a logistic regression model to establish the relationship between access to a particular opportunity and a set of

---

as wage employees; b) have an unpaid job (internships or apprenticeships, or are unpaid workers in the family business); c) are self-employed or employers; d) are unemployed but are looking for a job; e) are unemployed and declare not to look for a job for *temporary* reasons (temporarily sick/injured; waiting for an answer; have already found a job that will commence in the near future).

[14]The ILO Statistical Decent Work Indicators consider excessive working time a working week with more than 48 hours. In order to take into account the differences in the structure of the various labour markets, we adopt a more flexible classification.

pre-determined circumstances. Inequality of opportunity is then measured as the contribution of these circumstances to the inequality index. By construction, the measure of inequality (D-index) can change according to the specific set of variables that are used to define the groups. In particular, the D-index can only increase or remain constant when additional circumstances are added to the initial set of circumstances. In other words, our estimates represent the lower-bound of inequality of opportunity[15]. The circumstances used to measure the inequality of opportunity in the labour market are the same as for the analysis on earnings: whether the place of birth is rural or urban, gender, parental educational attainment and membership to the Communist Party, and ethnicity (belonging to a minority group).

On average, inequality of opportunity for having a *decent* job is 50% higher than inequality of opportunity for having any job (Figure 2.6). Both these measures tend to be lower in countries with lower unemployment rates. This could reflect differences to workers bargaining power. As demand for labour falls, employers can more easily make arbitrary trade-offs between workers without sacrificing productivity. Workers who are then unjustly discriminated against do not have leverage to make demands to employers, because employers can easily replace them. Increases in employment may therefore be one way to decrease inequality of opportunity. Additionally, when inequality of opportunity for having a *decent* job is higher (compared to any job) the difference between male and female labour force participation is also larger (Figure 2.7). This might suggest that women are less likely to participate in the labour force in countries where access to better-quality jobs is constrained by circumstances. In principle, the lower female participation could be the result of discouragement that follows the unsuccessful search for a job that matches their (on average) higher (with respect to men) level of education. Given that our definition

---

[15]See Luongo (2011) for a discussion and Ferreira and Gignoux (2011) for a formal proof of the lower-bound result. The intuition is that the specific set of circumstances upon which we rely to split the population into groups, and thus to estimate inequality, is a subset of all possible circumstances. The existence of unobserved circumstances guarantees that the estimates of inequality of opportunity could not be lower if one were to consider the full set.

of a *decent* job has to do uniquely with the contract arrangements and the working time, this is not the more convincing interpretation. It rather seems to relate to the security of the job. Women, who in many cases still bear the burden of caring for children and the family, may require more stable job, with more flexible working hours.[16]

In general, parental background remains the most important circumstance for explaining inequality of opportunity in the labor market, and it is even more so in the countries that joined the European Union (Figures 2.8 and 2.9). Parents membership to the Communist Party is on average more important for getting a *decent* job than any job (Table 2.6, Panel (a) vs. Panel (b)), suggesting the persistence of networks dating back to pre-transition times (in Western European comparator countries this effect is predictably absent). The evidence on the overwhelming contribution of parental education to the inequality of access to opportunities in the labour market is in line with previous findings by Abras et al. (2013). Although not directly comparable, their results highlight the importance of father's education for explaining inequality of opportunity in having a good job[17].

2.4.3 Inequality of opportunity for education

Equitable access to tertiary education is often seen as a first hurdle that countries must tackle in order to reduce inequality. Since 1980, income gains in the United States have accrued almost exclusively to those with tertiary education. Meanwhile lower-skilled workers have not seen real wage increases since 2003 (Goldin and Katz, 2009; Acemoglu and Autor, 2011). Higher and better education is associated not

---

[16]See Bender et al. (2005) for a discussion on how women and men differently value flexibility at work; and Neyer (2006) for a discussion on the interplay between family, fertility and employment policies in Europe.

[17]Abras et al. (2013) use data from the Life in transition Survey conducted in 2006, focus on a reduced sample of countries and have a different definition of having an opportunity in the labor market, based on the available information contained in the data.

only with higher employability, income and wealth, but also with a better health status, higher civic engagement and higher level of social trust (Cutler and Lleras-Muney, 2008; Campante and Chor, 2012; Easterbrook et al., 2016 Moreover, the benefits of education spread beyond the individual level. Improvements in human capital are beneficial to firms, industries and the economy as a whole Blundell et al., 1999; Prskawetz et al., 2007.

Following the global trend in education, educational attainment levels of the population in the transition region have improved significantly over the last decades (Barro and Lee, 2013).[18] Despite the increase in the percentage of individuals getting higher education and the persistent skills mismatch[19], returns to a tertiary degree are still high in most of the 34 countries, and comparable to what has been found in some Western European countries, such as Spain and Netherlands, but lower than Eastern Europe in the early years of transition (Badescu et al., 2011; Bartolj et al., 2013). Individuals with a tertiary degree earn, on average, 29% higher income than those with a secondary (or lower) degree, and the returns are robust to the inclusion of several controls (Table 2.4).

Given the profound effect of market liberalization on the education systems in former communist countries, it is interesting to study the level of inequities in access to education for two separate cohorts: the younger cohort (those who started education after 1989) and the older cohort. With the transition from plan to market, tertiary education in post-communist countries went from being free, although competitive, to often having a non-trivial cost. Even where education remained nominally free, scholarships to cover the cost of living, generous before transition, have been effectively phased out, resulting in much higher costs of being a student.

---

[18]In 2015, almost 29% of people aged 25-54 had attained a tertiary level of education, compared with 19% of those aged 55-74. According to Eurostat statistics, comparable figures for the EU-28 are 32.6% and 20%, respectively.

[19]In the region, the average percentage of people under the age of 30 who are overqualified for their job has steadily increased over the last decade, rising from 12.5% to 15.1% (for Reconstruction and Development, 2016)

Furthermore, the previously strong and heavily controlled link between education and jobs effectively disappeared, and new skills where rewarded in the labour market (Guriev and Zhuravskaya, 2009).[20] The younger cohort would have faced the option to start tertiary education in the early 2000 s, by when education systems had been reformed and many Central and South-Eastern European countries had stronger prospects of EU membership.

Figure 2.10 shows the change in educational attainments across the two cohorts. For the younger cohort, the percentage of people completing some tertiary education is almost double that of the older cohort and overall schooling inequality, as proxy by the standard deviation of educational attainment, is slightly higher. While the total level of inequality in achievement might reflect differences in effort, we are interested in analyzing how much of that inequality is explained by pre-determined circumstances beyond individual's control. Table 2.5 (columns (d) and (e)) reports the estimates of the inequality of educational opportunity for the two cohorts[21], considering place of birth, parental background, ethnicity and number of books at home during childhood as circumstances, which the individual cannot be held responsible for. The estimates range between 0.15 (Azerbaijan) and 0.39 (Russia) for the older cohort, and between 0.17 (Slovenia) and 0.52 (Bulgaria) for the younger cohort. Figure 2.11 provides the same results graphically. Clearly, inequality of educational opportunity has generally increased: pre-determined circumstances matter more for educational attainment in the young cohort than they do for the older cohort. On the other hand, no clear regional pattern emerges , neither in terms of sub-regions (South-eastern -, Central -, Eastern Europe, Central

---

[20]See Brunello et al. (2010) for a discussion about the changes in the economic and education systems of Eastern Europe after the fall of the Iron Curtain. Interestingly, they find evidence that having obtained primary and/or secondary degrees (but not tertiary) under communism is penalized in the economies of the late 2000s. Confirming their findings, results in Table 2.4 suggest that tertiary education acquired under the communism is not penalized by the radical transformation of the Eastern bloc economies.

[21]The value of inequality of opportunity for tertiary education is equal to the Dissimilarity index calculated on the distribution of the logistic regression's predicted probabilities of having tertiary education.

Asia) nor to the extent of being a member state of the European Union. Slovenia, Poland, Russia and Latvia are the only four countries where the inequity in accessing tertiary education for the younger cohort is lower than for the older ones (Figure 2.11). On average, inequality of educational opportunity for the younger cohort is lower in the countries that joined the European Union during the 2000s, than in the non-EU member states.

The decomposition of inequality of opportunity into partial shares by individual circumstances is presented in Figure 2.12. The shares of the six predetermined circumstances add up to the total inequality levels in Table 2.5 (columns (d) and (e)). The highest portion of inequality of opportunity in education is attributable to family educational and cultural resources. For example, parental education typically explains half of the inequality in educational attainment. A third of the inequality is explained by the availability of books in the house during childhood. Being born in a rural area is also an important determinant of inequality in educational attainment; it accounts for about 10% of the estimated inequality of opportunity, on average. Parental membership in the Communist Party and ethnicity prove to play a minor role across countries. Of particular note is the increased role of parental background for educational attainment in the young cohort. This increased dependence on parental education can be explained by the fact that parents with tertiary education gained more from transition: They have been better positioned to send their kids to universities and bear the associated costs, in terms of both university fees and foregone income of children. In contrast, the importance of parents Communist Party membership has almost halved.

Unlike in the case of income and employment, when it comes to education, women are more likely to obtain a tertiary degree than men. This reverse gender gap is more prevalent in the younger cohort. This evidence may simply reflect the long term legacy of the soviet organization, where men were encouraged to attend

vocational schools supplying the industrial and agricultural sectors, while women were more likely to gain higher education (Terama et al., 2014). It could also be the result of inequality of opportunity in the labour market that induces women to select into jobs that require more education. Autor et al. (2010), focusing on the United States, finds that women with less than a college degree experienced a more dramatic decline than men in the share of their employment in middle-skill occupations, between 1979 and 2007.

2.5 Conclusions

The transition from a planned to a market economy in the post-communist countries was accompanied by expectation of greater and fairly distributed opportunities for all. These expectations have been only partially met. In fact, inequality of opportunities in Eastern Europe and Central Asia is higher than in many Western European countries, being nevertheless restrained in comparison with the US and other developing countries. On average, post-communist European Union member states display lower levels of "unfair" inequality, especially in the acquisition of labour income and in accessing tertiary education.

Given the importance of working-poor phenomenon, especially in developing countries, we investigate not only the inequity in getting a job, but also in getting a "decent" job. Inequality of opportunity for getting a "good job" is significantly higher than for finding any kind of employment. Importantly, parental membership to the Communist Party still plays a role in explaining inequality, particularly in having a "decent" job, suggesting the persistence of networks dating back to the pre-transition period. In order to study the evolution of inequality of opportunity over time, we estimate the level of unfair inequality in accessing tertiary education separately for a younger cohort (individuals who started their schooling after the

fall of the Berlin wall), and for an older cohort (the rest of our sample). Our results point in the direction of an increase of inequality of opportunity, mostly in non-EU members.

Among the circumstances that explain inequality of opportunity, parental background is the stronger driver. Parents' education plays a particularly important role in determining children's educational attainments. Gender is another noteworthy source of inequality of opportunity, especially in labour income acquisition, as a substantial gender wage gap persists. Moreover, female participation falls considerably when inequality of opportunity for "decent" jobs is higher. Disadvantages deriving from being born in a rural area is relatively more modest in the countries under study.
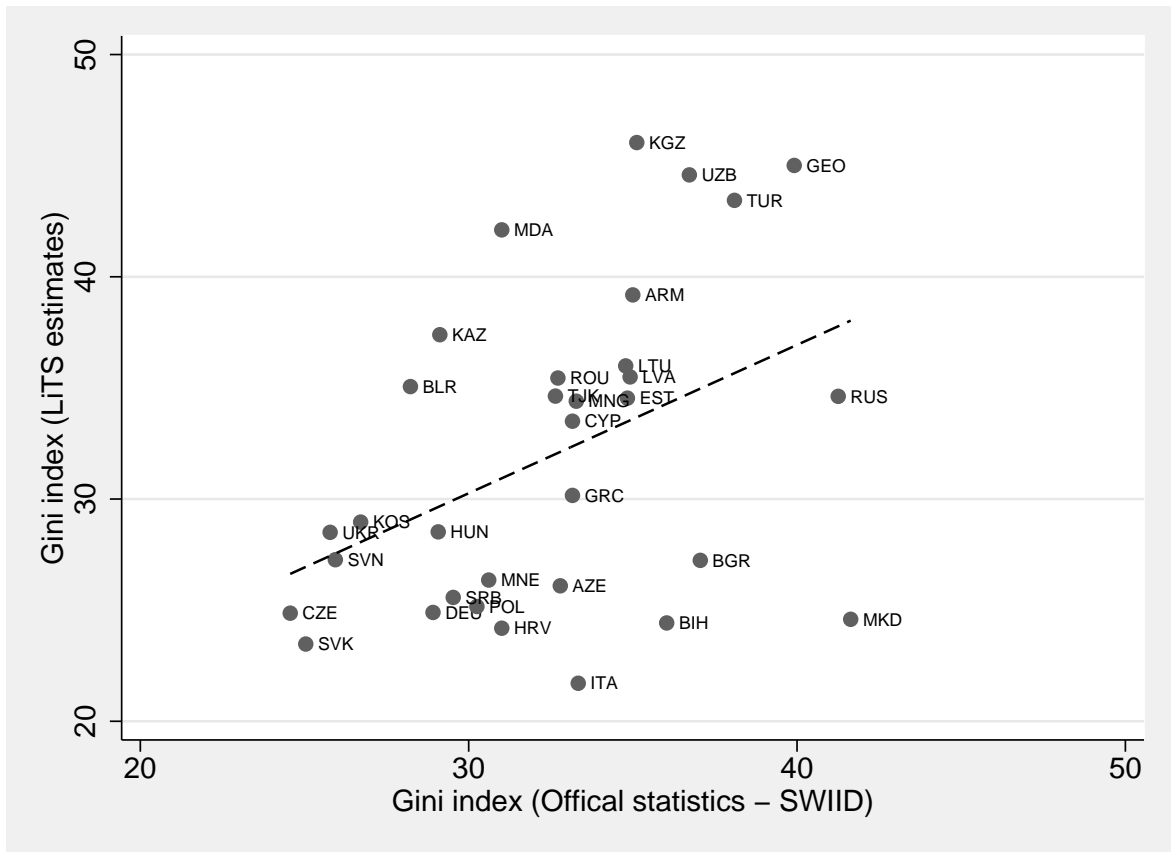
Figure 2.1. Comparison between the Gini index calculated on the winsorized distribution of earnings from the 2016 LiTS and the Gini index from the World Bank statistics. The correlation between the two measures of inequality is 0.42.
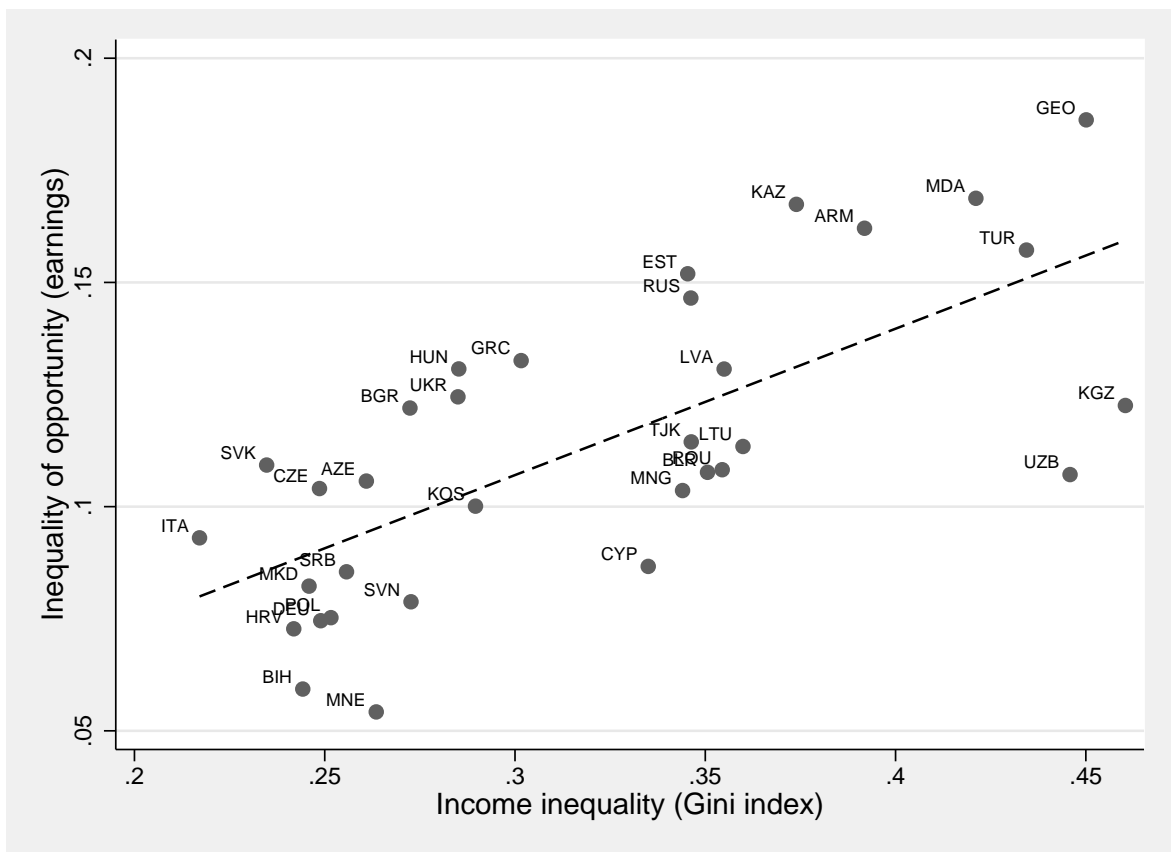
Figure 2.2. Relationship between overall income inequality, as measured by the Gini index of the winsorized distribution of earnings from the 2016 LiTS, and inequality of opportunity for earnings.
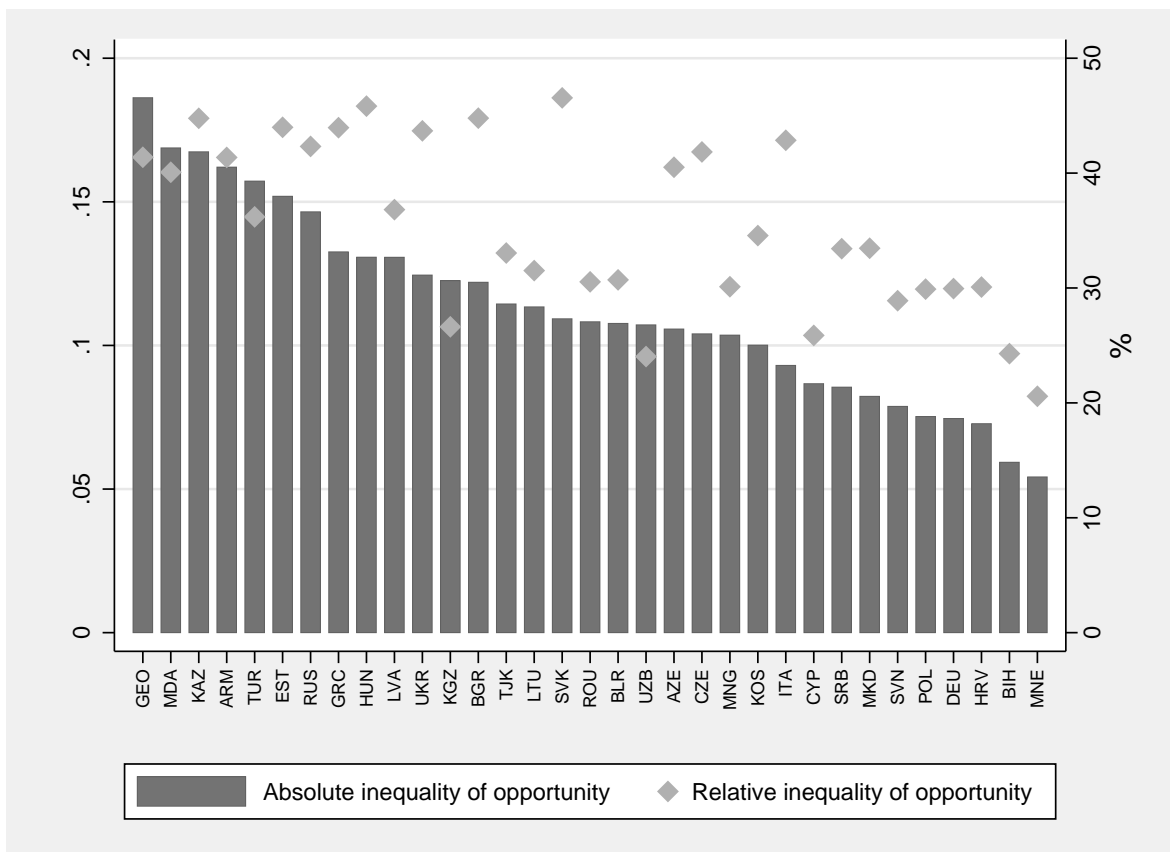
Figure 2.3. Absolute versus relative measures of inequality of opportunity for earnings. The relative measure is given by the ratio between the absolute measure and the Gini index of the overall distribution of earnings.
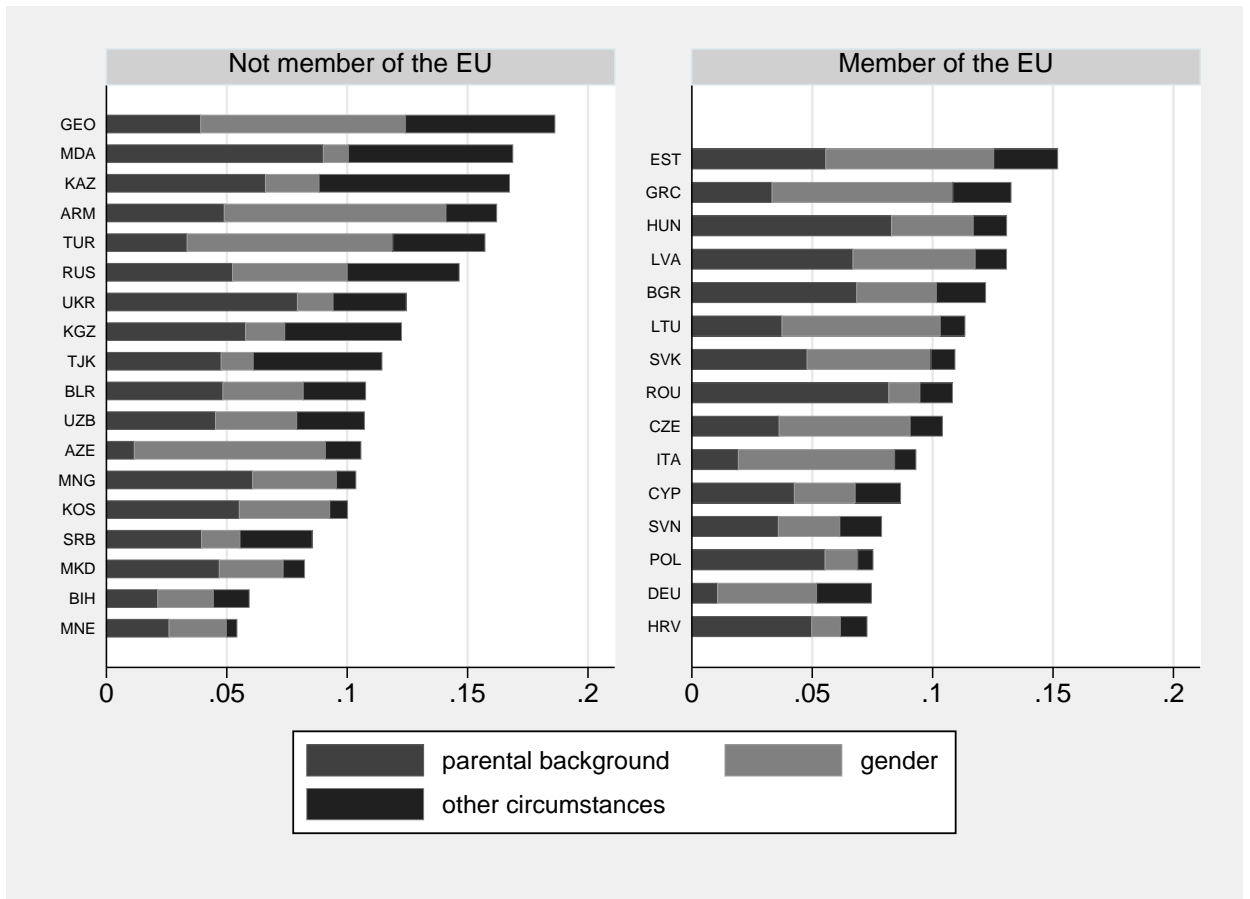
Figure 2.4. Relative contribution of circumstances to inequality of opportunity. "Parental background" includes both mother's and father's educational attainment and parental membership to the Communist Party. "Other circumstances" groups together birthplace (whether rural or urban) and being part of a minority.
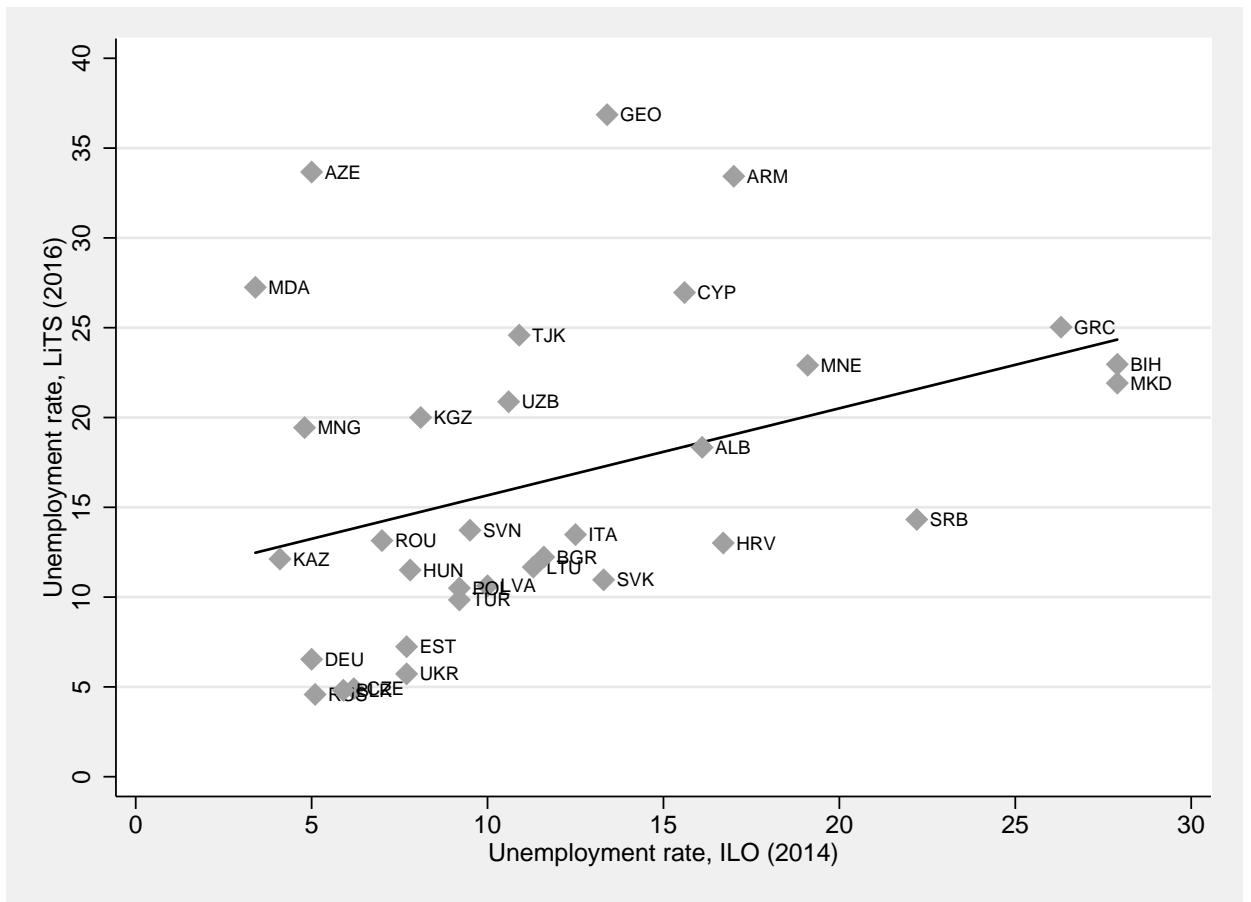
Figure 2.5. Comparison between the unemployment rates calculated from 2016 LiTS and from the ILO statistics.
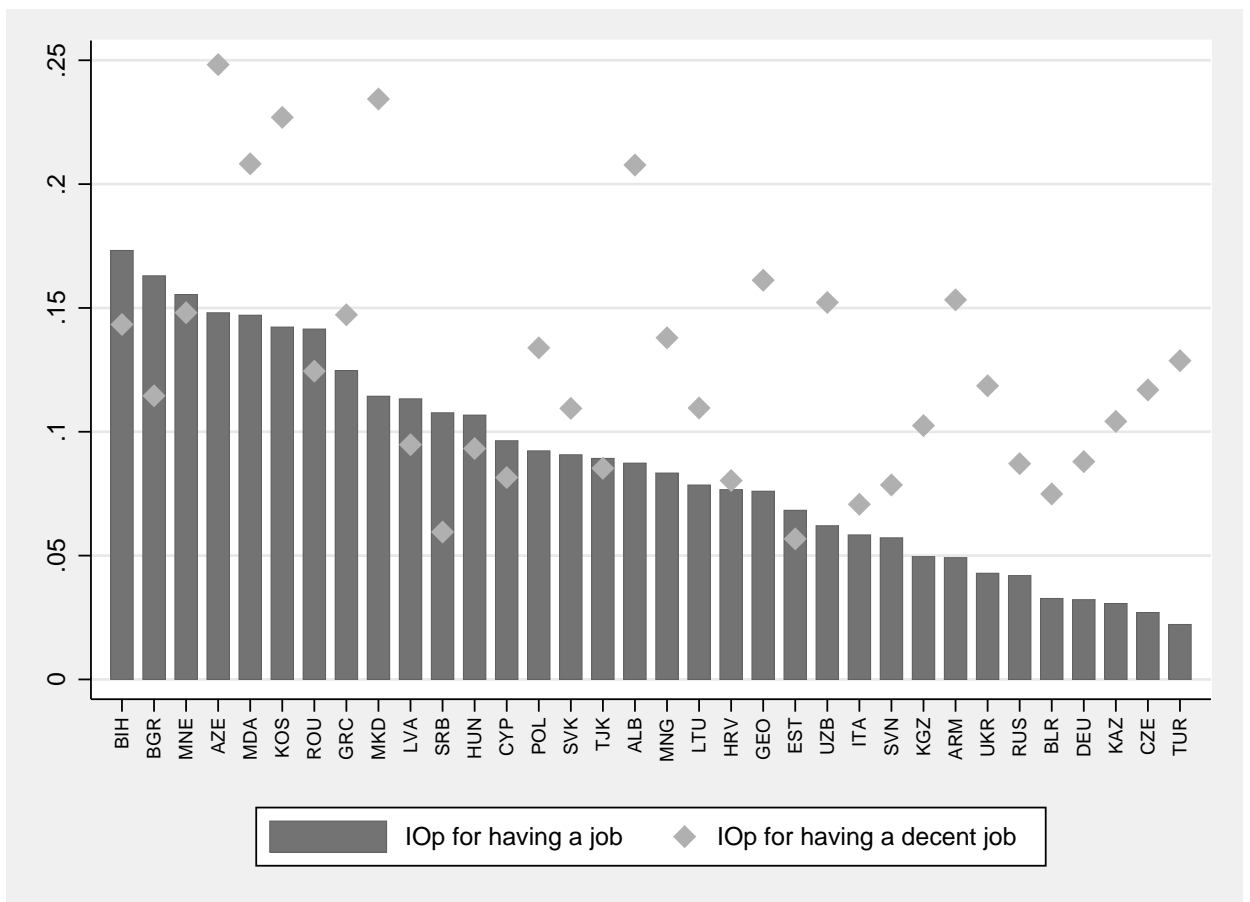
Figure 2.6. Level of inequality of opportunity for having a job and for having a "decent" one.
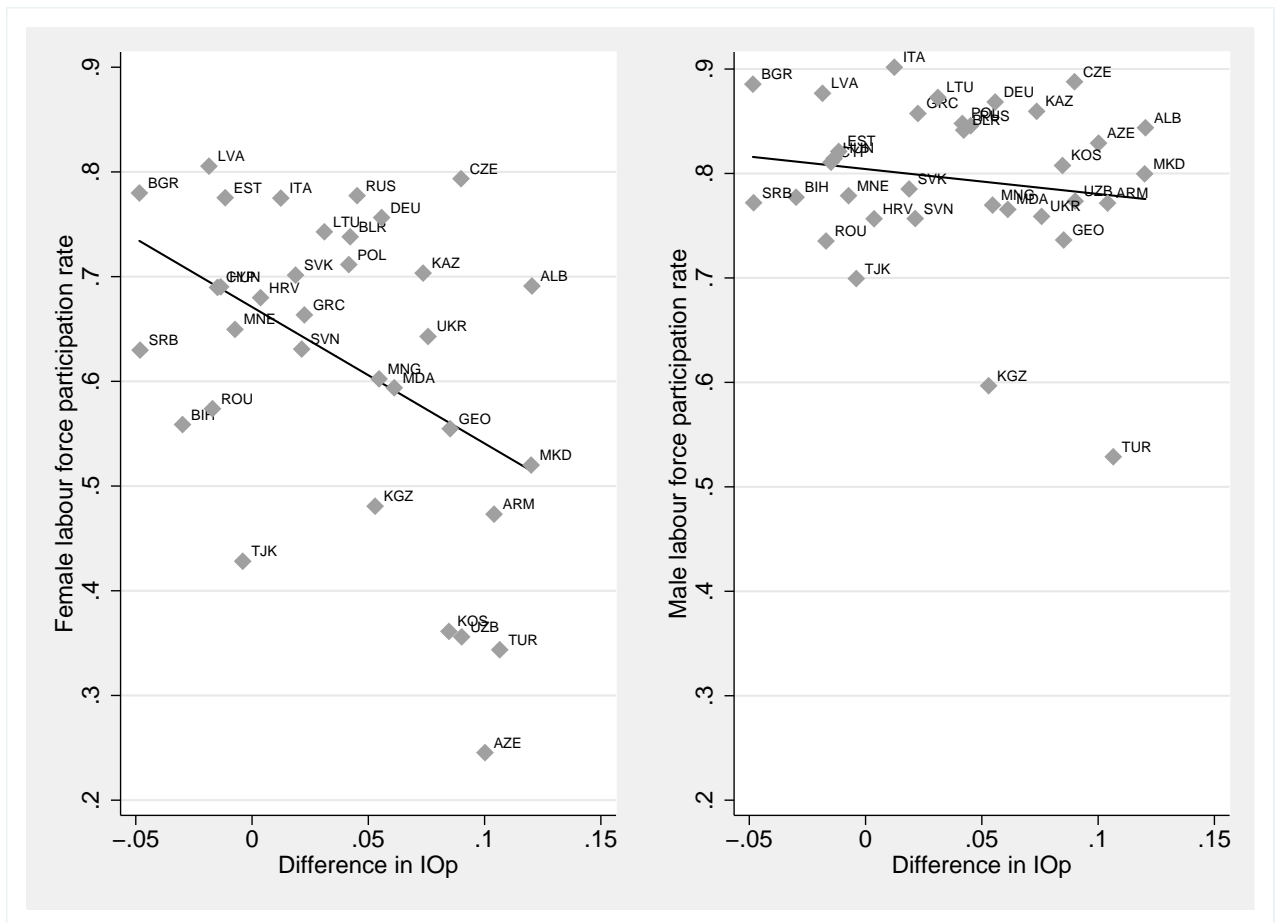
Figure 2.7. Relationship between the labour force participation rate, separately for women and men, and the difference in the level of inequality of opportunity for having a decent job and inequality of opportunity for having any job.
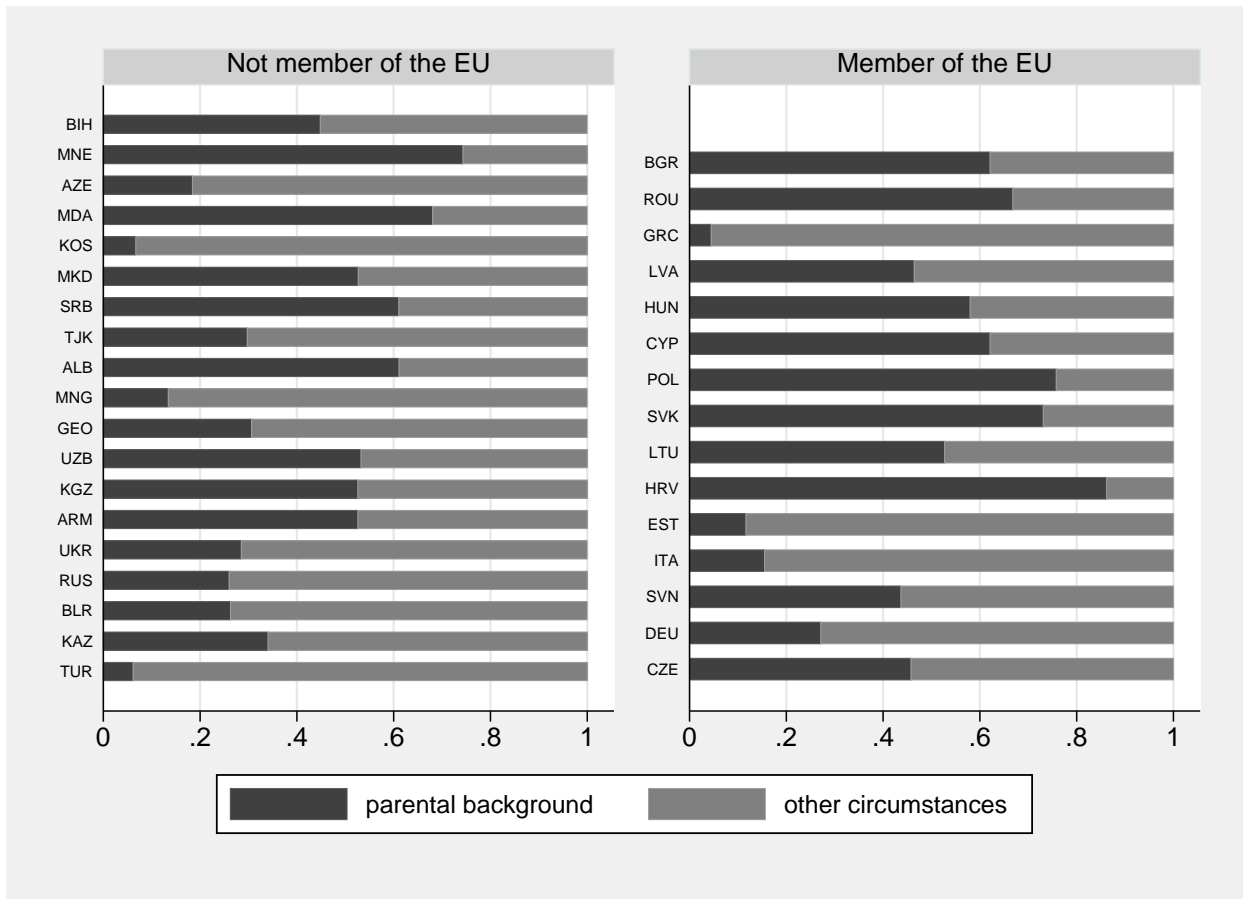
Figure 2.8. Relative contribution of circumstances to inequality of opportunity for getting any kind of job. "Parental background" includes both mother's and father's educational attainment and parental membership to the Communist Party. "Other circumstances" groups together birthplace (whether rural or urban), gender, and being part of a minority.
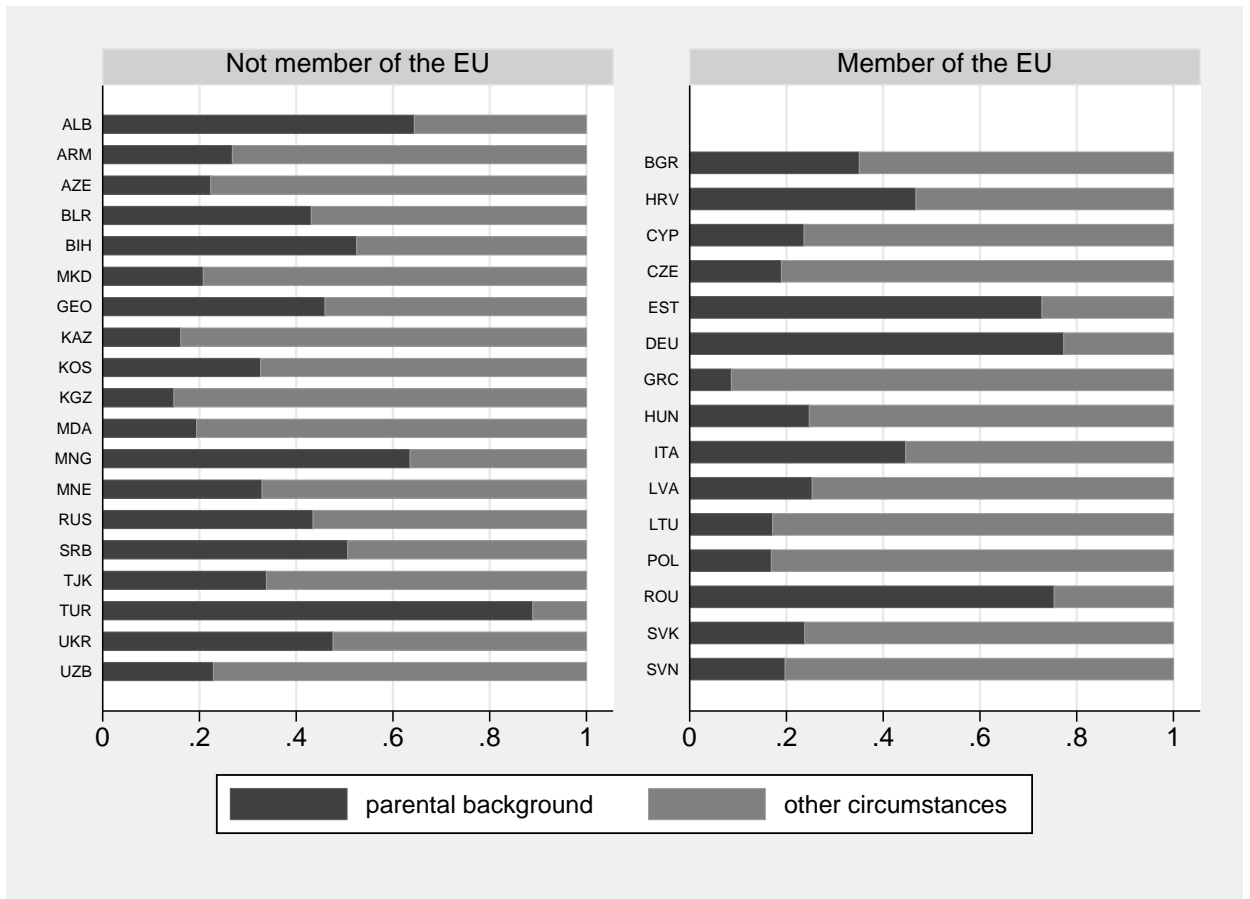
Figure 2.9. Relative contribution of circumstances to inequality of opportunity for getting a *decent* job. "Parental background" includes both mother's and father's educational attainment and parental membership to the Communist Party. "Other circumstances" groups together birthplace (whether rural or urban), gender, and being part of a minority.
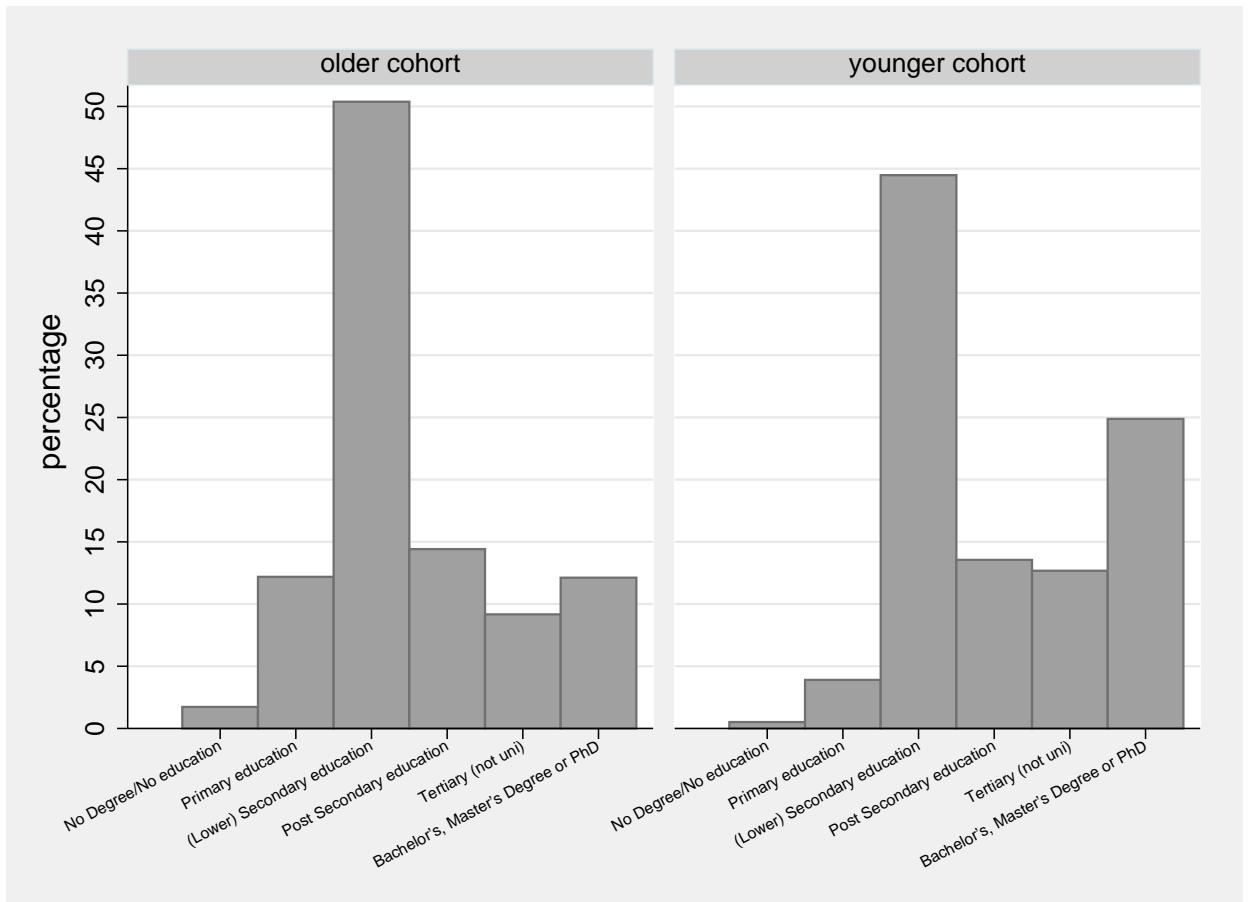
Figure 2.10. Educational attainments by cohorts. The younger cohort comprises all individuals who started education after 1989, while the older cohort groups together everybody else.
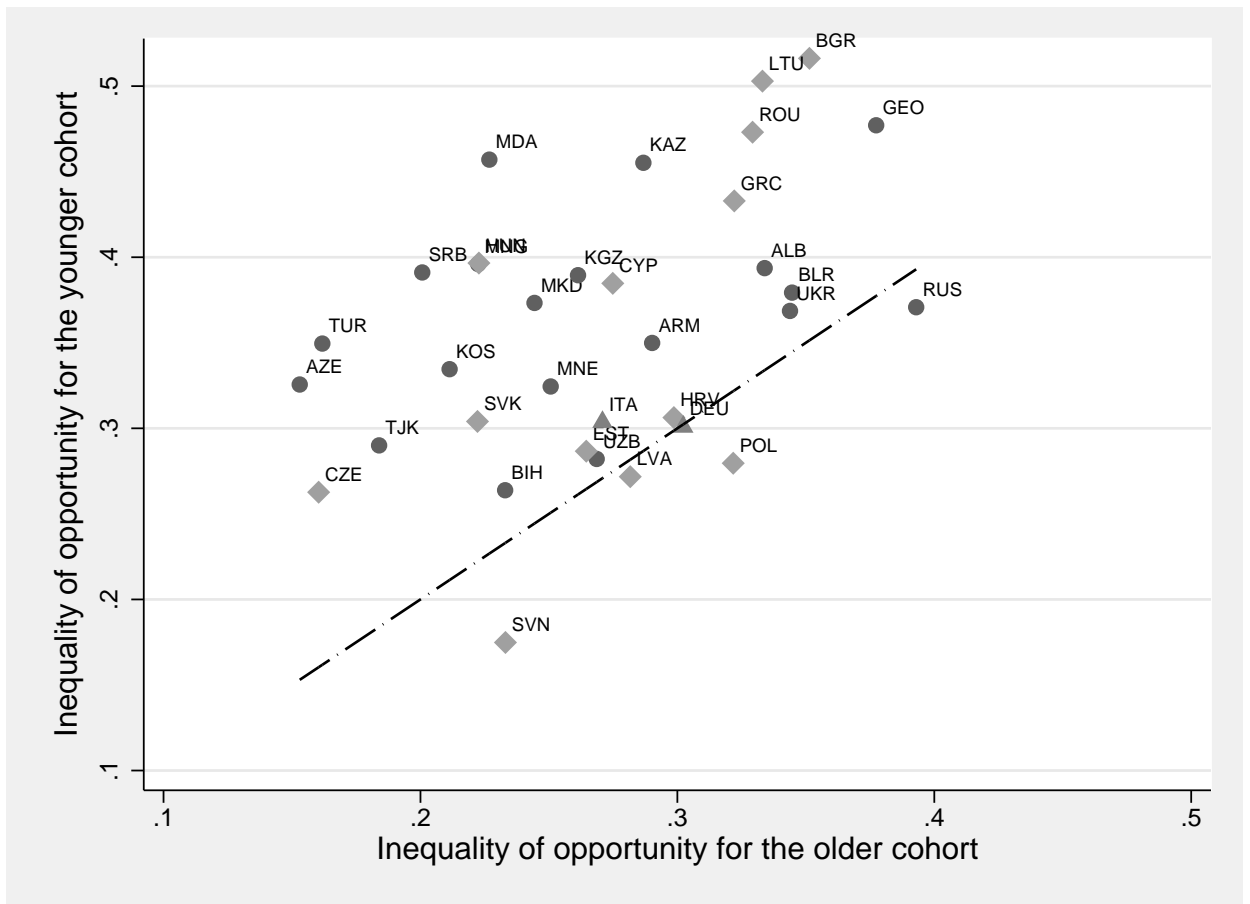
Figure 2.11. Inequality of opportunity for tertiary education for the two cohorts.
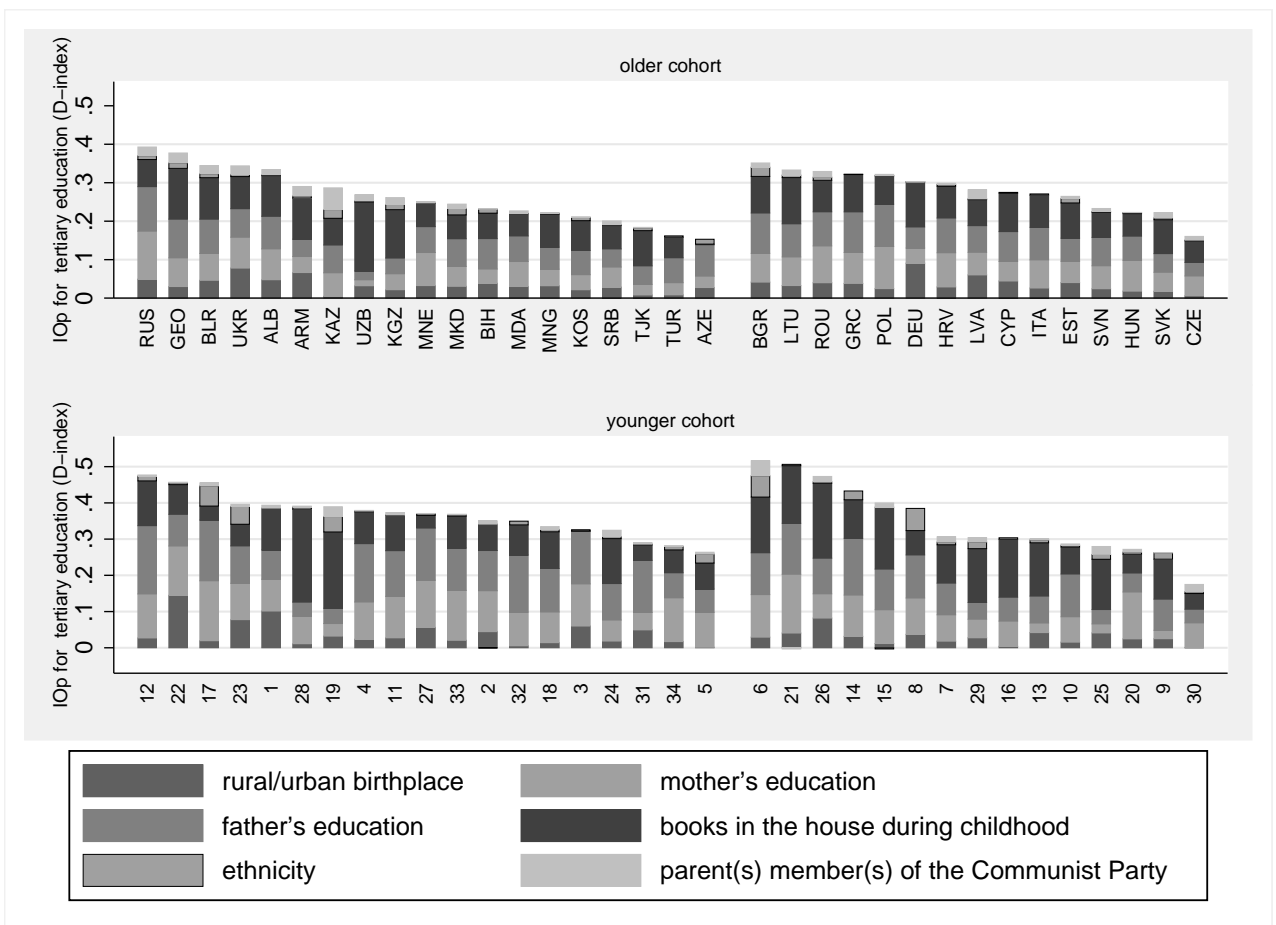
Figure 2.12. Relative contribution of circumstances to inequality of opportunity for getting some tertiary education, separately for the younger and for the older cohort.

Table 2.1. Missing data in individual self-reported earnings

Country	% self-employed and employers	% working and reporting earnings	% working but not reporting earnings	% working individuals (18-64)
Armenia	7.6	69.6	22.8	38.4
Azerbaijan	3.7	76.6	19.7	31.7
Belarus	4.0	68.2	27.8	74.6
Bosnia	8.3	59.6	32.1	51.6
Bulgaria	7.4	72.7	19.9	73.1
Croatia	7.0	62.8	30.2	62.3
Cyprus	12.2	77.2	10.6	54.3
Czech Republic	9.9	62.3	27.8	79.6
Estonia	6.6	77.9	15.5	73.8
Georgia	15.1	63.9	21.0	39.4
Germany	13.0	46.6	40.7	76.6
Greece	30.2	54.7	15.1	56.2
Hungary	4.9	64.7	30.4	66.2
Italy	13.9	75.8	10.3	72.6
Kazakhstan	6.0	63.5	30.5	66.7
Kosovo	10.1	66.7	23.2	37.4
Kyrgyzstan	15.1	55.5	29.4	42.8
Latvia	7.7	71.4	20.9	74.8
Lithuania	7.6	80.5	11.9	70.8
Macedonia	7.8	66.3	25.9	51.5
Moldova	9.2	76.7	14.1	49.4
Mongolia	19.7	62.7	17.6	54.6
Montenegro	8.6	68.4	23.0	54.7
Poland	10.2	44.4	45.4	68.6
Romania	3.9	67.5	28.6	56.0
Russia	4.6	68.4	27.0	76.7
Serbia	8.3	56.9	34.8	59.9
Slovak Republic	11.1	59.1	29.8	65.8
Slovenia	11.0	59.4	29.6	60.1
Tajikistan	13.0	60.4	26.6	41.6
Turkey	12.5	53.3	34.2	40.0
Ukraine	6.5	41.1	52.4	65.0
Uzbekistan	14.3	67.6	18.1	43.2

Note: Only individuals aged 18-64 have been considered in calculating percentages. Figures in the first three columns are calculated on the number of working individuals. The last column displays the percentage of the latter in the population of people aged 18-64.

Table 2.2. Comparison of characteristics between the group of those reporting labour income and the group of those not reporting labour income (all countries)

	Reporting	Non-reporting	Min	Max
Gender	.50	0.53	0	1
Age	40.90	40.61	18	64
Education	0.34	0.36	0	1
Marital status	0.60	0.62	0	1
Health status	0.67	0.71	0	1
N	14,877	5,307		

Note: "Education" is a dummy variable taking value 1 if individuals have some tertiary education.

Table 2.3. Comparison of the distribution of log-earnings (for those reporting labour income) and the distribution of predicted values (for those not reporting labour income)

Country	Mean		Standard deviation	
	log earning	fitted values	log earning	fitted values
ARM	13.40	13.56	0.97	0.35
AZE	8.12	8.16	0.55	0.35
BLR	17.70	17.78	0.61	0.23
BIH	9.10	9.11	0.51	0.20
BGR	8.84	8.96	0.55	0.25
HRV	10.88	10.90	0.50	0.24
CYP	9.30	9.44	0.82	0.37
CZE	12.08	12.16	0.59	0.21
EST	8.99	9.05	0.68	0.40
MKD	12.12	12.21	0.57	0.25
GEO	7.96	7.93	1.13	0.42
DEU	10.09	10.18	0.56	0.27
GRC	8.80	8.83	0.70	0.32
HUN	14.07	14.12	0.61	0.33
ITA	9.56	9.60	0.56	0.27
KAZ	13.26	13.25	1.07	0.41
KOS	8.28	8.25	0.56	0.37
KGZ	11.33	11.40	1.08	0.26
LVA	8.47	8.62	0.74	0.41
LTU	8.53	8.63	0.69	0.30
MDA	10.12	10.17	0.99	0.42
MNG	15.30	14.41	0.81	1.08
MNE	8.34	8.36	0.60	0.29
POL	10.15	10.19	0.51	0.26
ROU	9.60	9.63	0.70	0.26
RUS	12.29	12.37	1.19	0.36
SRB	12.83	12.81	0.70	0.25
SVK	8.79	8.83	0.51	0.27
SVN	9.20	9.28	0.52	0.29
TJK	8.84	8.93	0.75	0.32
TUR	9.80	9.81	0.88	0.41
UKR	10.19	10.22	0.73	0.24
UZB	15.58	15.77	0.82	0.43

Table 2.4. Returns to tertiary education

	(1) log earnings	(2) log earnings
tertiary education (d)	.352***	.293***
	(23.695)	(14.514)
gender (d)	-.253***	-.265***
	(-19.300)	(-19.471)
years of experience	.021***	.020***
	(7.806)	(6.627)
years of experience^2	-.001***	-.001***
	(-8.657)	(-7.633)
tertiary education communism (d)		.002
		(.054)
marital status (d)		.027**
		(2.258)
birthplace (rural/urban) (d)		.080***
		(5.235)
ethnic minority (d)		-.074***
		(-3.727)
father's tertiary education (d)		.047**
		(2.287)
books		.046***
		(6.921)
parental membership Communist Party (d)		.059***
		(2.985)
Constant	10.701***	10.559***
	(332.185)	(262.036)
Observations	1.41e+04	1.33e+04
R^2	.066	.075
R^2 adj.	.063	.072
F-stat	247.13	97.53

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
(d) indicates dummy variables.

Table 2.5. Inequality of opportunity for earnings, employment and education

Country	Obs	IOp earnings (a)	Obs	IOp any job (b)	Obs	IOp decent job (c)	Obs	IOp education (older) (d)	Obs	IOp education (younger) (e)
ALB			952	0.087 (0.006)	780	0.208 (0.016)	999	0.334 (0.021)	300	0.394 (0.021)
ARM	298	0.162 (0.048)	643	0.049 (0.009)	414	0.153 (0.027)	1020	0.290 (0.022)	247	0.350 (0.037)
AZE	339	0.105 (0.015)	630	0.148 (0.013)	418	0.248 (0.033)	730	0.153 (0.037)	360	0.326 (0.040)
BLR	604	0.107 (0.015)	921	0.033 (0.003)	878	0.075 (0.007)	1068	0.345 (0.021)	278	0.379 (0.036)
BIH	386	0.059 (0.015)	822	0.173 (0.008)	624	0.143 (0.011)	918	0.233 (0.030)	258	0.264 (0.031)
BGR	549	0.122 (0.012)	854	0.163 (0.007)	751	0.115 (0.009)	1178	0.351 (0.026)	171	0.516 (0.077)
HRV	450	0.073 (0.012)	804	0.077 (0.006)	696	0.080 (0.009)	1057	0.299 (0.022)	227	0.306 (0.033)
CYP	401	0.087 (0.018)	705	0.096 (0.009)	516	0.082 (0.015)	1217	0.275 (0.027)	175	0.385 (0.033)
CZE	549	0.104 (0.013)	928	0.027 (0.003)	884	0.117 (0.009)	1168	0.160 (0.030)	216	0.263 (0.046)
EST	480	0.151 (0.014)	650	0.068 (0.005)	606	0.057 (0.006)	1109	0.265 (0.018)	122	0.287 (0.041)
MKD	387	0.082 (0.017)	741	0.114 (0.009)	578	0.234 (0.014)	1002	0.244 (0.026)	241	0.373 (0.030)
GEO	272	0.186 (0.035)	669	0.076 (0.011)	421	0.161 (0.022)	1116	0.377 (0.016)	215	0.477 (0.029)
DEU	469	0.075 (0.015)	1084	0.032 (0.003)	1012	0.088 (0.005)	1029	0.302 (0.019)	322	0.301 (0.030)
GRC	343	0.133 (0.020)	838	0.125 (0.010)	628	0.147 (0.017)	1189	0.322 (0.020)	201	0.433 (0.039)
HUN	427	0.131 (0.012)	742	0.107 (0.006)	656	0.093 (0.010)	1148	0.223 (0.027)	173	0.397 (0.041)
ITA	622	0.093 (0.013)	934	0.058 (0.006)	806	0.071 (0.008)	1195	0.271 (0.023)	193	0.304 (0.056)
KAZ	541	0.167 (0.025)	955	0.031 (0.005)	835	0.104 (0.011)	1004	0.287 (0.017)	284	0.455 (0.029)
KOS	324	0.100 (0.020)	729	0.142 (0.012)	469	0.227 (0.017)	882	0.211 (0.033)	287	0.335 (0.035)
KGZ	309	0.123 (0.030)	699	0.050 (0.007)	547	0.103 (0.021)	983	0.261 (0.026)	328	0.389 (0.027)
LVA	428	0.130 (0.016)	650	0.113 (0.006)	587	0.095 (0.009)	947	0.282 (0.017)	169	0.272 (0.030)
LTU	526	0.113 (0.014)	736	0.079 (0.006)	650	0.110 (0.011)	1065	0.333 (0.014)	180	0.503 (0.035)
MDA	367	0.169 (0.022)	662	0.147 (0.009)	474	0.208 (0.017)	954	0.227 (0.027)	220	0.457 (0.032)
MNG	451	0.104 (0.021)	893	0.083 (0.007)	699	0.138 (0.020)	980	0.222 (0.025)	351	0.396 (0.024)
MNE	445	0.054 (0.015)	802	0.155 (0.009)	614	0.148 (0.015)	873	0.251 (0.025)	283	0.324 (0.027)
POL	346	0.075 (0.012)	868	0.092 (0.005)	775	0.134 (0.011)	1120	0.322 (0.019)	274	0.280 (0.029)
ROU	385	0.110 (0.019)	647	0.141 (0.007)	559	0.124 (0.011)	1115	0.329 (0.021)	171	0.473 (0.034)
RUS	619	0.147 (0.023)	928	0.042 (0.003)	887	0.087 (0.012)	935	0.393 (0.017)	276	0.371 (0.028)
SRB	377	0.085 (0.020)	725	0.108 (0.006)	620	0.060 (0.009)	1010	0.201 (0.030)	224	0.391 (0.041)
SVK	424	0.109 (0.014)	793	0.091 (0.006)	707	0.109 (0.009)	1220	0.222 (0.026)	166	0.304 (0.055)
SVN	333	0.079 (0.013)	632	0.057 (0.006)	547	0.079 (0.008)	1179	0.233 (0.025)	126	0.175 (0.051)
TJK	348	0.114 (0.029)	750	0.089 (0.007)	557	0.085 (0.022)	855	0.184 (0.024)	306	0.290 (0.043)
TUR	307	0.157 (0.029)	636	0.022 (0.004)	570	0.129 (0.030)	721	0.162 (0.044)	466	0.350 (0.029)
UKR	307	0.124 (0.021)	794	0.043 (0.004)	742	0.119 (0.014)	1129	0.344 (0.018)	247	0.369 (0.025)
UZB	379	0.107 (0.021)	700	0.062 (0.007)	552	0.152 (0.020)	864	0.269 (0.022)	266	0.282 (0.047)

Note: Gini index is used as an inequality measure for earnings, while Dissimilarity Index (D-Index) is used for employment and education (binary variable outcomes). Bootstrapped standard error computed with 99 replications in parenthesis.

Table 2.6. Decomposition of inequality of opportunity for employment into the relative contribution of circumstances

Country	Birthplace	Mother's education	Father's education	Gender	Ethnic minority	Communist Party
			Panel (a): Having any kind of job			
ALB	0.275	0.336	0.241	0.113	0.000	0.034
ARM	0.121	0.124	0.049	0.320	0.033	0.352
AZE	0.092	0.043	0.032	0.619	0.105	0.109
BLR	0.274	0.172	0.043	0.247	0.214	0.048
BIH	0.026	0.231	0.207	0.305	0.217	0.010
BGR	0.107	0.265	0.319	0.020	0.251	0.037
HRV	0.015	0.380	0.423	0.083	0.033	0.059
CYP	0.303	0.301	0.320	0.069	0.008	
CZE	0.020	0.298	0.157	0.117	0.405	0.002
EST	0.103	0.045	0.043	0.153	0.627	0.029
MKD	0.037	0.149	0.363	0.237	0.199	0.015
GEO	0.005	0.148	0.112	0.677	0.012	0.047
DEU	0.518	0.233	0.037	0.090	0.121	0.002
GRC	0.158	0.022	0.023	0.704	0.093	
HUN	0.066	0.341	0.226	0.082	0.272	0.012
ITA	0.086	0.048	0.107	0.734	0.024	
KAZ	0.641	0.056	0.025	0.005	0.014	0.259
KOS	0.072	0.027	0.017	0.851	0.009	0.023
KGZ	0.090	0.052	0.013	0.269	0.115	0.461
LVA	0.044	0.326	0.123	0.291	0.197	0.015
LTU	0.311	0.259	0.232	0.047	0.113	0.037
MDA	0.267	0.252	0.367	0.008	0.045	0.061
MNG	0.217	0.077	0.017	0.079	0.569	0.041
MNE	0.142	0.203	0.505	0.009	0.105	0.036
POL	0.152	0.264	0.221	0.046	0.044	0.272
ROU	0.237	0.232	0.377	0.014	0.081	0.059
RUS	0.413	0.082	0.173	0.041	0.286	0.005
SRB	0.280	0.319	0.277	0.101	0.006	0.014
SVK	0.068	0.429	0.273	0.056	0.144	0.029
SVN	0.023	0.101	0.288	0.384	0.156	0.049
TJK	0.031	0.161	0.015	0.126	0.537	0.121
TUR	0.030	0.014	0.048	0.304	0.605	
UKR	0.476	0.083	0.101	0.236	0.003	0.101
UZB	0.222	0.053	0.249	0.104	0.141	0.230

Panel (b): Having a decent job

Country	Birthplace	Mother's education	Father's education	Gender	Ethnic minority	Communist Party
ALB	0.147	0.330	0.310	0.204	0.004	0.004
ARM	0.089	0.069	0.133	0.640	-0.020	0.066
AZE	0.374	0.040	0.165	0.368	0.035	0.018
BLR	0.013	0.026	0.072	0.433	0.123	0.333
BIH	0.345	0.071	0.205	0.084	0.046	0.248
BGR	0.116	0.142	0.134	0.110	0.423	0.075
HRV	0.040	0.272	0.178	0.367	0.126	0.018
CYP	0.007	0.164	0.072	0.608	0.148	
CZE	0.067	0.020	0.043	0.728	0.016	0.127
EST	0.132	0.638	0.081	0.001	0.136	0.009
MKD	0.254	0.085	0.097	0.045	0.493	0.026
GEO	0.051	0.106	0.237	0.449	0.041	0.117
DEU	0.023	0.506	0.232	0.082	0.121	0.035
GRC	0.423	0.028	0.058	0.087	0.404	
HUN	0.442	0.024	0.012	0.237	0.072	0.211
ITA	0.097	0.267	0.180	0.268	0.188	
KAZ	0.359	0.025	0.003	0.089	0.388	0.133
KOS	0.067	0.063	0.208	0.071	0.536	0.055
KGZ	0.339	0.052	0.083	0.262	0.252	0.012
LVA	0.119	0.046	0.136	0.596	0.033	0.071
LTU	0.104	0.046	0.090	0.650	0.074	0.035
MDA	0.150	0.044	0.017	0.629	0.027	0.133
MNG	0.150	0.221	0.388	0.135	0.079	0.026
MNE	0.084	0.127	0.194	0.451	0.136	0.008
POL	0.003	0.023	0.085	0.826	0.002	0.061
ROU	0.192	0.337	0.130	0.045	0.008	0.287
RUS	0.061	0.243	0.132	0.471	0.033	0.059
SRB	0.013	0.279	0.056	0.243	0.237	0.171
SVK	0.023	0.042	0.012	0.731	0.008	0.184
SVN	0.009	0.069	0.018	0.709	0.086	0.110
TJK	0.469	0.035	0.298	0.135	0.057	0.004
TUR	0.017	0.276	0.613	0.088	0.006	
UKR	0.073	0.120	0.135	0.420	0.030	0.222
UZB	0.056	0.190	0.021	0.690	0.024	0.017

Chapter 3

Ethnic Diversity and Labour Market Outcomes: Evidence from Post-Apartheid South Africa

3.1 Introduction

A growing body of literature studies the link between diversity and the economic performance of regions and countries. One aspect of diversity which has increasingly attracted the interest of social scientists is ethnic diversity. Conceptually, the relationship between ethnic diversity and economic development can be both positive and negative, as is summarised in Alesina and La Ferrara (2005). On the one hand, in highly heterogeneous communities, new and innovative ideas are more likely to emerge and to consolidate into what Sobel et al. (2010) calls "cultural capital" (i.e. cultural and artistic creativity). Diversity is also potentially beneficial to technological and scientific innovation due to the complementarity of skills of individuals from different ethnic backgrounds (Fainstein, 2005; Eraydin et al., 2010).

44

On the other hand, more ethnically fractionalized communities can experience slower economic development measured by GDP per capita (Easterly and Levine, 1997) or city size growth (Glaeser et al., 1995), as ethnic diversity may be related to high social costs which are reflected in lower level of trust and participation in social activities (Alesina and La Ferrara, 2000, 2002), inefficient public goods provision (Alesina et al., 1997) and higher inequality (Alesina et al., 2016). The ethnic cleavage may also be detrimental to the establishment of a culture of inclusiveness and tolerance which is favorable to economic growth.

A general perspective from macro-level empirical studies (mostly cross-country studies), however, is that ethnic diversity is negatively associated with economic opportunities especially in African countries featured by high ethnic fragmentation (Michalopoulos, 2012; Michalopoulos and Papaioannou, 2013)[1]. Ethnic fragmentation harms the economic performance in these countries as it is associated with the under-investment of public goods (Michalopoulos and Papaioannou, 2013), conflict (Amodio and Chiovelli, 2017) and the collective action failures resulting from the difficulties in imposing social sanctions in diverse places (Miguel and Gugerty, 2005).

Much less is known about the micro-level evidence on how ethnic diversity affects individuals' outcomes, especially labour market performance which is of great importance in driving economic development. For example, increase in employment can significantly reduce inequalities (Anand et al., 2016). There is some firm-level microeconometric evidence on the direct effect of ethnic divisions on workers' productivity in Kenya which documents that upstream workers undersupply downstream workers at the sacrifice of total output if these people come from different

---

[1]More research in developed world finds support for the positive side of diversity (Andersson et al., 2005; Niebuhr, 2010; Ottaviano and Peri, 2006). The relationship between diversity and economic performance can also be non-linear. For example, Nikolova et al. (2013) use data from the post-soviet states and show that entrepreneurship is increasing in ethnic heterogeneity at low level of diversity, while it loses its positive impact when diversity reaches a certain threshold.

ethnic groups (Hjort, 2014). Some other literature looks at how entrepreneurs from a specific ethnic group make use of their ethnic networks to develop social capital and mobilise resources (Iyer and Shapiro, 1999), but they are not directly linked to ethnic diversity. In general, how the level of ethnic fractionalization affects labour market outcomes in African countries and how it affects people's choice between self-employment and being an employee remain unclear.

This paper investigates how within-black ethnic diversity affects black individuals' labour market outcomes in post-Apartheid South Africa. We focus on how their employment rate responds to the composition of black ethnic groups in the district of their residence[2]. Post-Apartheid South Africa provides a unique and interesting setting for the study of the diversity-labour market nexus. On the one hand, ethnic identity remains distinct even after generations of integration and interaction among different ethnic groups. This is because ethnicity became a salient concept during Apartheid (from 1948 to 1994) when pervasive racism, discrimination and segregation were meant to guarantee the power of the whites. The Apartheid government deteriorated inter-ethnic relationship by reinforcing the ethnic solidarity to prevent black ethnic groups from forming a coalition to fight against the white government (Gradin, 2014). Therefore distinctive features and identities between ethnic groups are not completely mitigated by hundreds of years of integration.

On the other hand, the Apartheid regime has largely destroyed both the regional path dependence in demand of black labour and the intergenerational occupational persistence in labour market outcomes by compressing the educational and job opportunities of the black South Africans universally. The Apartheid government imposes strict labour regulations to prevent the black South Africans from performing semi-skill and skilled jobs or running their own business in "white" areas.

_____

[2]There is literature about ethnic diversity at the workplace level, which shows the complementarities between workers from different cultural backgrounds as a rationale for the existence of a global firm (Lazear, 1999b). This is beyond the scope of this paper as we do not have access to firm-level data reflecting the ethnic composition in the workplace.

Therefore the post-Apartheid era is the first time since early 20th century when the majority of the black South Africans could freely make decisions on occupations and set up their own business. Thus contemporaneous labour market outcomes of the black might convey less information on the persistence in regional labour demand or inherited abilities, but is more related to their own experiences and living environments including the ethnic composition of their communities.

Baseline results, based on 1996 and 2001 census data, show that black individuals are more likely to be employed in a more ethnically diverse district (measured by intra-black ethnic diversity), especially more likely to work as an employee (as opposed to setting up their own business).

One challenge in interpreting this as a causal relationship is that the formation of ethnic diversity in a district may not be random. For example, if a district has more job opportunities or higher levels of development, it will attract people from different ethnic backgrounds and they will be more likely to be employed simply due to potentially higher labour demand in those districts. Or if people with some specific characteristics (i.e. higher ability) are attracted by more ethnically diverse districts, they might also perform better than less-abled counterparts wherever they go. A simple OLS regression will therefore amplify the effect of ethnic diversity on employment [3].

We therefore turn to an instrumental variable strategy, which relies on the historical origins of blacks' settlements (known as "homelands"). In particular, our instrument exploits the fact that assuming the magnitude of migration decreases with the distance between the original homelands and the destination districts outside those homelands, a district tends to host a more diverse population if it is equally distant to multiple homelands. On the contrary, a district becomes more homogeneous if

_____

[3]OLS may also underestimate the effect if individuals with lower ability move to more diverse places if it is easier to find jobs in those places.

it is relatively close to one homeland but far away from the rest. Importantly, the equidistance to multiple homelands remains a strong predictor of ethnic diversity even after controlling for the proximity of the district to the closest homeland. This further confirms that what can be captured by this instrument is not purely the absolute distance to these homelands but the equal distribution of distance to multiple homelands.

In our main IV regressions, one standard deviation increase in ethnic diversity index in 1996 (2001) decreases unemployment (including those who are economically inactive) by 0.0262 (0.044) point, which is 4.27% (6.92%) of the average unemployment rate in 1996 (2001). This positive association is more prevalent among the black ethnic groups with relatively larger population size and among people with lower levels of education.

We propose a model of a coordination game in the spirit of literature on social interaction to explain the unusual finding of the positive effect of ethnic diversity in an African country. Especially it can explain why only groups with relatively larger population size respond to ethnic diversity in our empirical results. As inter-ethnic communication is more costly (because one needs to cross language or cultural barriers for example) than intra-ethnic connection and that people get positive utility from social connection with diminishing returns (because they can get tired from social life), we document that in a more ethnically diverse place people have to communicate with a larger proportion of individuals outside their own ethnic group to maintain a certain level of social connection. Therefore it is more necessary for them to invest in social skills to be prepared for inter-ethnic connection. Their labour market outcomes will improve accordingly as these additional social skills can help them in job search, either by reducing search cost or by improving their productivity.

We then show why only groups with larger population size respond to ethnic di-

versity, starting with the initial condition where everyone in the district invests in social skills and participates in inter-ethnic communication. Groups with larger size are more likely to deviate from this coordination in a homogeneous place (where they dominant in group size) because they can get enough social connection purely by intra-ethnic communications. This is less likely to be the case if the district is more diverse where they are no longer the dominant group. For groups with smaller size who heavily rely on inter-ethnic connection, they do not have the incentive to deviate and will always participate in inter-ethnic interaction and invest in social skills regardless of the ethnic composition of the district.

This paper contributes to the literature in three ways. Firstly, we propose an instrumental variable to capture ethnic diversity, in a setting where two commonly established identification strategies might not be feasible. The first approach relies on the exogenous change of ethnic diversity in the time dimension, for example due to the implementation of new jurisdictions (Alesina et al., 2016). The second approach is based on natural or quasi-experiments which directly affects the level of ethnic diversity. For example, Algan et al. (2016) explore an exogenous allocation of public housing in France at the apartment block level and Dahlberg et al. (2012) make use of a policy on the compulsory allocation of refugees in Sweden. In South Africa, however, ethnic diversity does not change dramatically over time, which means there is not enough time variation to identify changing levels of diversity. It is also hard to find proper natural and quasi-experiments due to the political sensitivity of ethnic topics in this country.

Furthermore, our instrument has advantages over some other instruments (not necessarily instruments for ethnic diversity) which explore geographical features. For example, distance to certain places is largely used as an instrument for migration but whether this is orthogonal to economic conditions has been challenged[4]. By

---

[4]For example, a place close to an economic centre might get the positive spillover from the centre; or a place close to the road might perform better than others so that the demand for road

construction we control for the distance to the closest homeland and explore the remaining variation in equidistance to multiple homelands, which could be less problematic than distance itself. Alternatively, one can use the historical ethnic diversity directly as an instrument for contemporary diversity level, as is exploited in Miguel and Gugerty (2005) who use the historical distribution of ethnic residence in two districts in Kenya as an instrumental variable to study ethnic diversity and public goods provision. Such a historical distribution of ethnic settlements might also be correlated to other factors. For example, they find that places where several settlements intersect are in lack of sufficient public goods provision. This might however not be because there are fragmented ethnic groups but just public policies are less effective at the border between different districts in general (whether or not these districts represent different territories of ethnic groups). Our instrument mitigates this violation of exclusion restriction by focusing on districts outside these settlements instead of the settlements themselves. More importantly, by construction we can have places relatively far from all homelands but still with reasonably high ethnic diversity level as long as they are equidistant to all homelands. These places are less likely to be affected by the initial conditions of original homelands.

In addition, our identification strategy can easily be generalised to studies on other types of diversity. For example, replacing homelands with individuals' countries of origin, one can instrument the ethnic composition of immigrants in Europe or the U.S. with a measure of equidistance to multiple home countries (Alesina et al. (2015) implements an approach similar to this).

Secondly, we contribute theoretically to the mechanism through which ethnic diversity affects economic performance. The theoretical analysis in this paper, together with literature on the importance of social skills in employment, provides a new perspective on how ethnic diversity potentially affects labour market outcomes in

---

is higher in this place.

a positive way in South Africa (potentially in other developing countries as well). Traditional explanations indicating why diversity improves labour market performance, such as knowledge spillover, skill complementarity and discrimination, are not completely compatible with our empirical evidence[5]. We therefore mainly relate to literature on social interactions in communities with different levels of diversity for an explanation consistent with our results.

There are two key differences between our model and several models documenting social interactions in response to diversity in current literature. On the one hand, unlike models relying on the intrinsic ethnic-specific parameters of taste, preference or discrimination (for example, Morgan and Vardy (2009) shows minority candidates produce noisier signals of their ability), we show that ethnic diversity still affects people's decision in investments in social skills without documenting those assumptions. This is in line with the recent finding that ethnic diversity can be independent of cultural diversity (Desmet et al., 2017).

On the other hand, unlike Glaeser et al. (1992) which requires that communication is more extensive or the amount of social connection is larger in more diverse places (Alesina and La Ferrara, 2000), in our model the overall level of social interaction does not necessarily increase with ethnic diversity (total social interaction is the sum of both intra- and inter-ethnic connections). Ethnic diversity results in more investments in social skill because inter-ethnic communication is more costly (or requires more skills) than intra-ethnic connection.

The mechanism in our paper is the closest to, yet distinct in important aspects from two existing papers. In the story in Lazear (1999a), he finds that immigrants to the U.S. have higher English proficiency when there are smaller proportions of people from their native country in the communities in their destination. Our paper also documents that people are incentivised to learn another language for more poten-

---

[5]Detailed discussion is in the theoretical section of the paper.

tial communication partners (in our story we generalise "language" to a broader concept of social skill). The key difference is that they focus on the assimilation of the immigrants to the U.S and therefore the majority group (i.e. the U.S. native) do not respond to the diversity of the population in those communities. However, both the theoretical model and empirical findings in our paper show that only groups with larger size (analogue to the U.S. native) are affected by ethnic diversity of the districts whereas smaller groups (analogue to the minority group of immigrants in the U.S.) behave indifferently between ethnically diverse and homogeneous places[6]. What generates this difference is that his model is featured by unilateral assimilation of the immigrants to the U.S. while in our model social interaction and skill investments are bilateral. This makes more sense when we study a highly fragmented society where no ethnic group has dominance in group size (even the largest ethnic group makes up only around 25% of the whole population in our sample). Also due to strong ethnic identities, groups with smaller size will invest in a common or official language (groups with larger size do so as well) rather than the language of the large group.

In another model on social interactions between different groups, Alesina and La Ferrara (2000) assume that individuals prefer to communicate with people with similar income, race or ethnicity and conclude that homogeneous communities have higher levels of social capital. Instead of making the assumption of group-based preference directly, we treat this as an implicit implication of the model and explain that the reason for people's preference towards groups similar to them is lower intra-ethnic communication cost.

Moreover, our mechanism expands the literature on the importance of skill composition in labour market by linking skill mix to ethnic relations. Researchers in labour economics have highlighted the importance of skill mix (Acemoglu and Au-

---

[6]We control for the proportion of the black over the whole population in our analysis and focus on within-black communication.

tor, 2011) and social skills in production function (Deming, 2017). For example, our finding that individuals especially low-educated black people are motivated to invest in social skills in addition to human capital is consistent with the idea that when the labour market size is small, workers tend to invest more for the breadth instead of the depth of skills as there is less worker specialisation in certain tasks (Kim, 1989).

Recent literature also argues that higher communication skills in the workplace can facilitate people's trading of tasks based on each other's comparative advantage, therefore increasing their productivity (Deming, 2017). Taking a step back, we provide some insight on how to motivate the acquisition of social skills in preparation for the labour market. Our mechanism shows that this could potentially be achieved by encouraging the ethnic diversity of their communities.

Thirdly, we contribute to the literature on South African labour market by emphasising another dimension of inter-group relations in addition to black-white divisions. Studies on South Africa have been focusing on the interaction and segregation between black and white populations while each group within the black population is implicitly seen as being homogeneous. However, inter-ethnic relationship within the black population can also affect labour market outcomes. Especially, in the specific context of post-Apartheid South Africa, finding a positive effect of ethnic diversity on employment may be particularly unexpected after almost half a century of Apartheid which contributed to the erosion of inter-ethnic relationship and social capital.

Furthermore, major obstacles to contemporary unemployment in South Africa can potentially be well-tackled by looking at within-black ethnic diversity. Banerjee et al. (2008) propose that the stagnancy of the high-unemployment rate among the black in post-Apartheid South Africa might be mainly due to high search cost in job hunting and little growth in the informal sectors. On the one hand, social skill

acquisition can reduce the high search cost in an ethnically diverse district. On the other hand, as the informal sector is not powerful enough to generate more employment opportunities, black South Africans still rely heavily on jobs in formal sectors where social skills can be important as these jobs require more skill complexity. Coupled with the empirical finding that the high returns to education is only important for high-school graduates and above (Wittenberg, 2002), our story shows that ethnic diversity and investment in social skill can perform as substitutes for formal human capital accumulation which especially benefit the less educated.

The paper unfolds as follows. In Section 2, we provide a historical overview of the pattern and formation of ethnic diversity as well as summary statistics of labour market in South African context. In Section 3, we describe the data sources used for the analysis and how we construct the variables of interest. Section 4 details the empirical methodology used for the analysis, focusing on the instrumental variable and its validity. In Section 5 we comment on the results about how ethnic diversity affects labour market outcomes in post-Apartheid South Africa and how this impact differs across sub-groups. Section 6 proposes a plausible theoretical model to explain the above empirical results and rule out some alternative explanations. Finally we draw some conclusions and policy implications in Section 7.

3.2 Institutional Setting

3.2.1 Ethnic groups in South Africa and the formation of ethnic diversity

None of the black ethnic groups in contemporary South Africa are indigenous in this country. These groups migrated from eastern and central Africa to southern Africa starting from centuries ago, as part of the so-called "Bantu migration".

Before explaining the narratives, two concepts should be made clear. The first is "homeland" which refers to the original settlements of those ethnic groups when they first moved to South Africa. The second is "white areas" or "white South Africa"[7] which refers to places in South Africa outside those homelands. Many years after arrival in South Africa, those black people moved out of their original homelands and ended up in these "white areas" due to different reasons, mainly the pressure of conflicts with the British and Dutch colonisers as well as other ethnic groups. Therefore, "white areas" are not areas where only white people reside, but places outside original black homelands (the proportion of the black over the whole population can be large in those "white areas").

Based on Mwakikagile (2010) and Gradin (2014), we provide historical narratives on the mass migration of ethnic groups from central Africa towards South Africa, the original settlements of these ethnic groups and the migration of these people out of their homelands to "white areas" in South Africa. The timeline about the history of the settlements and migration of the black ethnic groups outside their own settlements up till the time of South Africa's independence can be found in the upper panel of Figure 3.1.

The indigenous groups in South Africa are San and Khoikhoi (both are "coloured" groups) residing in the southwestern and southeastern coast about 2000 years ago. Around 700s A.D., black Africans had settled in the northern part of what is South Africa today[8]. They were members of different Bantu ethnic groups who had moved southward from East-Central Africa (the Great Lake district around Congo) and spoke related languages.

Ethnicity-specific information on the Bantu migration from eastern and central Africa towards South Africa and the formation of ethnic diversity in South African

---

[7]It became an official terminology during the Apartheid regime.
[8]Some argue it is as early as the third century (Gradin, 2014).

"white areas" is summarised in Appendix A1. The table contains information on the time of their migration into South Africa, geographical location of original homelands, time of migration outside homelands and the Bantustans assigned to them during Apartheid (which will be explained in constructing our instrumental variable). For example, Zulu are believed to be descended from a leader named Zulu born in the Congo Basin area. In the 16th century, they migrated south and eventually settled in the eastern part of South Africa, an area now known as Kwazulu-Natal. The Zulu empire, in the 1800s helped with their vast migration and expansion of territory.

One indication from the narratives is that Africans had settled in the country long before Europeans arrived. For example, the diaries of shipwrecked Portuguese sailors attest to a large Bantu-speaking population in present-day Kwazulu-Natal by 1552. In 1652 Jan van Riebeeck and about 90 other people set up a permanent European settlement as a provisioning station for the Dutch East India Company at Table Bay on the Cape of Good Hope, beginning the era of European colonisation.

Due to the pressure from the potential conflicts with white colonisers and the other ethnic groups, the nine black ethnic groups began to move out of their homelands or change their territories. By the early 1700s, there were already some African groups migrating into the interior of the country to shield themselves from European domination. By 1750 some white farmers, known as Boers, expanded to the region where they encountered the Xhosa and Zulu. Starting from 1789, a series of wars and conflicts over land and cattle ownership broke out between the Boers and the black ethnic groups. In early 1800s the British replaced the Dutch at the Cape as the dominant force. The Boers, defeated by the British, migrated eastwards into today's Kwazulu-Natal and Free state where the conflicts between the Boers and Zulu people continued. Many other ethnic groups have encountered similar conflicts.

The destination of their migration is not well-documented. This information, however, can be reflected from today's distribution of ethnic groups across South Africa. This pattern of migration will also affect today's distribution of ethnic diversity. For example, a place would be more diverse potentially if more ethnic groups moved in. Details will be shown in the next section. One thing which has to be emphasised here is that in most of the cases the key driving force of emigration from ethnic homelands is mainly the conflict either with the white or with other ethnic groups and less likely to be the economic benefits in the destination.

Importantly, further evidence shows that the mass migration both from central to southern Africa and from homelands to "white areas" took place mainly before the spur of industrialisation and modern economy. The discovery of mineral resources is a milestone in the economic development and transformation towards modern South Africa. Diamonds were first discovered in 1867 along Vaal and Orange rivers, and in Kimberley in 1871. In 1886, gold was first discovered in Witwatersrand, around today's Johannesburg, which stimulated trade and construction (building infrastructure for example) in large dimensions. This timeline confirms that the mass migration largely occurred before the rise of industrial sectors. This means the migration from homelands to "white" areas, although not completely random, may not be purely driven by the higher economic prosperity in the destination which attract more diverse migrants.

In 1910 the Union of South Africa was established, which declared the superior socio-economic status of the white politically and created a white-dominated society. Since then racial discrimination has been a prominent feature of South African society even before the official institution of Apartheid, and the mobility of the black was largely restricted.

Summary demographic statistics about the nine ethnic groups are reported in Table 1.1 for 1996 data and Table 1.2 for 2001 data. The patterns of the distribution of

population share among these nine groups and their labour market outcomes are similar in these two years. In both 1996 and 2001 there are three out of nine ethnic groups (Xhosa, Zulu and South Sotho) who have relatively larger population size (i.e. their share of the whole population is over 20%). We define them as *large* groups. Another two ethnic groups have smaller size (Tswana and North Sotho), and are therefore defined as *medium* groups. The remaining four ethnic groups have much smaller population share (less than 5%) and are defined as *small* groups.

3.2.2 The role of Apartheid in shaping inter-ethnic relations and labour market outcomes

Since mid-1900s, inter-ethnic relationships and labour market outcomes have been significantly shaped by the Apartheid regime and related regulations. The regime reinforced the ethnic identity and destroyed much of the path dependence in the opportunities in education and labour market for the black. The timeline of the Apartheid regime can be found in the lower panel of Figure 3.1.

Starting in 1948, the ruling Afrikaner National Party (NP) implemented a program of *apartness* and formalized a racial classification system, which transformed into official *Apartheid* by the 1951 *Bantu Authorities Act* and 1953 *Bantu Self-Govern Act*. Each individual living in South Africa belonged to one of the four races (White, Indian, Colored, Black), which essentially defined an individual's social and political rights. In addition, the government over-emphasised the differences among the various ethnic groups, in the spirit of the *"dividi et impera"* principle. The ethnic segregation, on top of the racial separation, was aimed at guarantying the political and economic supremacy of the white minority. This exacerbated division of ethnic groups served as a tool for the white to control the black in an easier way (Gradin, 2014).

With the introduction of the *Promotion of Black Self-Government Act* in 1959, the government delimited a number of scattered rural areas as "native reserves" for blacks (called "Bantustans"), one for each ethnic group. The designated areas for the reserves amounted to 13 percent of the total South African territory, while the blacks accounted for more than 75 percent of the total population. Blacks' land ownership was restricted, as well as their ability to freely move and settle in the white South Africa. Internal migration was severely regulated until the repeal of the *Pass Laws Act* in 1986. With the forced removal of the blacks from the "white areas" of South Africa, the Bantustans became over-densely populated territories, where land was overgrazed and afflicted with serious soil erosion. The economic development of these reserves never materialized, leaving their inhabitants in acute poverty (Christopher, 2001). In 1970, the regime promulgated the *National States Citizenship Act*, which provided citizenship to blacks in their homelands. The ultimate aim was to create a number of ethnicity-based independent states.

In conclusion, the Apartheid regime used separation along racial lines and ethnic lines as a fundamental device for the demarcation of physical and social boundaries for all interactions.

One thing which needs to be pointed out is that Apartheid did not shift the big picture of the magnitude and distribution of ethnic diversity in these "white areas", despite the campaign of forced-removal during this time. This is proved by the high correlation of district-level ethnic diversity between 1996 and 1985 (the correlation is 0.918, calculated from 1985 and 1996 census by the authors). Therefore in this paper we link contemporaneous distribution of ethnic diversity to the location of historical homelands without incorporating the forced removal of black people during Apartheid into the story.

The Apartheid regime also severely limited the allocation of job opportunities and resources among the black (Posel, 2001). The *Bantu Education Act* of 1953 en-

sured that non-whites received a substandard quality of education, while access to occupation was regulated by the 1956 *Industrial Conciliation Act.* Whites were authorized to determine the racial allocation of jobs (Mariotti, 2012) and to reserve certain professions, especially in the manufacturing sector, for themselves. In particular, the black were banned from semi-skilled and skilled occupations. Similarly, blacks were not allowed to run their own businesses in white areas. In fact, only with the advent of the democracy, in 1993, non-whites were able to make their free occupational choices. This, together with the reallocation of industries, changed the industrial and occupational structures in white areas, which partly weakened the path-dependence in regional demand of black labour. Moreover, the intergenerational occupational persistence, which has been shown to be particularly relevant for employment (Sørensen, 2007; Pasquier-Doumer, 2012; Magruder, 2010), does not represent a very important issue in the early post-Apartheid era. In other words, blacks may rely more on resources outside their families in helping overcome the entry barriers to jobs (barriers such as information about trade partners and market opportunities, informal credit and insurance arrangement).

3.2.3 Labour market in post-Apartheid South Africa

Labour market prospects for the black actually have worsened since the end of Apartheid in 1994, featured by the rise in unemployment and increased proportion of discouraged workers (Bhorat and Oosthuizen, 2005; Leibbrandt et al., 2009). Based on 1996 census data, over 60 percent of the working-age black population are either unemployed or out of labour force. A large share of the unemployed in 2005 have never worked in their life. To make things worse, skill-biased technological changes lead to an increase in capital-labour ratio in late 1980s and the whole 1990s, further reducing demand for unskilled labour. At the same time, real wage has been stable or decreasing between 1995 and 2005 (Banerjee et al., 2008). The increase

in the supply of unskilled labours, together with the shrinkage in labour demand due to skill-biased technical change as well as the exodus of the white (who are the owners of capital and factories) largely leads to this persistent unemployment issues in the contemporary South African labour market (Banerjee et al., 2008). Furthermore, there is very limited informal employment rate in South Africa, which is only 7.7% - 9.7% based on various measures of informality in September 2004 Labour Force Survey (Heintz and Posel, 2007), possibly because there is also entry barriers in those informal sectors (Kingdon and Knight, 2004). This means the formal wage-employed sector is still the main force in absorbing increased labour supply.

Summary statistics on labour market outcomes based on 1996 and 2001 census data confirm this pattern. In Table 1.1 and Table 1.2, in the overall sample, less than 40% are employed over the whole working-age black population, among which self-employment rate is particularly low (3.2% in 1996 and 2.3% in 2001). The slight rise in unemployment rate from 1996 to 2001 is consistent with the current finding that unemployment rates peaked between 2001 and 2003 in South Africa (Banerjee et al., 2008).

There is, however, large heterogeneity among different ethnic groups. In general groups with medium and small sizes are more active in the labour market and more likely to be employed, both in self-and wage-employed jobs. This indicates that groups with smaller size are in general more active in the labour market and more competitive in job search, which can be explained by the theoretical model later on in the paper.

3.3 Data

For our empirical analysis, we make use of different data sources. We rely on census data for main analysis. There are three years of census data in the Post-Apartheid area: 1996, 2001 and 2011, all of which are the 10% sample from the original national sample in publicly available sources. We do not use 2011 census as both the classification and boundary of magisterial districts have changed dramatically after 2001, making it less reliable to match the new system of magisterial districts in 2011 to the older ones. More importantly, in publicly available 2011 census data, there is no information on which magisterial district each individual resides in.

The unit of analysis is the Magisterial District (MD)[9]. It is particularly convenient to use the MD as a small-scale geographical unit for comparative analysis, given that all other administrative divisions have been revised and re-demarcated repeatedly since the first democratic elections in 1994. It also provides a reasonably large geographical unit to define labour market. Our final sample consists of 210 districts in 2001 census (205 in 1996 census), which were the "white" areas outside the historical homelands. Take 2001 census as an example. The excluded districts are either part of the homelands and thus had distinct political status and partially different laws and labour market regulations (124 districts)[10], or districts where the black population in 2001 accounted for less than 1% of the overall population (11 districts), or they cannot be matched with 1985 census data that is employed in

---

[9]We calculate the ethnic diversity of the magisterial districts where individuals reside in. There are three reasons why we do not use district of work for the main analysis. Firstly, the mechanism we provide in this paper regarding how ethnic diversity affects labour market outcomes is more related to the districts where one resides (i.e. places where one has social interaction even before entering the labour force) than where one works, which we will explain in the theoretical model. Secondly, the correlation between district of work and district of residence are very high so that they provide similar information. Thirdly, more than half of the black population are unemployed or out of labour force. Therefore the information on their district of work is unavailable and has to be replaced by the information on district of residence, making the district-level information among this group and that among the employed people less comparable

[10]The boundary of the homelands does not coincide with the boundary of contemporary MD. Taking a conservative method, we define district with less than 10 % overlap with homelands as "white" districts.

the instrumental variable approach (9 districts)[11].

Status in employment. Using both 1996 and 2001 census data, we construct an individual-level binary variable for employment. The dummy takes value 1 if one is unemployed or economically inactive and 0 if one is employed (either self-employed or an employee). Among workers who are employed, we also consider the allocation of them between self-employment and wage-employment jobs. More in details, an individual is considered to be self-employed if s/he declares to be either self-employed, employer or worker in the family business. To do this, we create another dummy variable only for employed people. It equals 1 if one is self-employed and takes value 0 if s/he declares to be an employee. We only consider working-age black population (15-64 years old).

Ethnicity. The ethnolinguistic group each individual belongs to is identified using the information on the first language they speak in the 1996 and 2001 census. There are nine black ethnic groups in the country: Xhosa, Zulu, Swazi, Ndebele, North Sotho, South Sotho, Tswana, Tsonga, and Venda. Following Desmet et al. (2012), we rely on Lewis' *Ethnologue* tree of ethnolinguistic groups (Lewis et al., 2009) to build our measures of ethnic diversity[12]. For each magisterial district and census year, we calculate the relative shares of each ethnic group within the black population and combine them into ethnic diversity index: the *fractionalisation index*[13]. Universally used in the empirical literature on ethnic diversity (Desmet et al., 2017; Easterly and Levine, 1997; Alesina et al., 2003; Alesina and La Ferrara, 2005), the ethno-linguistic fractionalisation index (ELF) is a decreasing transformation of the

---

[11]OLS regression results remain unchanged if we include the nine districts which cannot be matched with 1985 census data.

[12]The nine black ethnolinguistic groups of South Africa belong to the Niger-Congo language family and correspond to level 11 in the tree of ethnolinguistic groups.

[13]We consider another index: polarization index in the robustness check. It has been proved that fractionalisation index performs better in explaining economic outcomes than polarisation index (Alesina et al., 2003)

Hirschmann-Herfindahl concentration index and is defined as

$$ELF = 1 - \sum_{k=1}^{K} s_k^2$$

where s_k is the population share of ethnolinguistic group k and K is the overall number of groups. Intuitively, the index measures the probability that two individuals who are randomly drawn from the population belong to different ethnic groups. Larger value of the fractionalisation index indicates higher diversity in the magisterial district.

Figure 3.2 shows how ethnic diversity, measured by the ELF index, is distributed in the districts in 1996 in our sample. Districts in darker colours are those with higher ethnic diversity. There is large variation in ethnic diversity levels across South Africa. In general, districts in the northeastern part of the country are more ethnically diverse than those in the southwestern part. In addition, some districts in the middle part of the country are the most ethnically diverse ones. These patterns will be explained when we construct instrumental variables. Districts coloured in white are those inside original homelands, with less than 1% of the black population or cannot be matched to 1985 census data.

Demographic, socio-economic and geographical controls. From the censuses, we also derive a number of controls, which we introduce in our regressions either at the individual level or as aggregated information at the district level. Individual characteristics include gender, age, educational attainment, marital status, whether one's father is alive. Among the district-level controls, we consider population density, proportion of the black, proportion of people working in manufacturing and service sectors, whether the district is mainly rural or urban, and whether there is a river and road crossing the district. Additionally, we introduce other geographical factors, which are particularly relevant to potentially shaping the economic activities of a region. Starting from the Mineral Resources Data Sys-

tem[14], we compute the density of mine for each district. In order to account for the agricultural suitability of land, we use the measure of terrain ruggedness from Nunn and Puga (2012)[15]. As a proxy for the economic development at the local level, we use the National Oceanic and Atmospheric Administration night-time light satellite images data for 1996 and 2001 (Michalopoulos and Papaioannou, 2013)[16]. We also include the number of conflicts in each district.

The rationale of taking into account these control variables will be discussed in the section about empirical model specification. Details on the sources of data and methods in constructing district-level control variables are presented in the Appendix A2.

Before looking into the data, it is worthwhile to point out some differences in information collected in 1996 and 2001 census. Firstly, 1996 census distinguishes between those who are unemployed and out of labour force (i.e. economically inactive) while 2001 census combines these two categories. We thus conduct analysis separately as well as jointly for these two groups in 1996 data, and compare the results based on the joint group with the corresponding results using 2001 census.

Secondly, information on working hours is only available in 2001 census data. We thus focus on 2001 census in calculating hourly income. In addition, a drawback of the income information in the census data in both years is that it asks income from all possible income sources, including labour market income, social grant and other sources like bonus, rent or interest. As a result, another dataset (i.e. Labour Force Survey) is required to calculate more precise measurement of wage, which will be

---

[14]Mineral Resources Data System, MRDS, is a collection of reports describing metallic and nonmetallic mineral resources throughout the world. Spatial data is available at: https://mrdata.usgs.gov/mrds/.

[15]We also tried the measure of slope from the same data source. The results are very similar. We do not include ruggedness and slope at the same time as they are highly correlated (the correlation is larger than 0.9), which potentially leads to multicollinearity issues in regressions.

[16]Night-light data is at 30-second grid level. Here we take the average night-time light density within each magisterial district by summing up the night-light measure over these grids and divide it by area of the district.

discussed in the empirical results.

Thirdly, 1996 census data asks information on both first and second language spoken at home whereas 2001 census only asks people about the first language they speak. Therefore, we only look at 1996 census to test our channel of social skill acquisition using proficiency of a second language as a proxy for social skills.

Fourthly, for migrants in each district, we have full information on the exact year of their migration to the current magisterial districts only in 1996 census. In 2001 census only migration between 1996 and 2001 is recorded. Therefore, in 1996 census data non-migrants are defined as those who either never moved or moved within magisterial districts and migrants are defined based on cross-district migration. In 2001 census non-migrants are those who did not migrate between 1996 and 2001 or migrated within magisterial districts while migrants are people who moved across districts between 1996 and 2001.

Table 2.1 and 2.2 compare districts whose ethnic diversity is above and below the medium level of ethnic fragmentation in 1996 and 2001, respectively. The column "ttest" shows the p-value corresponding to the t-statistics on the difference between districts with high and low ethnic diversity. In both years more diverse places perform significantly better in all indicators of employment, including employment rate, proportion of self-employed people and employees over the whole working-age black population. Among those people who are employed, there is some difference among sectors and occupations. In 1996 census places with higher diversity have larger proportion of people in the manufacturing sector and less in the service sector and this pattern will change once we include our control variables in regressions. Districts with larger ethnic diversity also have less proportion of people in the unskilled occupations among all workers. The similar pattern holds in 2001 census as well.

The negative correlation between unemployment and ethnic diversity at district level is further confirmed in Figure 3.3 where we plot the proportion of unemployed (including economically inactive) people over the whole working-age black population against ethnic diversity in each district. The downward-sloping line between these two variables is observed in both 1996 and 2001.

3.4 Empirical Methodology and Specification

3.4.1 Baseline model specification and potential bias

We study the relationship between ethnic diversity among the black population living in "white areas" of South Africa and their labour market outcomes. In particular, we examine whether the within-black ethnic diversity affects blacks' employment. We start by examining the cross-sectional evidence and investigate the relationship separately for year 1996 and 2001. For both of the years we specify our linear probability model as follows:

$$Empl_{ikdp} = \alpha + \beta ELF_{dp} + \gamma \mathbf{X}_{ikdp} + \delta \mathbf{Z}_{dp} + v_{ikdp} \tag{3.1}$$

where $Empl_{ikdp}$ is a dummy variable for the labour market outcome for individual i of ethnicity k in district d in province p, taking value 1 if one is employed, and 0 otherwise. We also report the results for wage-employment, self-employment (including self-employed, employer and working in the family business) and the substitution between wage-employment and self-employment within the subsample of the employed people. ELF_{dp} takes the value of the within-black index of ethnic

diversity (i.e. fractionalisation index computed in Section 3.3[17]) in district d in province p. X_{ikdp} is a vector of individual-level characteristics (age, gender, educational attainment, marital status, whether one's father is alive). The last one could be a proxy for family financial and non-financial support. Z_{dp} is a set of both time-varying demographic and economic controls as well as time-invariant geographical characteristics at the district level, which will be explained in more detail below.

Unobservables which potentially affect employment rate are included in the term v_{ikdp}. In particular, v_{ikdp} incorporates unobservables at province, ethnicity, district and individual levels, and therefore can be decomposed into the following items:

$$v_{ikdp} = \theta_p + \lambda_k + \sigma_d + e_i + \epsilon_{ikdp} \tag{3.2}$$

ϵ_{ikdp} is the random error term. θ_p is province fixed effect which mainly controls for historical path dependence in job opportunities in each province, as well as province-level fiscal variables including social grant provision and policies on taxation and redistribution. There is also evidence that there is inequality between ethnic groups (Alesina et al., 2016) and that the gaps between different ethnic groups lie in their demographic structure, location, education and labour market outcomes (Gradin, 2014). Therefore we introduce λ_k, ethnic group fixed effects, which allows us to control for mechanical compositional effect and ensures we are comparing individuals from the same ethnic group across districts exposed to different levels of diversity.

Cross-sectional estimates suffer from omitted variable bias originating from σ_d and e_i. For example, the existence of a local economic centre in the district could both create the demand for labour and encourage diversity, in that job opportunities attract individuals from other districts with different ethnic backgrounds. Or more

---

[17]We use the results about polarization index as a robustness check.

energetic individuals with higher work spirits, who are intrinsically more likely to be employed than the average population, may sort to more diverse districts which have more active atmosphere. In these cases, our results will suffer from upward bias as both ethnic diversity and employment rate are positively correlated with the unobserved district and individual characteristics.

To address the concern that the results are driven by these confounding factors, we first include a rich set of district controls Z_{dp} to limit the information in unobserved items. To account for market size effects, we introduce the population density and urban/rural status of the district. As proxies for local economic development, we use the average night-time light density across 30-second grid areas within each district, and the share of blacks in the district population. For the industrial structure of the district which potentially leads to differences in labour intensity of firms, we control for the proportion of people employed in manufacturing and service sectors. Furthermore, to control for the direct spillover from homelands, we include the distance to homelands which were severely deprived by the Apartheid government. To control for the potential cost of ethnic diversity like conflicts, we add the number of violence in each district in the corresponding years, as conflict has been proved to be associated with ethnic diversity (Amodio and Chiovelli, 2017) and potentially job opportunities for the black (for example, there might be more closure of factories in more turbulent districts). Finally, to control for agricultural suitability and other geographic factors relevant for the local economic activities we use the terrain ruggedness, the existence of a river and a road crossing the district and the density of mineral resources.

The remaining district-level omitted variables are included in σ_d. Our results will be biased if they are correlated with employment rates. All this will be dealt with using the instrumental variable discussed later on.

Unobserved information at the individual level in e_i might also bias the OLS result.

We therefore cluster standard errors at the district level to allow for correlation of the error term cross individuals in the same district. Furthermore, as a robustness check, we conduct regressions only on people who are born and remain in the districts (i.e. native people) as well as those who only migrated within districts. If the main results still hold among the native, the potential selection of people moving into places with different levels of diversity based on individual-level criteria will not largely drive the whole story. This will be discussed in more detail in the empirical results.

The relationship between ethnic diversity and labour market outcomes can also be investigated at the district level. Then model (3.1) would change accordingly. $Empl_{dp}$ would represent the proportion of individuals in unemployment, wage employment and self-employment in district d in province p and the ethnicity fixed effect would be removed. The set of individual characteristics X_{ikdp} should therefore be aggregated at the district level (e.g. average education in each district). The district-level regression becomes:

$$Empl_{dp} = \alpha + \beta ELF_{dp} + \delta \widetilde{\mathbf{Z}_{dp}} + \theta_p + \sigma_d + \epsilon_{dp} \tag{3.3}$$

Here $\delta \widetilde{Z_{dp}}$ include both the individual-level variables in X_{ikdp} aggregated at the district level, and the original district-level variables in Z_{dp}. Similarly, after controlling for province fixed effect θ_p, the remaining σ_d is still a source of omitted variable bias which will be dealt with using the same instrumental variable approach.

As individual-level regressions contain more information (especially ethnic-specific characteristics captured by ethnicity fixed effects), we mainly report results based on individual-level regressions in our analysis whilst presenting the results of district-level regressions for robustness check.

3.4.2 Instrumental variable approach

Our instrument for ethnic diversity exploits the historical origins of the location of blacks' homelands. As is explained in the institutional setting, the nine black ethnic groups moved long ago from the northern territories of the African continent and settled in different regions of today's South Africa, with one ethnic group occupying one settlements (i.e. defined as "homelands"). Assume the magnitude of migration from the homelands to outside districts decreases with the distance between them and distance is the only determinant in migration. When they moved out of these homelands to the outside districts (i.e. "white" districts which we are focusing on in this paper), the territories that are equidistant to multiple homelands are more likely to be inhabited by individuals with different ethnic origins, and therefore the ethnic diversity will be the highest. On the contrary, places only close to one homeland and far away from the rest become ethnically homogeneous as they have one group dominant in population size migrating from the closest homeland. Visually, this prediction is confirmed by the distribution of ethnic diversity in South Africa in 1996 (Figure 3.3). As is shown before, places with relatively higher diversity are not necessarily places at the border or close to economic centres of the country, but are those in the middle and northeastern part of the territory surrounded by multiple homelands.

We therefore need an instrument for each district to capture their equidistance to all the original homelands. Our instrumental variable strategy proceeds in two stages. First, similar to Alesina et al. (2015), we estimate a parsimonious gravity model of migration based on 1985 census data (i.e. pre-1994 distribution of ethnic groups). We aim at predicting the level of within-black ethnic diversity in each white district d, solely as a function of a factor that is plausibly exogenous to labour market outcomes of the blacks: the distance of the district to the homelands. Second, we start from the predicted stocks to construct a diversity index. Specifically, we

estimate:

$$N_{dk85} = \alpha + \beta_1 Dis_{dk} + \gamma_k + \epsilon_{dk85} \tag{3.4}$$

where N_{dk85} is the actual stock of individuals belonging to ethnic group k in (white) district d in 1985; Dis_{dk} is the bilateral Euclidian distance between the centroid of district d and the closest border of homeland for ethnic group k[18]; and γ_k is the homeland fixed effect. The determinants in our model are the ones traditionally employed in the related literature (Mayda, 2010; Beine et al., 2013; Ortega and Peri, 2014; Dumont et al., 2010). In particular, the physical distance between two districts (the homelands and the white areas) accounts for the migration costs, while the homeland fixed effects take into account common shocks in living conditions in the original settlement and the stock of population of each ethnic group in homelands, which can also influence migration decision. Following Santos Silva and Tenreyro (2006), we estimate the model by using the pseudo poisson maximum likelihood (PPML) estimator, which better suits the count data in the dependent variable [19].

By imposing a universal β_1 to all ethnic groups, we assume that the per-unit migration cost is the same for everyone, regardless of their ability and ethnicity. In addition, by ignoring any characteristics of the destination (e.g. population size, economic development and job opportunities) in the gravity model, we impose the condition that the benefit of migration is also the same for everyone. Therefore by

---

[18]The reason why we use the centroid of the districts instead of capital city is that capital cities are not well-defined at the magisterial district level. We use the border instead of the centroid of the homeland because the shape of the homeland is highly irregular and scattered. Furthermore, the distribution of population within homeland is highly uneven, making the centroid of homeland a less reliable measure in capturing the distance between the destination and the location of potential migrants from homeland.

[19]We do not control for the population size in the destination in the gravity model as it might be endogenously determined by the level of economic development in the destination which potentially affects the flow of migrants into the destination. Here our aim is not to get the most precise estimate of bilateral migration but to construct the counterfactual number of migrants in each district under a hypothetical setting where bilateral migration is only determined by distance between the original homeland and destination.

construction our predicted number of migrants from each homeland is only determined by the distance between homeland and destination.

In principle, the migration stocks could be predicted by 1996 and 2001 data. Nevertheless, we prefer to use the 1985 census data to rule out the selection of migration resulting from the movements of the black population after 1994 (this happened even as early as the repeal of the Pass Law in 1986). In fact, as previously documented (Section 3.2), while blacks were not allow to choose their place of residence during Apartheid, after 1986 they could freely migrate and decide where to resettle. Therefore, the distribution of ethnic groups in 1985 is less affected by the simultaneous change of labour market conditions and blacks' selection into "white areas" after 1994. Another reason why we use the 1985 distribution of the black population is that the equidistance to different homelands is a feature which stays relatively stable over time. By sticking to 1985 data we can construct an instrumental variable whose value stays the same between 1996 and 2001 to make the IV regression results in these two years more comparable [20].

Using the predicted stocks $\widehat{N_{dk}} = \widehat{\alpha} + \widehat{\beta_1} Dis_{dk} + \widehat{\gamma_k}$, we calculate the predicted share of ethnic group k in the black population of district d and construct the instrument for the fractionalization index ELF:

$$\widehat{ELF} = 1 - \sum_{k=1}^{K} \widehat{s_k}^2 \quad \text{with} \quad \widehat{s_k} = \frac{\widehat{N_{dk}}}{\sum_{k=1}^{K} \widehat{N_{dk}}} \tag{3.5}$$

The same instrumental variable approach with the same model specification at the first stage can be applied to district level regressions.

The remaining challenge is to find a proper measure of the original homelands

---

[20]In reality we do not find much variation in fragmentation index between 1996 and 2001, which means ethnic diversity stays relatively stable over time.

for each ethnic groups. As there is no document about the exact location and boundary of these homelands, we use the territories of Bantustans during Apartheid as proxies for these original homelands. As is discussed in the institutional setting, with the ascent of the apartheid regime, the white-dominated government of South Africa designated specific territories as pseudo-national homelands (i.e. "native reserves", called "Bantustans" in the official documents) for the country's black African population. The Bantustans were organized on the basis of ethnic and linguistic groupings and were a major administrative device for the exclusion of blacks from the "white areas" of South African. The location of the Bantustans is based on the government's knowledge and documents about the historical location of homelands of each ethnic group. Ten Bantustans were created for these nine ethnic groups (there are two Bantustans for Xhosa people - Transkei and Ciskei) and other groups each occupies one Bantustan[21].

To verify that the location and territory of Bantustans can be treated as proxies for the original homelands for the black people, we compare the distribution of these Bantustans and the "Murdock map". This map, drawn by an anthropologist George Murdock in 1953[22], provides the information on what the dominant ethnic group is in each geographical unit on the map of the whole African continent at the end of the 19th century. As reflected in the Murdok's map (panel (a) in Figure 3.4) (each colour represents a certain group dominating the corresponding place in terms of population size), up to the end of the 19th century, each of the nine groups have occupied some specific areas of the country. Although the map does not reveal the location of original homelands and the boundary of the geographical units on the map does not coincide with the border of magisterial districts in South Africa, it

---

[21]Therefore we treat Transkei and Ciskei as one homeland in the gravity model. When we calculate the distance between each district and the original homeland of Xhosa people, we measure the distance between each district and Transkei and Ciskei respectively and choose the smaller one.

[22]The map has been digitized by Nathan Nunn, starting from "Tribal Map of Africa" which is a fold out map from the book "Africa: Its peoples and Their Culture History" by George Murdock, 1959.

roughly implies the spatial distribution of each ethnic groups in South Africa as a joint result of the distribution of original homelands and centuries of emigration from these original settlements.

Comparing Murdock's map in panel (a) and the distribution of Bantustans under the Apartheid system in panel (b) in Figure 3.4, we can find large overlaps of the Bantustans designated to each ethnic group with the region where the same group have dominated historically in Murdock's map. For example, places around the Bantustan designed for Tswana people (the dark green part in panel (b)) are also the places dominated by Tswana people (labeled with the same dark green colour) at the end of the 19th century in Murdock's map in panel (a). Therefore, it is reasonable to use the distribution of Bantustans as proxies for the location of original ethnic homelands.

The map in Figure 3.5 presents the value of predicted diversity index together with the distribution of Bantustan across the country. The white places with slashes are either places which cannot be plausibly considered as "white" South Africa of our interest as they have more than 10% overlap with Bantustans, or places which cannot be matched with 1985 census data. The spatial pattern of predicted value of ethnic diversity in this figure is similar to the distribution of ethnic diversity in Figure 3.3 based on the real data. Again, places with the highest predicted ethnic diversity are those amid multiple homelands (mainly in the middle and northeastern part of the country). A more important feature is that the distance to the closest homeland (proxied by Bantustans) does not completely determine the level of predicted ethnic diversity. That is to say, places with the highest diversity are not necessarily the closest to a particular homeland, which is particularly prominent for the districts around the Bantustans of Transkei, Ciskei, Kwazulu and Bophuthatswana. We will discuss this in more detail in the next section.

Test of validity of the instrumental variable

Identification requires the instrument to capture the ethnic diversity pattern observed in 1996 and 2001 and to be uncorrelated with any other determinants of the blacks' labour market outcomes. The first condition is satisfied provided that: 1) The historical distribution of ethnic groups within the country varies with and is closely related to the distance of the destination region ("white" district) from multiple homelands, and 2) Apartheid did not overturn the historical pattern. As for the second condition, the non-randomness of blacks' homelands could cast doubts on its fulfillment. The proximity to the Bantustans might well be correlated with unobserved factors other than diversity, affecting the blacks' labour market outcomes.

However, the instrument exploits the distance to *multiple* ethnic homelands as a predictor for diversity. As is mentioned above, the map in Figure 3.5 shows that districts with higher predicted diversity are the ones that are "equally" distant to multiple homelands, and not necessarily the ones that are the closest to a specific homeland. For example, although being contiguous to one of the Bantustans - Transkei (identified with the red color in Figure 3.5), districts in the South-East are among the most ethnically homogeneous areas because they are located at the periphery of other homelands. To further ensure that the instrument only captures the relative distance to multiple homelands and not the proximity to a single Bantustan, in the regression we control for the distance to the closest homeland. We argue that, conditional on proximity to a single homeland, the distance to multiple homelands is as good as random.

For a more rigorous test of the validity of our instrumental variable, we run regressions to show that the predicted ethnic diversity index is not correlated with many unobserved district-level characteristics, conditional on all the control variables in

our first stage regressions. Firstly, we test the correlation between the instrumental variable and potential job opportunities. According to agglomeration economics, economic centres, as clusters of economic activities, business and capital inflow, may act as the hub of job creation. Therefore, distance to economic centres may capture the potential job opportunities an individual is exposed to, based on the spillover of economic prosperity from the economic centres. There are five main economic centres in South Africa: Cape Town, Pretoria, Durban, Port Elisabeth and Johannesberg. In the validity test we calculate the distance from the centroid of each magisterial district to the closest economic centre and correlates it with predicted fragmentation index discussed above.

The second potential confounding factor is the economic activity of the white. On the one hand, as the Apartheid regime destroyed the self-employment opportunities, leadership and the training towards skilled occupations of the black in the "white" South Africa, the majority of the employers of wage-employed black people are the white. Although our main regressions focus on the black, the population size and the employment status of the white are also important in determining black people's employment rate, as they might be the providers of potential jobs to the black workers. On the other hand, the dominance and wealth of the white might potentially affect the migration decision of the early black migrants. Black people from different ethnic groups may move to a district where the white behave relatively better as there are more opportunities (or poorer as there is less stress/competition from the white) and thus the ethnic diversity of the black might be correlated with the behaviours of the white. We then calculate the employment rate of the white among their working-age population for each magisterial district in our sample and see if it relates to ethnic diversity of the black.

Thirdly, path dependence also matters in determining contemporary employment opportunities. As the distribution of black settlements is not completely random,

the equidistance to multiple original settlements might reveal some socio-economic characteristics besides the distance itself (i.e. customs, early conflict or the distribution of ancient civilisations) which have long-time impact on contemporary development. This persistence of particular socio-economic features is usually a concern in literature which constructs instrumental variables with geographical variables. However, in our special setting, the Apartheid regime before our sample period compressed the opportunities of education, job opportunities and residential choice nationwide among the black and potentially destroyed part of such historical path dependence. If we can show that the path dependence which potentially correlates with equidistance to homelands was largely destroyed by the Apartheid regime due to the shift in residential patterns and the re-allocation of economic activities both for the black and the white, we will be safer to claim that the historical persistence is not likely to affect contemporary employment opportunities directly. As there is no reliable data to reveal the employment pattern of the black during apartheid, we use the employment pattern of the white in 1980 as a proxy for the remaining path-dependence in employment close to the end of the apartheid and see if it correlates with our instrumental variable measured with 1996 and 2001 data. For the employment status of the white in 1980, we do not consider self-employment as the definition of self-employment is not quite clear under Apartheid regime and therefore has large measurement errors [23]. We also consider the population size of the white in 1980.

The fourth potential confounding factor is the magnitude of migration. It might be the case that a place with higher diversity of migrants is also a place attracting more migrants in magnitude. In other words, if a place attracts a larger migration pool, the composition of migrants is likely to be more diverse. Migrants behave differently from the native in many ways and are more selective in their own. If

---

[23]There are four census during Apartheid: 1960, 1970, 1980 and 1985 census. We only consider 1980 census as the data quality is higher than that in 1960 and 1970 census. Publicly available 1985 census data has no information on employment status.

places close to multiple homelands attract larger number of migrants due to the less migration cost incurred, even if we restrict the sample to the native people in each district, our result may still be biased once there is spillover effect from migrants to the native people. The effect can be positive if the migrants provide job opportunities to the native or it can be negative if these two groups compete for similar positions. Therefore, we need to show that our predicted diversity does not capture the magnitude of migration (calculated as the total number of migrants in each district), but the composition of their ethnicity.

Table 3 shows the results on the validity of the instrumental variable based on 1996 and 2001 census data. We regress a set of variables that potentially affect employment rate on predicted fractionalisation index conditional on all the control variables in the main regressions discussed above. Panel A, B, C and D present the tests on the relationship between predicted ethnic diversity and job opportunities, economic activities of the white, path dependence and magnitude of migrants, respectively. We obtain the coefficients of the tests by regressing the corresponding dependent variables (as reported in the table) on predicted ethnic diversity conditional on all the control variables in the main regression. These dependent variables include: distance to the closest economic centre, proportion of white people who are self-employed over the white population in 1996 and 2001, proportion of white people who are employees over the white population in 1996 and 2001, proportion of white people over the whole population in 1996, 2001 and 1980, proportion of white people who are employees over the white population in 1980 and the number of black migrants in each district. We do not find systematic relationships between these potential confounders and our instrumental variable, which means the predicted ethnic diversity can be considered as a valid instrumental variable.

Other potential threats to the instrumental variable

This section discusses some remaining potential threats to the instrumental variable which are not likely to be measured with available data.

Firstly, one may argue that the original distribution of ethnic homelands is not completely random. The fact that one place is close to multiple homelands at the same time might mean that these homelands are themselves close to each other. Similarly, one possible pre-requisite for a place to be close to only one homeland is that those homelands might be scattered and relatively far away from each other. If the whole region is equipped with better endowments (geography, climate or soil quality) than the others at the time of the Bantu migration from central Africa, this place could attract more than one ethnic groups to establish their homelands, whilst regions with only one ethnic homeland or regions where the distribution of homelands is more scattered might be less attractive in resources and endowments. Therefore, our instrumental variable - the predicted diversity index might just capture the distribution of homelands and the original endowments of the whole surrounding region.

This is not likely to be the case. for the following reasons. The first reason is that our instrumental variable captures the equidistance to different homelands conditional on the distance to the closest homeland. By construction places far away from all homelands can still have reasonably high predicted diversity, as long as it is of equidistance to all these homelands. These places are less likely to be affected by the original endowments and resources of ethnic homelands. The second reason is that we have already controlled for geographical endowments (ruggedness and river) in each district which are potentially correlated with their initial development by affecting their agricultural production. The third reason is that if our instrumental variable mainly captures the initial economic development and the endowments or

resources of the region rather than ethnic diversity, the predicted diversity index should be correlated with the labour market outcomes among both black and white population. However, as is shown in table 3, our instrumental variable is not systematically correlated with the employment rate of white workers. Therefore, it is unlikely that the initial endowments in the regions surrounding ethnic homelands challenge the exclusion condition of the instrumental variable.

Secondly, there is a possibility that districts close to multiple homelands might be the trading centres for people from those homelands whilst trade flows in districts close to only one homeland are less. This might also lead to the difference between these two types of places in the initial economic prosperity and the establishment of cities resulting from trade. Here we show this is unlikely to severely violate the validity of our instrumental variable. Our instrumental variable by construction allows for the case that a place far away from all homelands can be reasonably diverse if it is equidistant to different homelands. And this place is less affected by the initial trade flows among homelands. Furthermore, places with more initial trade flows might become contemporaneous economic centres due to the path dependence in city development and the accumulation of capital and labour. In our validity test we do not find a systematic pattern of the distance to the closest economic centre and predicted diversity index.

Thirdly, one may worry that certain events which attract diverse migrants might happen coincidentally in places close to multiple homelands. For example, the homeland for Tswana group (i.e. the Bantustan of Bophuthatswana) and places in Mpumalanga and Limpopo Province (in the northeastern part of the country) are rich in mineral resources. If our instrumental variable mainly captures the distribution of mineral resources, and if the discovery of mines in a district motivates people of diverse backgrounds to migrate into the district and at the same time boosts economic development, what can be reflected in the predicted ethnic diversity

is mainly the effect of mineral resources. In our analysis we have controlled for the density of the mines in each district. More importantly, narrative evidence reveals that the mass migration from central Africa (which can be dated back to the 11th and 12th century) and the emigration from homelands to "white" South Africa happened well before the discovery of mineral resources (mainly starting from the 19th century). Therefore, the discovery of mines and the related events are not likely to violate the validity of our instrumental variable.

First stage results

Table A0 in the Appendix reports the estimated parameters of the gravity model. It suggests that the distance between a white district d and an ethnic group's homeland is strongly negatively correlated with the size of the same ethnic group's population living in district d. Table 4 presents the first-stage regression of the instrument at the individual level both without and with province fixed effects, together with all control variables. Columns 1 and 2 (3 and 4) report the first-stage regression results based on 1996 (2001) census data. In both years the predicted fragmentation index $\widehat{ELF}$ is positively associated with the observed index ELF. The F-statistics is very high in all regressions (i.e. much larger than 10), indicating that the instrument is a very strong predictor of ethnic diversity.

Comparing column 1 (without province fixed effects) and 2 (with province fixed effects) reveals that the F-statistics decrease drastically from 367.1 to 24.93. Therefore a large part of the variation in predicted ethnic diversity comes from cross-province comparisons. However, even if we control for province fixed effects, there is still remaining variation in predicted diversity index within provinces and the instrument is still a strong indicator of real-world diversity index. Comparison between columns 3 and 4 confirms the same pattern in year 2001.

District-level regressions in Appendix Table A1 reveal the same pattern. Predicted ethnic diversity is positively and strongly correlated with the ethnic diversity index in real data. F-statistics of the instrument are still large in all regressions, and similarly they are larger without province fixed effects. All results consistently show that our predicted ethnic diversity index is strong enough as an instrumental variable.

3.4.3 Supplementary approach: district-level fixed effect

The fact that we have two-year cross-sectional census data and that the territory of magisterial districts stay stable between 1996 and 2001 motivate us to find a way to construct panel data at district level as a supplementary approach to the instrumental variable specification. From the district-level model specification (3.3), we realise that the main source of bias comes from the unobserved σ_d. Therefore an alternative way to instrumental variable approach to deal with this bias is to control for it directly by including district fixed effect based on a panel of districts. Therefore we construct a balanced panel by matching the magisterial districts between 1996 and 2002[24] and conduct the model (3.3) by adding magisterial district fixed effect σ_d directly. Any time-invariant variables in Z_{dp} and θ_p are dropped automatically. Instead we add time fixed effect u_t in the model[25].

---

[24]Among 205 magisterial districts in 1996 and 210 districts in 2001, 205 of them can be matched, given that we exclude districts with less than 1% of black people over the whole population.

[25]A potential further specification is to combine the above two approaches and rely on fixed effect-IV approach. The rationale to do this is that some district-level unobservables might change over time which cannot be captured by time-invariant σ_d. In this case, we have the first difference specification:

$$\Delta Empl_{dt} = \alpha + \beta \Delta ELF_{dt} + \delta \Delta \widetilde{\mathbf{Z}}_{dt} + \Delta f_{dt} + \epsilon_{dt}$$

Ideally we can find an instrumental variable for f_{dt}. A similar case to this specification can be found in Dustmann et al. (2017). However, this first-difference specification at district level with instrumental variable is not appropriate here because there is little variation in both the real-world ethnic diversity and the predicted ethnic diversity (i.e. the equidistance to different homelands does not change over time) over time, which is not sufficient for reliable statistical inference.

$$Empl_{dt} = \alpha + \beta ELF_{dt} + \delta \widetilde{\mathbf{Z}_{dt}} + \sigma_d + u_t + \epsilon_{dt} \tag{3.6}$$

We report the results of this district-level fixed effect model right after the main analysis.

3.5 Empirical Results

3.5.1 Ethnic diversity and labour market outcomes

Ethnic diversity on employment

Table 5 summarizes the main results on the effect of ethnic diversity (measured by fractionalisation index) on unemployment rate. The dependent variable is a dummy which equals 1 if one is unemployed or out of labour force and 0 otherwise (including people who are self-employed and employees). In 1996 census data which distinguishes people who are unemployed and out of labour force, we create dummies for unemployment and labour force participation and look at how they respond to ethnic diversity separately. Columns 1-6 report the results in year 1996 while columns 7-8 are for year 2001 when unemployed workers and people out of labour force are combined into one category in the original census data. Furthermore, panel A in Table 5 reports the results based on the cross-sectional OLS regressions at the individual level. Panel B in Table 5 provides the corresponding estimates based on the instrumental variable regressions. We provide results both without and with province fixed effects for comparison. All regressions control for the individual and district level characteristics including ethnicity fixed effects discussed above.

In most of the OLS and IV regressions in Table 5 the coefficients of ethnic diversity on unemployment (or labour force participation or these two outcomes altogether) are significantly negative, indicating that within-black diversity increases the rate of employment and labour force participation. Comparing panel A and panel B, the negative and significant coefficients of ethnic diversity remain in IV regressions in many columns. In panel B, comparing columns 2, 4 and 6 reveals that ethnic diversity increases employment mainly by decreasing the number of people who are actively looking for jobs but still unemployed, rather than bringing people into the labour force. Table 5 also shows that in most of the regressions the coefficients decrease after controlling for province fixed effects, meaning province-specific features can partly explain the response of employment rate to ethnic diversity.

We now calculate the magnitude of the effects of ethnic diversity on employment based on the results in columns 6 and 8. In panel A in column 6, one standard deviation increase in ethnic diversity index in 1996 is associated with 0.0215 point decrease in unemployment (including inactivity), which is 3.5% of the average unemployment (including inactivity) rate[26]. Similarly, in panel A in column 8, one standard deviation increase in ethnic diversity index in 2001 is associated with 0.0388 point decrease in unemployment (including inactivity), which is 6.1% of the average unemployment (including inactivity) rate[27]. Correspondingly, in IV regressions, one standard deviation increase in ethnic diversity index in 1996 (2001) decreases unemployment (including inactivity) by 0.0262 (0.044) point, which is 4.27% (6.92%) of the average unemployment (including inactivity) in 1996 (2001).

---

[26]It can be calculated that the standard deviation of ethnic diversity in 1996 is 0.2659. The coefficient of diversity index in panel A in column 6 is -0.081. Therefore one standard deviation in diversity index decreases unemployment by 0.081 * 0.2659 = 0.0215. From Table 1.1 we know that the average unemployment (including inactivity) rate among the black in "white" districts is 0.613. Therefore this point decrease is 0.0215/0.613= 3.5% of the average unemployment rate.

[27]It can be calculated that the standard deviation of ethnic diversity in 2001 is 0.2586. Therefore in 2001 one standard deviation in diversity index decreases unemployment by 0.150 * 0.2586 = 0.0388. From Table 1.2 we know that the average unemployment (including inactivity) rate among the black in "white" districts is 0.636. Therefore this point decrease is 0.0388/0.636= 6.1% of the average unemployment rate.

Comparing the magnitude of estimates in OLS and IV regressions in both years shows that the magnitude of the effects of ethnic diversity on employment rate increases largely between 1996 and 2001 (from 3.5% of the average unemployment rate to 6.1% in OLS and from 4.27% to 6.92% in IV) and IV estimates are slightly larger than OLS estimates. This can be explained by the fact that IV regressions capture LATE for workers at the margin of being affected by ethnic diversity. They might be the most responsive to ethnic diversity in considering their employment status.

Appendix Table A2 further breaks down employment status into two categories: self-employment and wage-employee. All the independent variables remain the same as those in Table 5. In columns 1 and 3 in Appendix Table A2, the dependent variable is a dummy which equals 1 if one is self-employed and 0 otherwise (including unemployed, inactive and wage employee). The dependent variable in columns 2 and 4 is a similar one which equals 1 if one is an employee and 0 otherwise. Again, panel A (B) reports the results for OLS (IV) regressions.

The results show that in the post-apartheid South African context, within-black ethnic diversity has a positive effect on the labour market outcomes of the blacks, mainly in wage-employment as is shown in columns 2 and 4. Specifically, one standard deviation increase in the fractionalisation index is associated with a 0.0226 (0.037) point increase in the wage-employment rate of the working-age black individuals in 1996 (2001), according to the OLS results. This corresponds approximately to a 6.4% (10.85%) increase of the average wage-employment rate among the population of reference in 1996 (2001). In IV regressions, one standard deviation increase in the fractionalisation index increases wage-employment rate by 0.027 (0.047) points in 1996 (2001), which is around 7.6% (13.78%) increase of the average wage-employment rate in 1996 (2001).

Similar to the patterns in Table 5, the effect of ethnic diversity on wage-employment

increases from year 1996 to 2001. IV estimators have slightly larger magnitude than OLS estimators for possibly the same reason. We do not find anything significant about self-employment rate. One plausible reason is that there is no enough variation in self-employment rate across districts for reasonable statistical inference as the self-employment rate in South Africa is very low in both years (2% - 3%) according to Table 1.1 and 1.2.

Table 6 further presents how ethnic diversity affects workers' choice between self-employment and being an employee. As self-employment rate is between 2% - 3% of the whole working-age black population, we drop self-employed people from the whole sample and investigate if ethnic diversity increases the probability of being an employee against unemployed in columns 1 and 3. The magnitude and significance of the coefficients on ethnic diversity index are very similar to those in the corresponding columns (columns 2 and 4) in Appendix Table A2. This shows that most of the effects of ethnic diversity on employment takes place in wage-employed jobs.

Columns 2 and 4 only include employed people and look at the allocation of these workers between self- and wage- employment. The dependent variable equals 1 if one is self-employed and 0 if being an employee. This is to investigate the effect of ethnic diversity on the potential substitution between self- and wage-employment among employed black population. We replicate the results of the main analyses by restricting the sample to people who are either wage-employed or self-employed (i.e. excluding the unemployed and the inactive). Although the self-employment rate might be too low for enough variations to generate significant statistical inference, we find that the coefficients of ethnic diversity are consistently negative in OLS and IV regressions in both years. That is to say, ethnic diversity helps unemployed individuals get into employment; a large fraction of those newly employed people opt for working for others as an employee.

The corresponding district-level regressions based on the model specification 3.3 are reported in the Appendix Table A3. In these district level regressions, the dependent variables are the proportion of working-age black people who are unemployed or inactive; who are wage-employed; who are self-employed and the proportion of people who are self-employed relative to employees (columns 1-4 and columns 5-8, for year 1996 and 2001 respectively), given the corresponding individual features aggregated at district level and district level controls. OLS (IV) estimators are shown in panel A (B).

The OLS and IV estimates reported in Table A3 confirm the positive impact of diversity on the employment of the blacks. And this positive impact mainly takes place in wage-employment. The effect on employment (and wage-employment) in OLS regressions is slightly smaller than the ones estimated with the individual-level regressions, while the magnitude of the effect in IV regressions is slightly larger than that in individual-level regressions[28].

Ethnic diversity on wage, income and working hours

In this section we replicate the above individual-level regressions (both OLS and IV) by replacing the dependent variables with other labour market outcomes, including working hours, hourly wage and monthly earnings. As information on working hours is only available in 2001 census data, we only conduct these analyses based on 2001 data. For data on working hours, if values of self-reported weekly working hour are larger than 80, we treat them as outliers and exclude them from regressions. In addition, we trim the income data by excluding values above 5 standard deviation of the mean income. Hourly wage is constructed by dividing monthly earnings by

---

[28]Columns 4 and 8 report the results on the effect of ethnic diversity on the rate of self-employment relative to wage-employment aggregated at district level by only including black people who are employed. Results in other columns are based on the whole working-age black population

monthly working hours (i.e. 4*weekly working hours).

Data on monthly income in 2001 census includes both labour market earnings and income from other sources such as dividend, rent or social grant. We first report the results based on these rough measures of monthly earnings and replicate the regressions with more precise data on labour market earnings and working hours.

Panel A in Table 7 reports the OLS and IV regression results on these labour market measures based on 2001 census data. Dependent variables include: log monthly income, log hourly income and weekly working hours. As self-employed workers and employees have very different determinants of working hours and earnings, and that ethnic diversity mainly increases wage-employment rate, we only focus on employees in all regressions[29]. Columns 3 and 6 indicate that ethnic diversity does not affect weekly working hours among the employees. Therefore the increase in employment in response to ethnic diversity comes from the extensive margin by increasing employability of unemployed and inactive people, rather than the intensive margin (measured by weekly working hours). And this extension of the extensive margin of labours is not achieved at the sacrifice of decreased intensive margin.

Columns 1, 2, 4 and 5 show some evidence on the increase in both monthly and hourly income among the black employees in response to higher ethnic diversity. As is stated above, information on income in census data incorporates all potential income sources. Therefore we need another dataset which asks information on labour market earnings in particular. We turn to October Household Survey 1996 to replicate all the results in Panel A [30]. We do not choose year 2001 because starting from year 1998 there is no information on the magisterial districts each

---

[29]There are more observations in columns 3 and 6 than others because there are missing values in income and we trim the income values above 5 standard deviation from the mean.

[30]It is an annual survey staring from 1993 (which was renamed as Labour Force Survey conducted twice a year from 2000 and became a quarterly survey from 2008). In 1996 survey 72890 individuals are covered, among which 16082 have information on work status.

individual lives in. The results are in Panel B in Table 7. Columns 3 and 6 confirm
that weekly working hours are not responsive to ethnic diversity. In columns 1, 2, 4
and 5 the effects of ethnic diversity on measures of labour market earnings are not
significant, possibly because the increase in employment can come from both the
supply and demand side of the labour market, or because the measures of nominal
earnings are not adjusted for price levels (as there is no price or living cost data at
the magisterial district level).

3.5.2 Supplementary approach: district-level fixed effects

As a supplementary approach to the instrumental variable approach, we provide
estimation results on district-level fixed effects models based on the model specifi-
cation (3.6) in Table 8. We construct a balanced panel between 1996 and 2001 (205
magisterial districts each). The measures of labour market outcomes (i.e. depen-
dent variables) are: proportion of people who are unemployed or inactive among
the whole working-age black population; proportion of employed workers among the
whole working-age black population (excluding self-employed people); ratio of the
number of self-employed workers versus employees and log monthly income among
employees.

Similar to the main IV regression results, higher ethnic diversity is associated with
higher employment, mainly in wage-employment but there is no significant corre-
lation between ethnic diversity and monthly income. In particular, in district fixed
effect regressions we find some evidence that more diverse districts are associated
with higher ratio of wage-employment in relation to self-employment.

The magnitude of coefficients in Table 8 are larger than those in Table 5 and Table
6, which can be explained by two possible reasons. Firstly, district-level regres-
sions do not include ethnicity fixed effect which is used to capture ethnicity-specific

unobservables which affect the labour market outcomes of each ethnicity such as the attitudes towards work and leisure and ethnic-specific skills. It is however not appropriate to include this fixed effect in the district-level regressions due to the potential multicollinearity problem, as the proportion of each ethnic group in a district is already a component of the ethnic diversity index (i.e. an item in Herfindahl Index).

Secondly, the relatively larger coefficients of panel regressions might reflect some time-varying district-level unobservables. For example, people are more likely to move to ethnically diverse districts as time goes by as a result of increased benefits in the destination (i.e. the economy of the districts with higher ethnic diversity might growth more rapidly than that in more homogeneous districts). In individual-level IV regressions, our instrumental variable is not likely to be correlated with the economic development in the destinations by construction (as the distance between homelands and destination is the only determinant in migration). Therefore the variation of these unobservables over times does not affect our estimates in IV regressions. However, as panel regressions with district-level fixed effects may lead to upward bias of the key estimator as they do not take into account these time-varying unobservables.

3.5.3 Heterogeneous effects of ethnic diversity on employment

Table 9 split the whole sample into several sub-samples to investigate the heterogeneity in the impact of ethnic diversity on labour market outcomes with individual-level regressions. In particular, we replicate the regressions in the main specification by carrying out the same analysis on these sub-samples. By excluding workers who are self-employed, we use a dummy dependent variable which takes the value 1

if one is an employee and 0 if one is unemployed or inactive[31]. Panel A and B in Table 9 replicate the same regressions in columns 1 and 3 in Table 6 by splitting the working-age black population into sub-samples. Panel C and D look at the allocation of employees among different sectors and occupations in response to ethnic diversity by regressing the probability of working in particular sectors or occupations on ethnic diversity index only among employees.

Panel A split the sample by educational levels. "High education" refers to people with more than 9 years of schooling (i.e. high school, college and postgraduate) while "low education" means no education, primary and junior high school education. We present both OLS and IV results in both years[32]. In 1996 the positive and significant effect of ethnic diversity on wage-employment rate only exists among low-educated working-age black population. The magnitude of the coefficients of ethnic diversity index is also larger among the low-educated group. In IV regressions in 2001 the positive effect of ethnic diversity still only holds for low-educated people. However, there is some difference in its magnitude between 1996 and 2001. From 1996 to 2001 the magnitude of the coefficient of ethnic diversity index increases largely from 0.05 to 0.12 for high-educated people while for low-educated people the increase is smaller (from 0.141 to 0.19). A more detailed split of the sample reveals that the increase in the magnitude of the effect of ethnic diversity on wage-employment rate takes place only among college graduates while for high-school graduates the coefficient is insignificant and the magnitude is still around 0.05 (results not shown in the table though).

Panel B split the sample by group size. As is shown in Table 1.1 and 1.2, we have three "large" groups whose population share is above 20%, two "medium" groups

---

[31]We also conduct the analysis with a dummy on whether one is unemployed (including inactive people) or not. The results are quite similar.

[32]The results are robust to other definitions of "high" and "low" educational categories. For example, we also split the sample into people with more and less than 7 years of schooling, and people whose years of schooling are above and below the mean value in the district where they live.

whose share is between 10% and 20% and the remaining "small" groups making up less than 5% of the whole black population. We look at these three groups separately and discuss how they are affected by ethnic diversity. The results reveal that only the group with "large" size are positively affected by diversity. None of the columns show that "small" groups response to ethnic diversity of the districts they live while evidence on the "medium" group is more mixed. It is not very likely that the results are purely driven by the lack of power of statistical inference due to smaller sample size. In all the regressions for "medium" and "small" groups, the t-statistics is far from being large enough to generate significant inference. Furthermore, in some regressions the coefficients of ethnic diversity are negative, especially for those in the "small" group in 1996.

Focusing only on the black people who are employees, we can investigate the allocation of these workers among different sectors. Both 1996 and 2001 census data provides information on the industrial sectors they work, which we classify into agriculture, manufacturing and service sectors. Panel C presents the results on this allocation. There is some evidence that ethnic diversity decreases people's chance of working in the manufacturing sector, conditional on all the control variables we have. This further confirms the idea that the employment opportunities generated from ethnic diversity are not purely driven by the expansion of manufacturing sector due to the revolutionary events like the discovery of mines.

We study the allocation of employees further by looking into occupations. In both 1996 and 2001 census for each worker there is information on the occupation classified into a detailed 3-digit code. We aggregate this 3-digit coding system into types of occupations based on their skill levels: manager, professional, clerk, service worker, craft worker, skilled worker in agricultural sector, machine operator and unskilled worker. The dependent variables in Panel D are dummies on whether one works in one of these occupations. According to the regression results, ethnic

diversity decreases people's chance of becoming a machine operator and increases their probability of being a manager, professional employee and clerk. One common feature is that occupations such as manager, professional and clerk require more language and social skills while the demand for social skills is the least among machine operators. This is closely linked to our mechanism through which ethnic diversity influences labour market outcomes, which will be discussed in the modelling part.

3.5.4 Robustness check

We conduct a series of robustness checks in this section to consolidate the result that ethnic diversity increases employment rate among working-age black population.

Firstly, we provide some further evidence on the argument that our result is not purely driven by the sorting of migrants. That is, we show that the positive correlation between ethnic diversity and labour market outcomes does not purely come from the migrants with higher abilities moving to more diverse places and therefore are performing better in job searching. We divide the whole working-age black population into three sub-samples with different levels of sorting: people who were born and stay in the district or people migrating within districts (i.e. "native" people); people moving across districts (i.e. "migrants"); immigrants moving from other countries ("immigrants")[33]. In Table 10 we run the same IV regressions[34] as those in the main analysis separately for these three groups in 1996 and 2001. The dependent variables include a dummy on whether one is unemployed and a dummy on whether one is an employee (excluding self-employed workers).

Columns 1 and 4 show that in both years ethnic diversity positively affects the labour market outcomes for native people who are the least likely to sort to places

---

[33]Note that "migrants" and "immigrants" in 2001 census data are those who move across districts or countries between 1996 and 2001, whereas in 1996 census they are the people whose last migration was across districts or countries.

[34]OLS regressions have very similar results. We only show the results about IV regressions here.

with higher ethnic diversity, as they were born in these districts and remained there, or moved within districts. The positive effect of ethnic diversity on employment also exists among immigrants in columns 3 and 6, the mostly selected sample based on ability and preference (although the number of immigrants in South Africa belonging to one of the nine ethnic groups is very small compared with the whole black population). Interestingly, there is no effect of ethnic diversity on employment among migrants across districts. As we discussed in the validity of the instrumental variable, there are two potential mechanisms of selection among migrants. Either the selection occurs in the original place, meaning people with higher ability choose to move out; or the selection takes place at the destination, meaning people sort to places with higher economic prosperity or job opportunities or more socially active environment when they decide where to move. The result about cross-district migrants here might suggest that the first selection mechanism is more important - migrants are of higher ability and therefore behave better wherever they end up, which indicates that the relationship between ethnic diversity and employment is not solely driven by the selection of destinations.

Another potential threat to the interpretation of our results as illustrating a positive impact of ethnic diversity on employment is the emigration of the white after the end of Apartheid. It has been observed that there has been a large emigration of the white out of South Africa after 1994 and that white people moved out of the country for the fear of the worsening economic conditions, weaker government capacity, or the revenge from the black after the nightmare of Apartheid. A place has higher within-black diversity might just indicate that the power of the white is weaker in these places (so that the black community can grow and attract people with a diverse background). If this is the case, there would be more white people emigrating from South Africa in a district with larger ethnic diversity index. The mass emigration of the white may lead to many job vacancies to be filled by black workers, consequentially improves the job opportunities of the black. If this story

is true, the correlation between ethnic diversity index and employment rate in a district cannot reflect the impact of ethnic diversity as ethnic diversity index here is just a proxy for the power of the white in the district.

We therefore regress the number of the white in 1996 and 2001 respectively and the difference in the number of white residence between 1985 and 1996 (or 1985 and 2001) on ethnic diversity index for each district, using the same set of control variables. We find in Table 11 that the ethnic diversity index is associated with neither the absolute number of the white population nor the difference in the white population before and after the end of Apartheid (which captures the emigration of the white). This confirms that ethnic diversity is positively related to employment not simply because these places have more job vacancies left by the white people who emigrated from the country.

Secondly, we use non-linear econometric methods to estimate the main regressions. Given that our outcomes are measured by binary variables, we replicate our results by estimating a logit model, a probit model and a probit model with the instrumental variable in both 1996 and 2001. Results are summarized in the Appendix Table A4. Marginal effects at average ethnic diversity index are reported in all columns. The positive effect of ethnic diversity on both employment as a whole and wage-employment in particular (excluding self-employed people in columns 4 - 6) is robust to these specifications. The magnitude of the marginal effects is very similar to those in Table A4 and Table 6 in baseline regressions. For example in logit regressions in 2001, the coefficient of ethnic diversity on unemployment is -0.152, which is roughly the same as the corresponding coefficient in OLS regressions in Table A4 (-0.150 in column 7). In IV regressions the magnitude in non-linear models is smaller than that in linear IV models but the significance remains the same. For example, in probit regressions with our instrumental variable based on 2001 census data, the coefficient of ethnic diversity on unemployment is -0.148 while

in the corresponding IV regression it is -0.171 (column 8 in Table A4).

Thirdly, we check if the main results are robust to an alternative measure of ethnic diversity. Apart from the fractionalisation index, some literature uses the polarization index (P) such as Montalvo and Reynal-Querol (2005). The index captures the deviation of the distribution of the ethnic groups from the bipolar distribution (which represents the highest level of polarization). Following the notations in defining fractionalisation index, the index is computed as:

$$P = 1 - \sum_{k=1}^{K} \left(\frac{1/2 - s_k}{1/2} \right)^2 s_k$$

The fractionalisation and polarisation index are highly correlated at low levels, while being uncorrelated and negatively correlated at intermediate and high levels, respectively (Montalvo and Reynal-Querol, 2005).

We use the same "equidistance" measure as an instrumental variable for ethnic diversity here. Following the same approach as that for fractionalisation index, we use predicted polarisation index obtained from the predicted stock of ethnic groups in each district as an instrumental variable for real polarisation index. After getting predicted population share of each ethnic group $\widehat{s_k}$ in each district based on the gravity model 3.4, we get the predicted polarisation index:

$$\widehat{P} = 1 - \sum_{k=1}^{K} \left(\frac{1/2 - \widehat{s_k}}{1/2} \right)^2 \widehat{s_k}$$

We use this predicted polarisation index as an instrumental variable for P from real data and conduct both OLS and IV regressions. We report the first-stage outcomes

in Appendix Table A5.1 and the individual-level regressions in Appendix Table A5.2. First-stage regressions show that the predicted polarisation index is a strong indicator of the real polarisation index, although the instrumental variable is less strong with province fixed effects.

Appendix Table A5.2 reports the results of the polarisation index on both employment rate in general and wage-employment rate in particular (again excluding self-employed people). The effect of polarisation index is strong and positive in all columns without province fixed effects while the corresponding effect with province fixed effects is weaker. As the polarisation index not only reflects the diversity of ethnic groups but is also weighted by the relative group size, it is harder to interpret the labour market outcomes in response to polarisation index than that to fractionalisation index.

Summary of empirical results. The whole empirical section above explains and consolidates the following results which are the basis for the theoretical model in the next section:

1. Ethnic diversity increases employment among the working-age black population and this mainly takes place in wage-employed jobs.

2. The positive effect of ethnic diversity on employment can only be observed among the ethnic groups with relatively large size.

3. Ethnic diversity affects low-educated working-age black people more than better educated ones.

3.6 How Does Ethnic Diversity Affect Employment: A Theoretical Model and Mechanism

We propose a plausible theoretical framework consistent with our empirical findings above to explain the positive effects of ethnic diversity on employment and the heterogeneity of the effects across sub-groups. More specifically, we focus on social skill investment which increases with ethnic diversity.

3.6.1 A plausible theoretical framework

The story is as follows. Assume that inter-ethnic communication requires more skills than intra-ethnic interaction. In a more diverse place, the necessity to communicate with individuals from different ethnic groups may motivate people to learn and practise more social skills. The acquisition of this extra skill, which is helpful in reducing coordination costs or increasing labour productivity (which we will discuss later on), could make individuals more competitive and increase their chances of finding jobs.

In more detail, people obtain utility from interacting with others. Establishing a relationship with someone from a different ethnic group requires more skills than that within the same ethnic group (this may be due to cultural barriers like language) between ethnicity. In a more ethnically diverse place people have to communicate with a larger proportion of individuals outside their own ethnic group to maintain a certain level of social connection. Therefore they put in more efforts in developing social skills as long as the benefit of interacting with a different ethnic group outweighs the cost of learning efforts. Social skills here can be of many types, including both cognitive skills like language and non-cognitive skills like communication skills or social attitudes. When these people enter the labour market, these skills are ben-

eficial to their labour market performance, in addition to human capital investment. We will explain this in more detail below.

What needs to be emphasised here is that more ethnically diverse places do not necessarily have more social interaction in general but the overall investment in social skills should be higher because a larger proportion of social interaction comes from inter-group connection and inter-ethnic interaction requires more skills than intra-ethnic communication.

The distinction between social connection and investment in social skills is analogue to the literature which differentiates social connectedness and network formation (Chay and Munshi, 2015). Their stories implies that there exists a threshold only above which social connectedness and network-based outcomes are positively correlated. Similarly, in our story, the level of social connection can be high in both ethnically homogeneous and diverse places, but investment in social skills is only high when a large proportion of this social connection takes place between ethnic groups as intra-ethnic communication is relatively costless.

Model setup

We provide a model of a coordination game to explain the mechanism. We assume that individuals gain utility from social interaction at the cost of investing in social skills. As the cost of communicating with a different ethnic group is larger than that with the same group, we normalise the cost of communication within each ethnic group to be 0 and set the costly investment in social skill for inter-ethnic interaction to be c per unit. We assume that the amount of investment in social skills x_i also equals the output of the investment (i.e. the amount of skills acquired). We have the following setup of a coordination game:

Players. Each group only differs in terms of their population size. Suppose there

are m ethnic groups in total. We denote these different groups as m different sets $N_1, N_2, \ldots N_m$, each with a group size n_k and $k = 1, 2, \ldots m$. The overall population in each district is N, so that $\sum_{k=1}^{m} n_k = N$.

Strategies. Each individual invests x_i in social skills. For simplicity we assume x_i is a binary variable which equals 1 (0) if i invests (does not invest)[35]. One can only participate in inter-ethnic social interactions if he invests in social skills. The total amount of people each individual i in ethnic group N_k with a group size n_k has access to in the inter-ethnic communication is calculated as $x_i \sum_{j \neq k} \sum_{q \in N_j} x_q$. There is complementarity between i's own investment in social skills and the overall investment level of people outside group k. If $x_i = 0$, i cannot benefit from social interaction even if everyone outside his group invests in social skills[36]. Therefore the total number of people interacting with i (both inside and outside his own group) can be calculated as $n_k + x_i \sum_{j \neq k} \sum_{q \in N_j} x_q$.

Utility. Individual i belonging to group k obtains utility from social interaction which depends on the size of his own groups n_k and the number of people he can reach in other ethnic groups, the latter relying on both his own investment in social skills and the efforts from other ethnic groups. The utility from overall social interaction is written as $f(n_k + x_i \sum_{j \neq k} \sum_{q \in N_j} x_q)$, which is assumed to be increasing at a diminishing rate. That is, $f' > 0$ and $f'' < 0$. The implication is that utility from social interaction increases as more people participate in communication, but this has a diminishing return as people get tired from social life when the number of contacts increases. We can thus write the net utility U_{ik} from overall social interaction for individual i in group k as follows:

---

[35]One can potentially treat x_i as a continuous variable or make x_i heterogeneous in communicating with different ethnic groups. For example, similar to Akerlof (1997), we can introduce the investment of x_{ij} if individual i is interacting with group j, and x_{ij} is a decreasing function of social distance between groups i and j. However, this binary setting of x_i is already enough to explain the key empirical findings about ethnic diversity discussed above.

[36]This is a reliable assumption in our setting as we later on use proficiency of English/Afrikaans as the second language as a proxy for social skills. One can communicate with people from another ethnic group only if both learn a second official language.

$$U_{ik} = f(n_k + x_i \cdot \sum_{j \neq k} \sum_{q \in N_j} x_q) - c x_i \qquad (3.7)$$

Equilibrium. In this paper we focus on pure strategy Nash equilibrium. Clearly the coordination game has multiple equilibria. For example, $x_i = 0, \forall i$ is a Nash equilibrium. This is because starting with this initial condition, no one has the incentive to deviate. In more detail, for an individual i in group k, his utility from social interaction is:

$$U_{ik} = \begin{cases} f(n_k) - c, & \text{if } x_i = 1 \\ f(n_k), & \text{if } x_i = 0 \end{cases}$$

Therefore individual i always gets higher utility by not investing in social skills. That is to say, in order for the social interaction to happen, there might be some initial efforts to stimulate communication.

As it is not possible to conduct comparative statics across different Nash equilibria in this setting, we only focus on the equilibrium where the number of individuals investing in social skills is as large as possible, and see how the equilibrium state changes in response to group size. By doing this, we can demonstrate how the maximum possible level of skill investment changes with ethnic diversity.

One important feature of this particular equilibrium is that to guarantee the maximum participation in inter-ethnic communication, individuals always choose to invest in social skills unless the net utility from doing so is strictly smaller than that from deviating. In other words, even if the individual is indifferent between investing and not investing, he will always choose to invest in social skills.

Social interaction, skill acquisition and distribution of group size

In this game, player i from group k chooses either $x_i = 1$ or $x_i = 0$ to maximise his total utility from social interaction, given the population size of each ethnic group as well as the investment of x among people outside group k. We derive two lemmas before proceeding to the effect of diversity on investment in social skills.

Lemma 3.1. *People from the same ethnic group choose the same amount of investment.*

Proof. Suppose player 1 and player 2 both come from ethnic group k with group size n_k. Without loss of generality we assume $x_1 = 1$ and $x_2 = 0$. We focus on the pure strategy equilibrium with the maximum number of skill investment. As both 1 and 2 maximise their utility from social interaction, we have:

$$\begin{cases} f\left(n_k + \sum_{j \neq k} \sum_{q \in N_j} x_q\right) - c \geq f(n_k), & \text{for player 1} \\ f\left(n_k + \sum_{j \neq k} \sum_{q \in N_j} x_q\right) - c < f(n_k), & \text{for player 2} \end{cases}$$

Clearly these two inequalities contradict each other. Therefore we must have $x_1 = x_2 = 1$ or $x_1 = x_2 = 0$. $\square$

Based on this, we have lemma 3.2:

Lemma 3.2. *People from different groups will choose the same amount of investment as long as the population size of these groups is the same.*

Proof. Suppose player i and player j come from ethnic group k and l, and $n_k = n_l$. Without loss of generality we assume $x_i = 1$ and $x_j = 0$. According to lemma 3.1, everyone from group k (l) chooses $x_i = 1$ ($x_j = 0$). As both i and j maximise their utility from social interaction, we have:

$$\begin{cases} f(n_k + n_l \cdot 0 + \sum_{p \neq k, p \neq l} \sum_{q \in N_p} x_q) - c \geq f(n_k), & \text{for player } i \\ f(n_l + n_k \cdot 1 + \sum_{p \neq k, p \neq l} \sum_{q \in N_p} x_q) - c < f(n_l), & \text{for player } j \end{cases}$$

When $n_k = n_l$, these two inequalities hold altogether if and only if $f(n_k + \sum_{p \neq k, p \neq l} \sum_{q \in N_p} x_q) - c > f(n_k + n_k + \sum_{p \neq k, p \neq l} \sum_{q \in N_p} x_q)$ for each possible x_q in group p. As $f' > 0$, $n_k \geq 0$, $c > 0$, this inequality cannot hold.

Therefore we must have $x_i = x_j = 1$ or $x_i = x_j = 0$. $\qquad\square$

Combining lemma 3.1 and lemma 3.2, we can link the size distribution of ethnic groups and social skill investments. To guarantee the maximal level of skill investment in equilibrium, we start with the initial condition where $x_i = 1, \forall i$ and study people's incentive to deviate from this condition.

We can derive the following two propositions.

Proposition 3.1. *Social skill investment increases with the number of different ethnic groups in a district.*

Proof. Consider a symmetric case where each group has the same groups size. In this case for any ethnic group k, we have $n_k = \frac{N}{m}, \forall k = 1, 2, \ldots m$. According to lemma 3.1 and lemma 3.2, everyone has the same social skill investment x, regardless of his ethnic group.

We can re-write the utility function of social interaction for an individual i in any ethnic group in the following way:

$$U_{ik} = \begin{cases} f(\frac{N}{m} + (N - \frac{N}{m})) - c, & \text{if } x_i = 1 \\ f(\frac{N}{m}), & \text{if } x_i = 0 \end{cases}$$

For $x_i = 1, \forall i$ to be a Nash Equilibrium, no player is going to deviate by choosing $x = 0$ instead. Suppose c satisfies $c < f(N)$, we have:

$$f(N) - c \geq f\left(\frac{N}{m}\right)$$

Since $0 < f(N) - c < f(N)$ and $f' > 0$, there exists a fixed n^* such that $f(N) - c = f(n^*)$. Given $f' > 0$ and $f(n^*) \geq f(\frac{N}{m})$, we have:

$$m \geq \frac{N}{n^*} \tag{3.8}$$

Therefore, the larger the m is, the more like the Nash Equilibrium $x_i = 1, \forall i$ will be maintained. $\qquad \square$

Fixing the total number of ethnic groups in a district, the even (uneven) distribution of these groups may also affect social skill acquisition. Now suppose the number of groups m is fixed but groups are not distributed evenly. We have the following proposition:

Proposition 3.2. *Social skill investment increases when the distribution of population size among different groups becomes more even.*

Proof. Since the total population size is N, we must have relatively larger groups k such that $n_k > n^*$ (Otherwise the overall population size is smaller than N).

Again starting from $x_i = 1, \forall i$ as the Nash Equilibrium. The utility of social interaction for individual j in group k is:

$$U_{jk} = \begin{cases} f(n_k + (N - n_k)) - c, & \text{if } x_j = 1 \\ f(n_k), & \text{if } x_j = 0 \end{cases}$$

Individual j in this group will deviate if:

$$f(N) - c < f(n_k) \Rightarrow n_k > n^*$$

$$(3.9)$$

As there always exists $n_k > n^*$ when group sizes are unevenly distributed, the largest group k will deviate and choose $x_k = 0$. For the remaining groups, suppose group l is the second largest group. Given the largest group deviates from the equilibrium $x_i = 1, \forall i$, the same logic shows that for group l to deviate as well, we must have:

$$f(N - n_k) - c < f(n_l)$$

$$(3.10)$$

Since $f' > 0$, we find that the motivation for deviating increases with group size. In particular, when group sizes are more unevenly distributed, more groups will have large size so that they will deviate from the equilibrium where everyone chooses to invest in social skills.

$\square$

Social interaction, skill acquisition and ethnic diversity

We prove from the above propositions that skill investment is higher when the number of groups is larger or the group size is more evenly distributed. And how does these relate to ethnic diversity?

We therefore decompose ethnic diversity index into the number of groups and the distribution of group share as follows. Suppose each group has a share s_i over the whole population.

$$ELF = 1 - \sum_{i=1}^{n} s_i^2 = 1 - \sum_{i=1}^{n}((s_i - \frac{1}{n}) + \frac{1}{n})^2$$

It is obvious to get the following decomposition:

$$ELF = 1 - \frac{1}{n} - \sum_{i=1}^{n}(s_i - \frac{1}{n})^2 = 1 - \frac{1}{n} - n \cdot var(s_i) \qquad (3.11)$$

Thus ethnic diversity increases with the number of different ethnic groups and decreases with the variance of group share (which increases if the distribution of group size is more uneven). Based on proposition 3.1 and proposition 3.2, we have the following proposition 3.3:

Proposition 3.3. *Social skill investment increases with ethnic diversity, either due to the increase in the number of different groups or the more even distribution of group size.*

Following proposition 3.2, we also have proposition 3.4:

Proposition 3.4. *Ethnic groups with relatively smaller group size are not affected by the ethnic diversity because they always participate in inter-group social interaction and invest in skills regardless of ethnic diversity.*

This is because when the initial condition is $x_i = 1, \forall i$, the small group will not deviate as long as their group size is below a certain level (regardless of the strategies of the large group). In other words, they always choose to remain the initial condition regardless of the decision of larger groups. Therefore the small group will in general have more inter-group social interaction and social skill investment than the large group but their social interaction is not affected by ethnic diversity of the district. The intuition is that as the small groups get relatively less utility from intra-group communication, they rely more on inter-group connection and therefore

are less sensitive to the incentive to deviate caused by changes in the level of ethnic diversity.

One thing to notice is that in our data "large", "medium" and "small" groups are defined by the group size in the national population while in the model "small" and "large" groups are defined at district level. However, definitions at these two levels are compatible in our data. A detailed investigation of the population share in each district in both 1996 and 2001 shows that in general groups with large population size at the national level are also the dominant group in ethnically homogeneous districts, while groups with small population share at the national level also makes up a very small part of the population in those districts. In diverse places the population size of these groups becomes more balanced.

Social skills and labour market outcomes

The social skills acquired through inter-group interactions in a diverse place might potentially improve workers' employment opportunities in several ways.

Less search cost in job hunting. Social skill lowers the cost of searching for potential jobs, therefore increasing labour supply. More social skills help individuals build closer and stronger intra-group contacts. For example, people with higher social skills are better at making use of networks and other methods in gaining job information or asking for referrals. Current literature shows that social network is an important factor in providing more job opportunities for low-educated labours both in South Africa (Magruder, 2010) and in other developing countries (Munshi, 2003).

Increased productivity of certain skills. Recent literature which incorporates different tasks in the production function (Acemoglu and Autor, 2011) and highlights the importance of social skills (Deming, 2017). Under the framework that low

and high-skill workers have their own comparative advantages in dealing with different tasks and the range of tasks performed by low-skill workers is determined by where their comparative advantages are, Deming (2017) explains that social skill increases the productivity of certain tasks by allowing workers with comparative advantages to trade their tasks, which leads to more efficient production. In our story, acquiring additional social skills may also potentially increase the productivity of certain tasks and increase the employment chances for low-skilled workers by allowing them to perform a wider range of tasks.

Overcoming skill deficit. A simple explanation on why social skill stimulates employment is that it works as a substitute for other skills required by employers. In particular, low-educated workers may lack skills necessary for certain occupations, which prevents them from getting the position. For example, if the candidate for the position of a salesman lacks necessary skills of communication, proficiency of additional language may compensate for this communication skills and guarantees him for the position. As the substitutability between social skill and skills acquired through formal education helps more people qualified for the positions they apply for in a more diverse place, the employment rate will increase accordingly. Skill acquisition from inter-group interaction here functions in a way similar to what is emphasised in related literature that community-based network can work as a substitute for endowments by helping individuals from disadvantaged families get out of low-skill occupational traps (Munshi, 2011).

In conclusion, social skill improves employment either by reducing search cost in the job hunting, increasing productivity of certain skills or substituting for some skills necessary for certain occupations. One implication of the above three channels is that ethnic diversity may be more influential to low-educated individuals, as they can be more restricted by the high search cost, low productivity and lack of certain basic skills.

Ethnic diversity, social skill acquisition and employment: empirical evidence

In this section we provide some evidence to show that social skill acquisition increases with ethnic diversity. There is no straightforward information in census data on social skills. The closest one we can approach is the information on second language at home, including whether or not one speaks a second language and which language they speak. A black person is considered to have some proficiency in a second language if he speaks either one of the nine ethnic languages or a common language (English or Afrikaans). Language is often considered as a cognitive skill which can be learnt from school. In this setting, however, controlling for educational background and investigating into the heterogeneity in the acquisition of language skills among sub-groups, we hope the proficiency of the second language can capture some information on the skills one acquires from inter-group interactions.

More importantly, whether one speaks a second language (and which language he speaks) reflects more of his investment in social skills than the inheritance of language skills from his parents. This results from a series of laws and regulations during the Apartheid regime. Firstly, inter-racial marriage was prohibited during Apartheid starting from 1949 when the Prohibition of Mixed Marriage Act came into effect. The act was repealed in 1985 by the Immorality and Prohibition of Mixed Marriages Amendment Act. In 1996 and 2001 census, parents and spouse of the working-age black people of our interest either lived through Apartheid when marriage between black and white (or black and coloured) was abandoned, or they got married before the independence of South Africa from the British colonisation when there was already informal racial segregation. Thus it is not very likely that the proficiency of English or Afrikaans among the current generation was purely obtained from their parents in the inter-racial marriage. Even among the black population, inter-ethnic marriage is also rare. As is discussed at the beginning

of the paper, inter-ethnic relationship was deteriorated during Apartheid so that marrying someone from another ethnic group is not a common case. Appendix table A6 shows that in 1996 census, the contemporary inter-ethnic marriage rate is less than 4%. This phenomenon is even more rare in the parental generation as their inter-ethnic marriage rate is only 1%. Although the sample is selected as only spouse and parents cohabiting with the household head are included in the census, this statistics can still reflect the low inter-ethnic marriage rate.

Furthermore, whether one speaks a second language is not very likely to capture the language proficiency of individuals before they decided to move out of the homelands. As is discussed in the institutional setting, there were almost no indigenous black people in the "white" areas in South Africa and the contemporary population in these districts are mainly the decedents of the migrants from different homelands before the arrival of white colonisers. Therefore it is unlikely that those ancestors learnt English or Afrikaans before migration. For contemporary migration, both 1996 and 2001 census data show that more than 50% of people never move from the time of birth up till the time of the census survey. Even among recent migrants, intra-district migration is much larger than inter-district migration.

To prove the channel in our theoretical model, we first show that ethnic diversity improves social skill acquisition (i.e. measured by second language proficiency) and then we demonstrate that higher social skill is correlated with higher employment rate conditional on ethnic diversity. As the information on second language proficiency is only reported in 1996 census data, we only show the results in 1996 census in this section.

Appendix table A6 also reveals that the proportion of people who speak a second language is not too small. Among the whole black population, around 22.5% speaks a second language, 8.7% (13.8%) of which speaks a common language (ethnic language). In the regression analysis we focus on the common language (English or

Afrikaans) instead of ethnic language as the former one is more related to labour market performance in wage-employment and less likely to reflect family inheritance as the ban on inter-racial marriage was more strict than inter-ethnic marriage during Apartheid[37].

We introduce a dummy variable on whether one can speak English or Afrikaans as a second language and regress it on ethnic diversity in 1996, conditional on the same set of control variables in the main analysis. Simple OLS regressions may suffer from the same problem as is discussed before. For example, there are two potential types of selection of migrants related to their language proficiency. Firstly, migrants with higher ability are able to move out of the homelands and these people might have already mastered a second language prior to migration. Secondly, migrants with better language proficiency choose to move to a more diverse area where there are more job opportunities. If the first type of selection is the case, people with higher ability than their counterparts in the original homelands can potentially move to both ethnically homogenous and diverse places. Thus we should not see any correlation between ethnic diversity and proficiency of second language if language skills are purely captured by the selection of migrants at the time of moving out of homelands. The second selection of migration comes from the fact that migrants with higher ability (including language efficiency) move to more diverse places as migrants who cannot speak a common language may find it difficult to communicate with people outside their ethnic groups. To deal with this selection, we use the same instrumental variable approach as is implemented in the main analysis (using predicted value of ethnic diversity in 1996 as an instrument for real ethnic diversity).

Table 12 shows both OLS and IV regression results about how ethnic diversity affects individuals' second language proficiency. Panel A, B and C investigates the results for the whole black population, the heterogeneity of the effects of ethnic

---

[37]But in regressions the proficiency of both common language and ethnic language can respond to ethnic diversity.

diversity by educational levels and group size. The coefficients in Panel A in both OLS and IV regressions are significantly positive, indicating that ethnic diversity increases the probability of learning a second language (English or Afrikaans). We break down educational levels into detailed categories to further capture the heterogeneous effects of ethnic diversity on skill acquisition: primary school (up to 7 years of schooling), junior school (up to 9 years of schooling), senior high school (up to 12 years of schooling) and college and above. The significant effect of ethnic diversity only exists among the group with the lowest level of education. In panel C, a comparison between groups with large, medium and small population size indicates that ethnic diversity has a strong and positive effect on language skills only among the ethnic groups with relatively large population size, which is consistent with proposition 3.4 in the model. In addition, the instrumental variable remains strong in both whole-sample and sub-sample regressions.

We then look at whether acquisition of social skills improves labour market outcomes by regressing employment probabilities on the proficiency of a second language (English or Afrikaans) conditional on ethnic diversity, as is presented in Table 13. The dependent variable in panel A is a dummy on whether one is employed or not (including unemployed and inactive) while in panel B the dependent variable equals 1 if one is an employee and 0 if one is unemployed or inactive. The independent variable in all these OLS regressions is a dummy on whether one can speak English or Afrikaans as a second language. Again we look at the whole sample, sub-samples across different levels of education and the difference among groups with large, medium and small population size.

In all regressions learning a second common language is positively and significantly associated with higher employment rate (both overall employment rate and wage-employment rate). Columns 2 - 5 in Table 13 show that the returns to a second language (English and Afrikaans) decrease with educational levels, indicating that

low-educated workers benefit more from additional language skills.

We focus on the evidence on investment in social skills rather than the overall level of social interaction because our theoretical model highlights that investment in social skills is higher in a more ethnically diverse place while allows for the possibility that there is no difference in the overall level of social interaction between ethnically diverse and homogeneous places. These two phenomena are compatible because inter-ethnic communication, which requires more skills than intra-ethnic connection, is only part of the overall social interaction.

A further empirical evidence on the mechanism of the model comes from the decomposition of ethnic diversity index into the number of ethnic groups and the variation of population share among these groups. According to proposition 3.1 and proposition 3.2 in the theoretical model, employment rate should increase with the number of ethnic groups in a district and decrease if the distribution of group size becomes more uneven.

Based on Equation 3.11, we decompose the Herfindahl Index (1-*ELF*) into two terms: reverse of the number of ethnic groups and the number of ethnic groups times the variance of each group's population share. According to the theoretical model, we should observe both of the two terms to be negatively correlated with employment. In other words, when we regress unemployment rate on these two terms, the coefficients for both should be positive, which is the case in Table 14. In both 196 and 2001, we report the association between these two decomposed terms on unemployment rate at individual regressions both for the whole sample and for the sub-sample (larger vs. smaller groups; highly- vs. low-educated individuals). Also consistent with proposition 3.4, these two factors only affect ethnic groups with relatively larger size. Columns 5 and 6 show evidence that the link between number of ethnic groups, variance in each group's population share and employment rate is stronger for low-educated black people.

Summary of the theoretical model and mechanism

In summary, diversity along ethnic lines could provide individuals with cognitive and non-cognitive social skills, which improves their employability. That is to say, even if ethnic diversity does not necessarily increase the amount of overall social interactions within a district, it may still motivate people in more diverse areas to learn and practise more skills such as a common/official language. This is because communication with individuals from different ethnic groups requires more efforts and skills than intra-ethnic interaction. The acquisition of this extra skill, which is helpful in reducing coordination costs or increasing productivity of certain skills, could increase individuals' chances of finding a job.

In our model, without imposing any intrinsic difference in taste, skills or attitudes between different ethnic groups, the tradeoff between the cost of and benefit from developing social skills leads to the conclusion that inter-ethnic social interaction and investment in social skills are the mostly likely to occur in a place with large number of different groups and a place where the distribution of group size is relatively even, both of which imply high ethnic diversity. It is because starting from an initial condition where everyone invests in social skills, less people deviate from this investment decision in the equilibrium state in a more ethnically diverse place. This effect occurs mainly among the ethnic groups with relatively larger group size. In the labour market, the acquisition of these extra social skills is helpful in lowering the barrier to formal jobs by reducing coordination and search costs, by increasing productivity of certain skills or by substituting for some necessary skills which are otherwise not available especially to low-educated people.

3.6.2 Ruling out some alternative explanations

Ethnic diversity might positively affect the labour market outcomes of the blacks through several channels. Here we rule out some alternative explanations through which ethnic diversity improves labour market outcomes based on our data and narratives.

Labour supply: skill complementarity. There might be some skill complementarities among different ethnic groups, as each may have their own comparative advantages in skills. For example, South Sotho are believed to have special skills as shaft-sinkers on the mines (Guy and Thabane, 1988). Therefore, diversity generates creativity and innovative environment by combining people with different skills. In this case, we can also expect diversity to affect differently individuals with different level of education. A priori, we would expect to find a stronger effect for the higher educated whose activities would benefit more from knowledge-sharing and problem solving. However, when we run OLS and IV regressions for people with high and low levels of education, we find in the above empirical part that the relationship between ethnic diversity as measured by the fractionalization index and employment is positive for the low-educated and not that obvious for the high-educated, pointing at the substitutability rather than the complementarity between education and ethnic diversity.

Furthermore, if ethnic diversity generates skill complementarity, it might also give birth to new occupations as new skills can be learnt from other ethnic groups and this creates opportunities for occupations which rely on otherwise infeasible tasks. Therefore, if ethnic diversity stimulates new ideas and skills, we may observe a larger range of occupations in a more diverse place. We regress the range of occupations in each district [38] on ethnic diversity. We take the results from 1996 census as an

---

[38]We measure the range of tasks by counting the total number of different occupations observed in each district. Occupations are counted in 3-digit code level.

example in Appendix Table A7. We do not find any positive relationship between diversity and potential new occupations in either OLS or IV regressions .

Labour supply: social grant. Social grant, such as Old Age pension, potentially dis-incentivise labour force participation in South Africa (Banerjee et al., 2008). At the same time there is a possibility that a more ethnically homogenous place is associated with higher level of public goods provision, which might include social grants. In particular, governments in a more ethnically homogeneous place might be willing to offer more social grants due to the nepotism towards the dominant group in that place or less coordination cost among ethnic groups. If the receipt of social grants dis-incentivise working-age people to enter labour force, this could also explain the association between higher ethnic diversity and higher employment rate. However, this is not the case in our setting for two reasons. Firstly, provision of social grants is mainly designed at the national level, which does not vary across magisterial districts. Secondly, we include province fixed effects to account for potential discrepancy of social grants at province level.

Labour demand: discrimination. Discrimination in the labour market is a potential reason why homogeneous places discourage employment, as employers deliberately prevent the minority groups from gaining job opportunities and therefore the demand for minority labours is declined (Goldberg, 1982). It has been proved that the disutility from discrimination against minority groups in the production network harms the productivity of co-workers (Hjort, 2014; Borjas and Bronars, 1989).

A more diverse place can reduce the discrimination against minority groups by encouraging higher level of tolerance and openness. As the chance of interacting and communicating with other ethnic groups increases in a more ethnically diverse place, discrimination in the labour market becomes less of an issue, either because employers have access to more information about the productivity and behaviours

of ethnic minorities, or because they are more open to people from different backgrounds.

If this story is the case, we would expect that ethnic groups with smaller size benefit more from increased ethnic diversity than those with relatively larger size, which contradicts our empirical evidence.

Labour demand: diversity of taste. Another potential driving force of labour demand might be the diversity of taste. As people from different ethnic groups have diversified tastes for consumption goods, the variety of consumption increases when a place becomes more ethnically diversified. This induces the diversity of production as well, resulting in higher variety of labour inputs in the production process. When different labour inputs are complementary in the production function, this love for variety of labours increases the total demand for labour, therefore improving workers' chance in the labour market. However, if this is the case, we should see the positive effect of ethnic diversity among both large and small ethnic groups, which also contradicts the empirical findings. There is also related literature about how greater diversification of sectoral demands reduces unemployment (Neumann and Topel, 1991). However this works under the condition that workers are mobile enough, which is not likely to be a prevalent case in South Africa where many black people locate far away from economic centres and the transportation cost is very high to them.

3.7   Conclusion and Discussion

This paper provides empirical support for the positive role played by within-black ethnic diversity and blacks' labour market outcomes in post-Apartheid South Africa based on an instrumental variable approach. We also propose a plausible theoretical model to explain how the need for inter-ethnic social interaction stimulates invest-

ment in social skills in more diverse places, making black workers better equipped for the labour market.

The finding reveals that ethnic identity, together with inter-ethnic relationship, is still a distinctive feature shaping people's social life and labour market in modern South African society. The distinction between ethnic groups does not fade away after years of integration, which might result from the Apartheid regime which reinforced ethnic identity. In addition, although the climate of hatred and mistrust generated by the Apartheid system had substantial repercussions on the social fabric, inter-ethnic connections still occur within the black population.

Our result is different from, yet can be reconciled with the association between ethnic diversity and inter-ethnic cleavages or the erosion of social cohesion. Firstly, most of those literature highlights the under provision of public goods and social capital in ethnically fragmented communities in developing countries (Alesina et al., 2016), or the conflict between different ethnic groups (Amodio and Chiovelli, 2017). Our story takes a different angel by focusing on skill investment motivated by social interaction. This can just be another side of inter-personal relations which can co-exist with conflicts or coordination problems. Secondly, we have shown in our model that the initial condition in skill investment is important in shaping the ultimate equilibrium. If the society starts from the situation where no one actively participates in inter-ethnic communication, benefits from inter-ethnic connection will stay at the low level forever. Therefore, societies where ethnic diversity is negatively associated with socio-economic indicators might have worse initial conditions in inter-ethnic interaction.

We also find the heterogeneous effects of ethnic diversity on labour market outcomes for different sub-groups. In particular, labour market outcomes of the ethnic groups with larger size and low-educated people are more responsive to ethnic diversity. The former indicates that our story is not likely to be the case where the minority

assimilates to the majority by integrating into their culture and language, nor is it the story that diversity alleviates discrimination against minority groups (in both cases only the small group will respond to diversity level). Rather, in our story groups with both large and small sizes participate in social interaction and invest in social skills in response to ethnic diversity.

The finding that low-educated people benefit more from ethnic diversity is different from several papers highlighting skill complementarity and knowledge spillovers (mostly in developed countries where diversity affects high-skilled labours more), but are analogue to findings in related topics where social interaction acts as a substitute for family background or formal education. For example, Munshi (2011) shows that community-based network can compensates for people's disadvantaged family background by bootstrapping these people out of low-skill occupational traps. This substitution mechanism might not always be efficient, though. As in Munshi and Rosenzweig (2006), community-based network continues to channel Indian boys into local language schools and consequentially traditional occupations which have lower returns than emerging white-collar occupations, suggesting a dynamic inefficiency.

Could any interventions be designed to increase employment opportunities for the black South African? As is presented in the theoretical framework, a successful intervention must encourage more inter-ethnic connection which can motivate people to invest in more social skills. It can be an efficient policy as we show that the initial investment in social skills is important to the ultimate equilibrium. Therefore, an attempt at fostering inter-ethnic communication in a more diverse society will have long-lasting effects on overall skill investments. Policies which directly improve black people's social skills may also be effective in preparing them for better employment opportunities.

These interventions to improve people's labour market performance have far-reaching

implications not only in different aspects of South African society but also in dealing with ethnic issues all over the world. On the one hand, reducing unemployment can have other important consequences on South African society. For example, it has been estimated that in contemporary South Africa a 10 percentage point reduction in unemployment lowers the Gini coefficient by 3 percent (Anand et al., 2016).

On the other hand, this paper can also shed light on dealing with inter-ethnic relations in other African countries or even developed countries. In recent decades, Western societies have also become considerably more ethnically diverse due to the net immigration flows and the growing presence of ethnic communities (Putnam, 2007), which gives rise to more social problems. For example, there is some negative evidence of ethnic diversity on the support for redistribution which in particular harms low-income earners (Dahlberg et al., 2012). Furthermore, current immigration policies in the US and the European refugee crisis also require urgent modification in policy interventions to improve inter-ethnic relationships and explore the positive impact of ethnic diversity on economic outcomes, to which our mechanism about inter-ethnic interactions can be generalised.

From 700s A.D.	1652	Before mid-1800s	1867	1910

Bantu migration from central and eastern Africa	First permanent European settlement	Emigration from homelands to "white areas"	First discovery of mines	Establishing the Union of SA
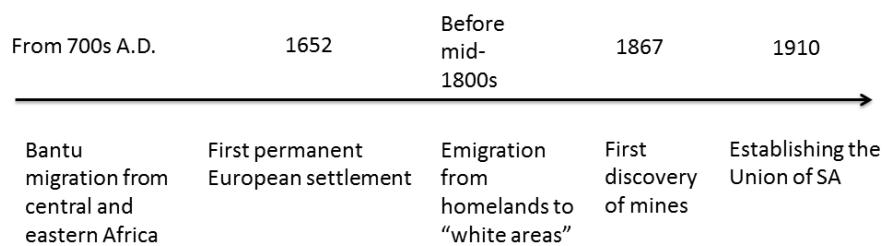
(a) Timeline of Bantu migration and early development in South Africa

1948	1951	1953	1976	1986	1994

Independence	Bantu Authorities Act	Bantu Self-Govern Act	Transkei Independence	End of Pass Law	End of Apartheid
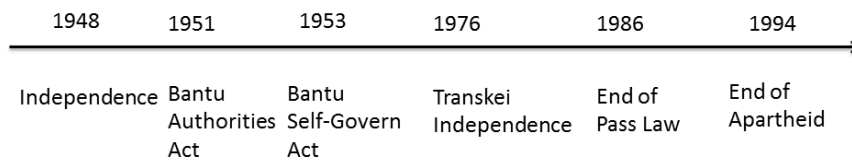
(b) Timeline of modern South Africa starting from Apartheid

Notes: The figures presents the timeline of important nodes in South African history: Bantu migration from central and eastern Africa, emigration of ethnic groups from original homelands, the White colonisation, the discovery of mines and Apartheid regime. Sources of narratives: Mwakikagile (2010) and Gradin (2014).

Figure 3.1. Timeline of Bantu migration, historical development and Apartheid regime in South Africa
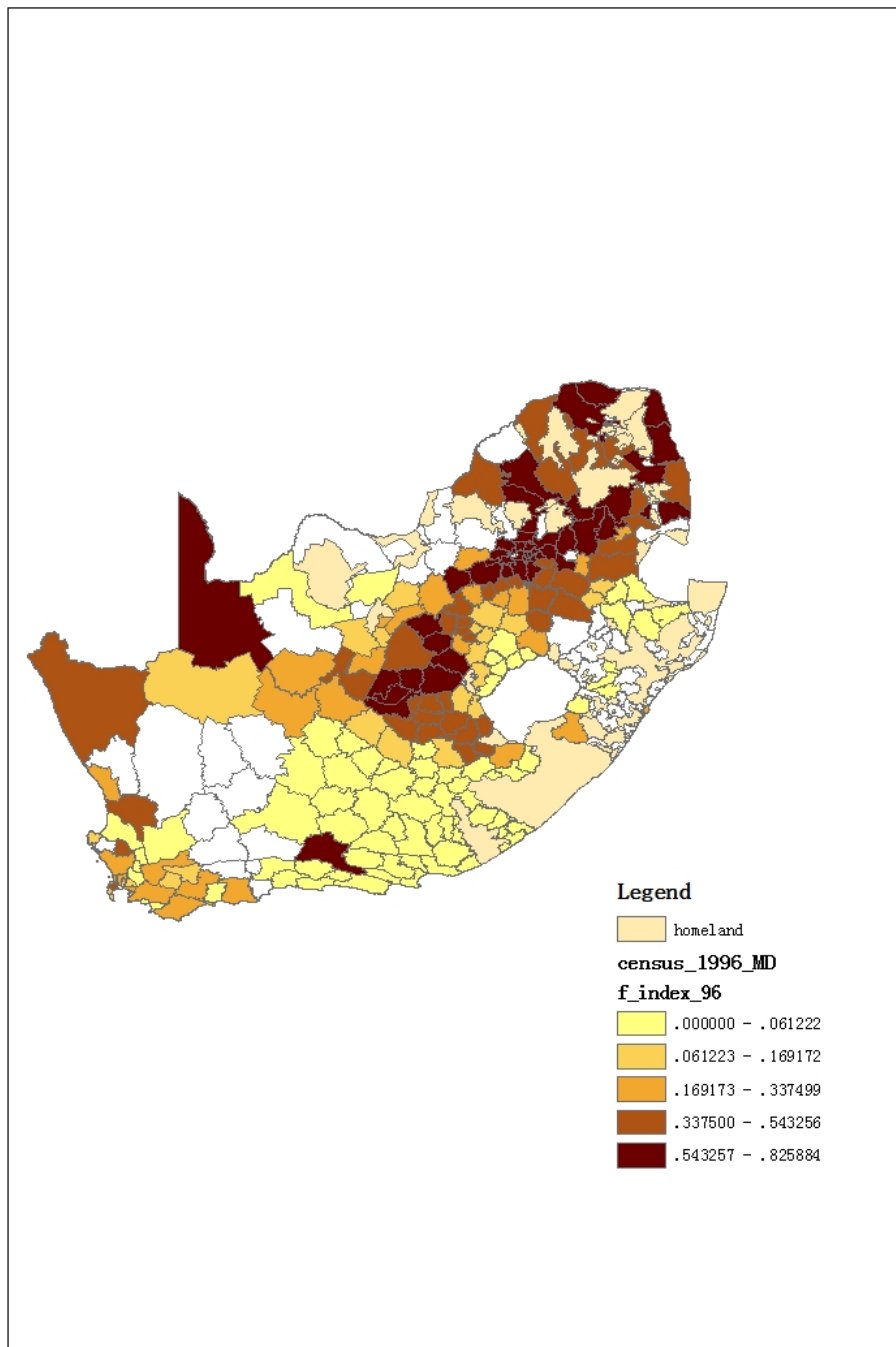
Notes: The figure presents the geographical pattern of ethnic diversity across South African districts in 1996. Within-black ethnic diversity is measured with Fractionalisation Index analogue to Herfindahl Index. The results are calculated by the authors based on 1996 data. The pattern of ethnic diversity in 2001 is very similar.

Figure 3.2. Distribution of ethnic fractionalization index in 1996
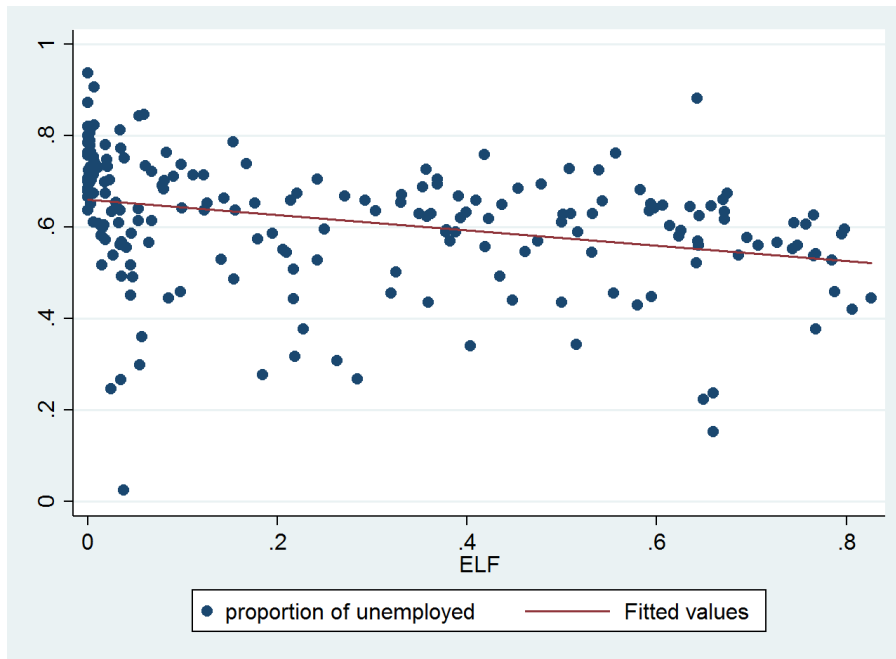
(a) Unemployment and ELF 1996
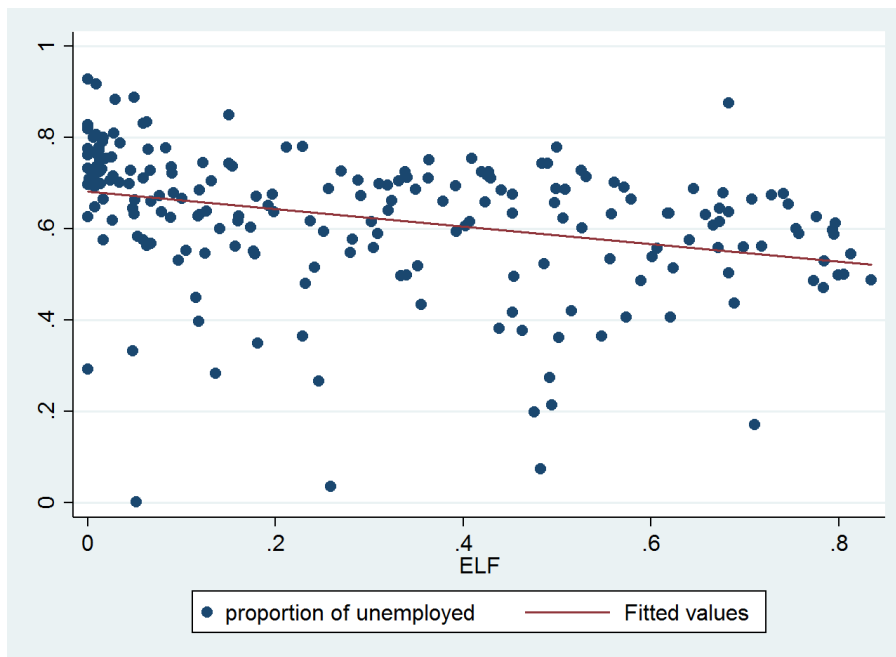
(b) Unemployment and ELF 2001

Notes: The figures present the results on the correlation between ethnic diversity and unemployment rate. Both are measured at the magisterial district level (therefore unemployment rate is calculated as the proportion of employed people over the whole working-age black population in a district). The results are calculated by the authors based on 1996 and 2001 census data.

Figure 3.3. The relationship between ethnic diversity and unemployment in South Africa in 1996 and 2001
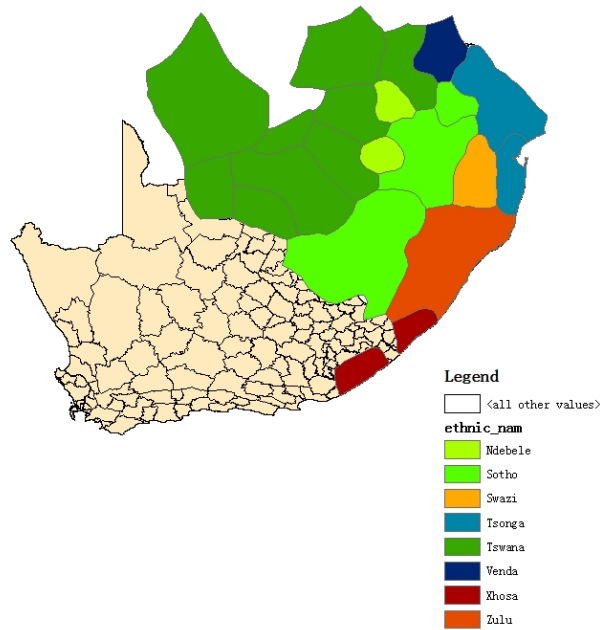
(a) Murdock's map
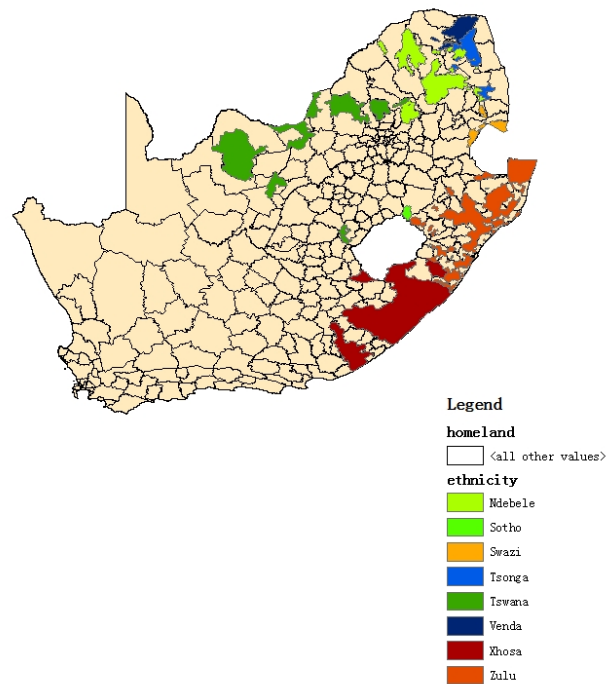
(b) Bantustan

Notes: The figures compares the distribution of ethnic groups in South Africa in Murdock map and the location of Bantustans as proxies for ethnic homelands. Murdock map comes from George Murdock's 1959 work which illustrates the dominant ethnic group in each geographical unit, which is highly consistent with the Bantustans for these ethnic groups assigned by the Apartheid government. This confirms that the location of these Bantustans can well reflect the spatial distribution of original homelands for those ethnic groups.

Figure 3.4. Comparison between the historical settlements od the black ethnic groups and Bantustans

Legend

homeland	census_1996_MD
▨ <all other values>	**f_index_pre**
NAME	☐ .480542 - .558344
🟩 Bophuthatswana	☐ .568333 - .622272
🟥 Ciskei	☐ .630734 - .679333
🟦 Gazankulu	☐ .681848 - .725466
🟧 Kangwane	☐ .727936 - .763769
🟧 KwaZulu	☐ .773882 - .801715
🟩 Kwandebele	☐ .802482 - .826406
🟩 Lebowa	☐ .829075 - .846912
🟩 Qwaqwa	☐ .849911 - .861714
🟥 Transkei	☐ .862397 - .875331
🟦 Venda	

Notes: The figures shows the spatial distribution of our instrumental variable for ethnic diversity - the predicted ethnic factionalisation index. Following the idea that district more (less) equidistant to multiple homelands are more (less) diverse, we first calculate the stock of each ethnic group in each district based on the distance in between with a gravity model. The instrumental variable is a fractionalisation index calculated based on the predicted stock of migrants.

Figure 3.5. Distribution of predicted ethnic fractionalization index

Table 1.1. Summary statistics of demographics and employment among black ethnic groups in 1996

	Population size	Share of the black population	Self employed	Wage employee	Unemployed	Inactive	Unemployed+inactive
Xhosa	2229452	0.252	0.027 [0.162]	0.312 [0.463]	0.273 [0.445]	0.388 [0.487]	0.661 [0.474]
Zulu	2073036	0.234	0.037 [0.189]	0.349 [0.477]	0.264 [0.441]	0.350 [0.477]	0.614 [0.487]
South Sotho	2009582	0.227	0.026 [0.159]	0.349 [0.477]	0.242 [0.428]	0.383 [0.486]	0.625 [0.484]
Tswana	1039138	0.117	0.026 [0.158]	0.384 [0.486]	0.224 [0.417]	0.367 [0.482]	0.590 [0.492]
North Sotho	770110.7	0.087	0.038 [0.192]	0.412 [0.492]	0.236 [0.424]	0.314 [0.464]	0.549 [0.498]
Tsonga	295688.6	0.033	0.075 [0.263]	0.434 [0.496]	0.247 [0.431]	0.244 [0.430]	0.491 [0.500]
Ndebele	185065.8	0.021	0.040 [0.195]	0.358 [0.480]	0.227 [0.419]	0.375 [0.484]	0.602 [0.490]
Swazi	181467.1	0.020	0.041 [0.198]	0.405 [0.491]	0.224 [0.417]	0.330 [0.470]	0.554 [0.497]
Venda	80189.34	0.009	0.048 [0.214]	0.497 [0.500]	0.202 [0.402]	0.252 [0.434]	0.455 [0.498]
Overall	8863729.5	1.000	0.032 [0.177]	0.355 [0.479]	0.251 [0.434]	0.361 [0.48]	0.613 [0.487]

Note: The number and proportion of each ethnic group in the whole black population are calculated in the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. Employment outcomes are calculated from individual-level 1996 census data. "Self-employed" refers to the proportion of self-employed people in each ethnic group over the whole working-age population of the corresponding ethnic group. Other labour market outcomes are calculated in similar ways.

Table 1.2. Summary statistics of demographics and employment among black ethnic groups in 2001

	Population size	Share of the black population	Self employed	Wage employee	Unemployed+inactive
Xhosa	3105625	0.249	0.017	0.299	0.684
			[0.130]	[0.458]	[0.465]
Zulu	2798132	0.224	0.025	0.331	0.643
			[0.156]	[0.471]	[0.479]
South Sotho	2531013	0.203	0.020	0.324	0.657
			[0.139]	[0.468]	[0.475]
Tswana	1373413	0.110	0.018	0.373	0.610
			[0.132]	[0.484]	[0.488]
North Sotho	1341608	0.107	0.027	0.396	0.577
			[0.163]	[0.490]	[0.494]
Tsonga	552403.3	0.044	0.048	0.421	0.531
			[0.214]	[0.494]	[0.50]
Ndebele	292188.3	0.023	0.029	0.370	0.601
			[0.168]	[0.483]	[0.490]
Swazi	324071.7	0.026	0.028	0.376	0.597
			[0.164]	[0.484]	[0.491]
Venda	172927.4	0.014	0.034	0.457	0.509
			[0.183]	[0.498]	[0.500]
Overall	12491382	1.000	0.023	0.341	0.636
			[0.149]	[0.474]	[0.481]

Note: The number and proportion of each ethnic group in the whole black population are calculated in the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. Employment outcomes are calculated from individual-level 1996 census data. "Self-employed" refers to the proportion of self-employed people in each ethnic group over the whole working-age population of the corresponding ethnic group. Other labour market outcomes are calculated in similar ways. The 2001 census data does not distinguish unemployed and economically inactive people.

Table 2.1. Summary statistics of ethnic fragmentation and labour market outcomes in 1996

	High ELF			Low ELF			
	Mean	S.d	Obs	Mean	S.d.	Obs	ttest
ELF	0.507	0.018	103	0.044	0.005	102	***
self employment	0.028	0.0019	103	0.021	0.00195	102	***
wage employee	0.4	0.012	103	0.32	0.0144	102	***
unemployed	0.57	0.012	103	0.658	0.014	102	***
agriculture	0.466	0.019	103	0.454	0.018	102	
manufacture	0.115	0.011	103	0.09	0.011	102	*
service	0.419	0.013	103	0.455	0.017	102	**
manager	0.0136	0.001	103	0.0117	0.002	102	
profession	0.07	0.004	103	0.082	0.0066	102	*
clerk	0.0318	0.0031	103	0.0198	0.0024	102	***
serve	0.0728	0.0035	103	0.063	0.0048	102	*
craft	0.107	0.0084	103	0.125	0.0116	102	
skilled agriculture	0.121	0.0062	103	0.107	0.0062	102	*
operator	0.088	0.005	103	0.062	0.004	102	***
unskill	0.495	0.0139	103	0.529	0.0124	102	**

Note: This table compares labour market outcomes in districts with relatively high (i.e. above the median value) and low levels of ethnic diversity. The sample is only for the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. Employment outcomes are calculated from individual-level 1996 census data. "Self-employed" refers to the proportion of self-employed people in each ethnic group over the whole working-age population of the corresponding ethnic group."Wage employee" and "unemployed" are calculated in similar ways. We only focus on people who are employed when comparing the allocation of workers across industrial sectors and occupations.

Table 2.2. Summary statistics of ethnic fragmentation and labour market outcomes in 2001

	High ELF			Low ELF			
	Mean	S.d	Obs	Mean	S.d.	Obs	ttest
ELF	0.527	0.016	105	0.077	0.007	105	***
self employment	0.0218	0.003	105	0.0185	0.002	105	
wage employee	0.396	0.013	105	0.315	0.014	105	***
unemployed	0.582	0.014	105	0.667	0.014	105	***
agriculture	0.338	0.024	105	0.376	0.023	105	
manufacture	0.183	0.017	105	0.096	0.008	105	***
service	0.478	0.019	105	0.527	0.023	105	*
manager	0.017	0.0026	105	0.0167	0.0046	105	
profession	0.082	0.0057	105	0.08	0.0058	105	
clerk	0.056	0.003	105	0.054	0.007	105	
serve	0.081	0.005	105	0.076	0.0047	105	
craft	0.059	0.005	105	0.084	0.0087	105	***
skilled agriculture	0.117	0.007	105	0.074	0.005	105	***
operator	0.108	0.0057	105	0.088	0.0051	105	***
unskill	0.48	0.0144	105	0.527	0.0147	105	**

Note: This table compares labour market outcomes in districts with relatively high (i.e. above the median value) and low levels of ethnic diversity. The sample is only for the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. Employment outcomes are calculated from individual-level 1996 census data. "Self-employed" refers to the proportion of self-employed people in each ethnic group over the whole working-age population of the corresponding ethnic group."Wage employee" and "unemployed" are calculated in similar ways. We only focus on people who are employed when comparing the allocation of workers across industrial sectors and occupations.

Table 3. Validity of instrumental variables

Dependent variable	[1] 1996	[2] 2001
Panel A: Job opportunities		
Distance to the closest economic centre	-274,037	-244,879
	(301,492)	(258,991)
Panel B: Economic activities of the white		
Share of white who are self employed contemporarily	0.231*	0.0309
	(0.138)	(0.136)
Share of white who are wage employed contemporarily	0.0966	0.187
	(0.169)	(0.159)
Proportion of white	0.337	0.149
	(0.220)	(0.141)
Panel C: Path dependence		
Share of white who are wage employed in 1980	-0.227	-0.244
	(0.216)	(0.220)
Proportion of white in 1980	-0.126	-0.716***
	(0.261)	(0.232)
Panel D: Migration		
Number of migrants	0.433	10,612
	(0.313)	(25,718)
District controls	YES	YES
Individual controls (district average)	YES	YES
Province fixed effect	YES	YES
Obs	205	210

Note: This table conducts validity test of the instrumental variable based on 1996 and 2001 census data. The sample is only for the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. All regressions are at the district level. We control for district-level variables especially geographical features, individual-level controls aggregated at district average and province fixed effects. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 4. First-stage regression results: individual level regressions

Dependent var.	[1] ELF 1996	[2] ELF 1996	[3] ELF 2001	[4] ELF 2001
Predicted ELF	2.012***	1.518***	2.022***	1.668***
	(0.105)	(0.304)	(0.110)	(0.280)
Edu	0.000720*	0.000110	0.000875***	0.000291
	(0.000404)	(0.000305)	(0.000321)	(0.000223)
Male	0.00283***	0.000412	0.00286**	0.00127**
	(0.00104)	(0.000624)	(0.00121)	(0.000640)
Age	5.27e-07	-5.56e-05	7.35e-05	-2.72e-05
	(8.56e-05)	(5.51e-05)	(7.63e-05)	(4.11e-05)
Married	0.00884***	0.00318**	0.00928***	0.00442***
	(0.00230)	(0.00159)	(0.00208)	(0.00141)
Father alive	0.00196*	0.000665	0.00380***	0.00159**
	(0.00111)	(0.000928)	(0.00127)	(0.000714)
Pop density	4.41e-05*	4.06e-05***	3.15e-05	3.12e-05**
	(2.65e-05)	(1.52e-05)	(2.25e-05)	(1.27e-05)
Urban	0.0384***	0.00880	0.0301**	-1.31e-05
	(0.0141)	(0.0113)	(0.0133)	(0.0106)
River	0.0720***	0.0832***	0.0466*	0.0622**
	(0.0273)	(0.0293)	(0.0254)	(0.0281)
Density mine	3.071**	0.371	3.025**	0.543
	(1.473)	(0.829)	(1.381)	(0.752)
Prop black	-0.141	-0.282***	-0.218	-0.434***
	(0.119)	(0.0591)	(0.178)	(0.0823)
Distance closest	-0.000269	-0.000485***	-0.000243	-0.000520***
	(0.000166)	(0.000171)	(0.000180)	(0.000152)
Ruggedness	0.00131	0.00661	-0.00699	-0.00331
	(0.00832)	(0.00814)	(0.00812)	(0.00659)
Per capita light	0.520**	0.353	0.871**	0.538
	(0.213)	(0.235)	(0.340)	(0.366)
Road	0.0613**	0.00810	0.0608*	0.0101
	(0.0287)	(0.0299)	(0.0325)	(0.0296)
Conflict	0.0127	0.0209**	-0.00173	-0.00496***
	(0.0176)	(0.00908)	(0.00338)	(0.00191)
Proportion manu	0.283**	0.288**	0.308***	0.252***
	(0.126)	(0.114)	(0.0856)	(0.0790)
Proportion service	0.596***	0.382***	0.349***	0.180**
	(0.119)	(0.131)	(0.0892)	(0.0864)
Ethnicity fixed effect	YES	YES	YES	YES
Province fixed effect	NO	YES	NO	YES
F-statistics of the instrument	367.1	24.93	336.8	35.53
R-squared	0.805	0.872	0.817	0.883
Observations	464,130	464,130	697,369	697,369

Note: This table reports the first-stage results of the instrumental variable based on 1996 and 2001 census data and report the F-statistics of the instrumental variable. The sample is only for the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. All regressions are at the individual level. We report all the control variables, both district-level variables especially geographical features and individual-level controls for socio-economic status. We control for ethnicity fixed effects and compare the results with and without province fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table 5. Ethnic diversity, unemployment and labour force participation: individual level regressions

Dependent variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
	unemployed		inactive		unemployed + inactive		unemployed + inactive	
	1996	1996	1996	1996	1996	1996	2001	2001
Panel A: OLS estimates								
Ethnic fractionalisation ELF	-0.0158	-0.0199	-0.151***	-0.0610**	-0.135***	-0.0810**	-0.152***	-0.150***
	(0.0115)	(0.0177)	(0.0208)	(0.0246)	(0.0200)	(0.0322)	(0.0250)	(0.0373)
Individual controls	YES	YES	YES	YES	YES	YES	YES	YES
District controls	YES	YES	YES	YES	YES	YES	YES	YES
Province FE	NO	YES	NO	YES	NO	YES	NO	YES
R-squared	0.033	0.033	0.194	0.195	0.152	0.153	0.174	0.175
Observations	464,130	464,130	464,130	464,130	464,130	464,130	697,368	697,369
Panel B: IV estimates								
Ethnic fractionalisation ELF	-0.0229	-0.141***	-0.167***	0.0427	-0.144***	-0.0984	-0.153***	-0.171**
	(0.0142)	(0.0471)	(0.0223)	(0.0602)	(0.0221)	(0.0756)	(0.0248)	(0.0863)
Individual controls	YES	YES	YES	YES	YES	YES	YES	YES
District controls	YES	YES	YES	YES	YES	YES	YES	YES
Province FE	NO	YES	NO	YES	NO	YES	NO	YES
F statistics of the instrument	367.1	24.93	367.1	24.93	367.1	24.93	336.8	35.53
R-squared	0.033	0.032	0.194	0.153	0.152	0.195	0.174	0.175
Observations	464,130	464,130	464,130	464,130	464,130	464,130	697,369	697,369

Note: This table reports results about the effect of ethnic diversity on unemployment rate at individual-level regressions based on 1996 and 2001 census data. The sample is only for the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. We control for district-level variables especially geographical features, individual-level controls aggregated at district average and ethnicity fixed effects. We compare the results with and without province fixed effects. Ethnic diversity is measured with fractionalisation index. We separate unemployed and economically inactive groups only for 1996 results as these two categories are combined in 2001 census. "Unemployed + inactive" is a dummy variable which equals 1 if one is unemployed or inactive and 0 if one is employed. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 6. Ethnic diversity and employment status: individual level regresions

Dependent variable	[1] wage employment 1996	[2] self/wage 1996	[3] wage employment 2001	[4] self/wage 2001
Panel A: OLS estimates				
Ethnic fractionalisation ELF	0.0850**	-0.0197	0.147***	0.013
	(0.0329)	(0.017)	(0.0374)	(0.0137)
Individual controls	YES	YES	YES	YES
District controls	YES	YES	YES	YES
Province FE	YES	YES	YES	YES
R-squared	0.1938	0.0109	0.1728	0.0081
Observations	449,200	180,535	681,529	253,809
Panel B: IV estimates				
Ethnic fractionalisation ELF	0.112	-0.052	0.176***	-0.041
	(0.0768)	(0.0398)	(0.0854)	(0.0354)
Individual controls	YES	YES	YES	YES
District controls	YES	YES	YES	YES
Province FE	YES	YES	YES	YES
F statistics of the instrument	25.01	23.55	35.82	32.14
R-squared	0.1938	0.0108	0.1728	0.0075
Observations	449,200	180,535	681,529	253,809

Note: This table reports results about the effect of ethnic diversity on employment and the allocation between self- and wage-employment at individual-level regressions based on 1996 and 2001 census data. The sample is only for the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. We control for district-level variables especially geographical features, individual-level controls aggregated at district average and ethnicity fixed effects. We also control for province fixed effects. Ethnic diversity is measured with fractionalisation index. In column 1 and 3 we drop self-employed people as they are a very small proportion of the whole working-age population. Column 2 and 4 are based only on the employed black people. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 7. Ethnic diversity, intensive margin and wage: individual level regressions

Dependent variable	[1] OLS log monthly income	[2] OLS log hourly wage	[3] OLS hour	[4] IV log monthly income	[5] IV log hourly wage	[6] IV hour
Panel A: Individual level, census data						
Ethnic fractionalisation ELF	0.366***	0.411***	-1.425	0.527***	0.455*	2.261
	(0.0722)	(0.0935)	(1.263)	(0.195)	(0.274)	(3.731)
Individual controls	YES	YES	YES	YES	YES	YES
District controls	YES	YES	YES	YES	YES	YES
Province FE	YES	YES	YES	YES	YES	YES
F statistics of the instrument				32.35	32.35	32.19
R-squared	0.345	0.314	0.053	0.344	0.313	0.052
Observations	228,256	228,256	232,533	228,256	228,256	232,533
Panel B: Individual level, LFS data						
Ethnic fractionalisation ELF	-0.0676	-0.0207	0.118	0.498	0.102	23.17
	(0.254)	(0.247)	(2.978)	(1.033)	(0.910)	(16.41)
Individual controls	YES	YES	YES	YES	YES	YES
District controls	YES	YES	YES	YES	YES	YES
Province FE	YES	YES	YES	YES	YES	YES
F statistics of the instrument				5.514	5.315	5.595
R-squared	0.481	0.473	0.054	0.479	0.473	0.018
Observations	3,615	3,478	3,660	3,615	3,478	3,660

Note: This table reports results about the effect of ethnic diversity on other labour market outcomes at individual-level regressions, including working hour, hourly wage and monthly earnings. We only report the result in 2001 as there is no information on hours of working in 1996 census. The sample is only for the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. We control for district-level variables especially geographical features, individual-level controls aggregated at district average and ethnicity fixed effects. We control for province fixed effects. Ethnic diversity is measured with fractionalisation index. All the columns only focus on employees. *** p<0.01, ** p<0.05, * p<0.1.

Table 8. Ethnic diversity and employment: district fixed effects models

Dependent variable	[1] unemploy + inactive	[2] wage employ	[3] self/wage	[4] log monthly income
Ethnic fractionalisation ELF	-0.291***	0.340***	-0.133*	-0.382
	(0.0709)	(0.072)	(0.0696)	(0.365)
Individual controls (district average)	YES	YES	YES	YES
District controls	YES	YES	YES	YES
R-squared	0.493	0.488	0.2436	0.730
Observations	410	410	410	410

Note: This table reports results about the effect of ethnic diversity on employment based on the district-level balanced panel. The sample is only for the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. We control for district-level variables which vary over time and individual-level controls aggregated at district level and province fixed effects. Ethnic diversity is measured with fractionalisation index. The dependent variable in olumns 1 is the proportion of unemployed over the whole working-age black population. Column 2 is defined in a similar way but we exclude those who are self-employed. Column 3 has the dependent variable which is the ratio of the number of self-employed to that of employees at district level. Column 4 only focuses on back people who are employed. *** p<0.01, ** p<0.05, * p<0.1.

Table 9. Heterogeneous effects of ethnic diversity on wage-employment: individual level regressions

	[1]	[2]	[3]	[4]
	OLS	IV	OLS	IV
Dependent variable	1996	1996	2001	2001
Panel A: By education				
High edu	0.0403	0.05	0.163***	0.12
	(0.0305)	(0.0889)	(0.031)	(0.0924)
Obs	151,944	151,994	291,307	291,307
Low edu	0.097***	0.141*	0.1365***	0.19**
	(0.035)	(0.078)	(0.042)	(0.09)
Obs	297,206	297,206	390,222	390,222
Panel B: By ethnicity				
Large	0.107***	0.124	0.152***	0.1827**
	(0.034)	(0.0816)	(0.031)	(0.0842)
Obs	320,901	320,901	459,108	459,108
Medium	0.023	-0.170	0.204***	-0.0374
	0.0737	(0.2234)	(0.06)	(0.213)
Obs	91,373	91,373	149,632	149,632
Small	-0.0215	-0.574*	0.038	0.217
	(0.0975)	(0.295)	(0.081)	(0.4079)
Obs	36,926	36,926	72,789	72,789
Panel C: By industrial sector				
Agriculture	0.0610*	0.0302	0.000105	0.0514
	(0.0334)	(0.0700)	(0.0326)	(0.0705)

Table 9 – continued from previous page

Dependent variable	[1] OLS 1996	[2] IV 1996	[3] OLS 2001	[4] IV 2001
Obs	165,605	165,605	180,227	180,227
Manufacturing	-0.0219**	-0.00688	-0.00801	-0.0118
	(0.00899)	(0.0190)	(0.00980)	(0.0196)
Obs	165,605	165,605	180,227	180,227
Service	-0.0390	-0.0233	0.00791	-0.0396
	(0.0276)	(0.0603)	(0.0299)	(0.0695)
Obs	165,605	165,605	180,227	180,227

Panel D: By occupation

	[1] OLS 1996	[2] IV 1996	[3] OLS 2001	[4] IV 2001
Manager	0.00380	0.0159*	0.00744**	0.0236***
	(0.00400)	(0.00965)	(0.00332)	(0.00764)
Obs	153,294	153,294	224,942	224,942
Profession	-0.0204	0.101**	-0.0170	0.0844*
	(0.0176)	(0.0425)	(0.0133)	(0.0489)
Obs	153,294	153,294	224,942	224,942
Clerk	0.0171***	-0.00244	0.0226**	0.0331
	(0.00578)	(0.0149)	(0.0102)	(0.0243)
Obs	153,294	153,294	224,942	224,942
Serve	-0.0249**	0.0331	0.0173	-0.0411
	(0.0118)	(0.0309)	(0.0156)	(0.0360)
Obs	153,294	153,294	224,942	224,942
Craft	-0.0203	-0.0360	-0.0401	0.0247
	(0.0272)	(0.0669)	(0.0261)	(0.0465)

Table 9 – continued from previous page

Dependent variable	[1] OLS 1996	[2] IV 1996	[3] OLS 2001	[4] IV 2001
Obs	153,294	153,294	224,942	224,942
Skilled agriculture	0.00620	-0.0677	0.0196	-0.0524*
	(0.0218)	(0.0427)	(0.0168)	(0.0308)
Obs	153,294	153,294	224,942	224,942
Operator	-0.0342**	-0.140***	-0.0391**	-0.0584
	(0.0156)	(0.0457)	(0.0172)	(0.0367)
Obs	153,294	153,294	224,942	224,942
Unskilled	0.0727**	0.0962	0.0292	-0.0139
	(0.0343)	(0.0714)	(0.0358)	(0.0637)
Obs	153,294	153,294	224,942	224,942
Individual controls	YES	YES	YES	YES
District controls	YES	YES	YES	YES
Province FE	YES	YES	YES	YES

Note: This table reports the main results about the heterogeneous effects of ethnic diversity on the probability of being an employee at individual-level regressions by subgroups in both 1996 and 2001 census. The sample is only for the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. We control for district-level variables especially geographical features, individual-level controls aggregated at district average and ethnicity fixed effects. We also control for province fixed effects. Ethnic diversity is measured with fractionalisation index. "High" ("Low") education is defined as years of schooling above (below) 9. All the columns in Panel C and Panel D only focus on employees to illustrate the allocation of employed workers across different industrial sectors and occupations. *** p<0.01, ** p<0.05, * p<0.1.

Table 10. Ethnic diversity and employment: separating native, migrants and immigrants

Dependent variable	[1]	[2]	[3]	[4]	[5]	[6]
	unemployed + inactive			wage employment		
	native	migrants	immigrants	native	migrants	immigrants
Panel A: IV estimates, 1996 census						
Ethnic fractionalisation ELF	-0.1615**	0.1216	-0.4639	0.1717**	-0.1108	0.5057*
	(0.0798)	(0.1284)	0.288	(0.0817)	(0.1300)	(0.292)
Individual controls	YES	YES	YES	YES	YES	YES
District controls	YES	YES	YES	YES	YES	YES
Province FE	YES	YES	YES	YES	YES	YES
F statistics of the instrument	26.20	17.81	10.56	26.23	17.83	10.86
R-squared	0.1907	0.1929	0.2998	0.1876	0.1962	0.307
Observations	305,458	128,215	4,657	296,864	122,956	4,283
Panel B: IV estimates, 2001 census						
Ethnic fractionalisation ELF	-0.153*	-0.248	-0.877*	0.159**	0.25	0.936*
	(0.0786)	(0.205)	(0.478)	(0.079)	(0.017)	(0.51)
Individual controls	YES	YES	YES	YES	YES	YES
District controls	YES	YES	YES	YES	YES	YES
Province FE	YES	YES	YES	YES	YES	YES
F statistics of the instrument	37.03	19.58	9.15	37.17	20.19	9.16
R-squared	0.1713	0.1960	0.2916	0.1682	0.1978	0.3147
Observations	568,260	119,696	20,390	556,296	116,089	19,250

Note: This table reports results about the effect of ethnic diversity on employment separately for native, migrants and immigrants at individual-level regressions based on 1996 and 2001 census data. "Native" is defined as people who were born in the district and never move out or within-district migrants. "Migrants" are cross-district migrants while "immigrants" are those who migrated from another country. The sample is only for the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. We control for district-level variables especially geographical features, individual-level controls aggregated at district average and ethnicity fixed effects. We also control for province fixed effects. Ethnic diversity is measured with fractionalisation index. In columns 4-6 we drop self-employed people as they are a very small proportion of the whole working-age population. *** p<0.01, ** p<0.05, * p<0.1.

Table 11. Ethnic diversity and the emigration of the white

Dependent variable	[1] num of white in 1996	[2] num of white in 1985	[3] diff: 96 - 85	[4] num of white in 2001	[5] num of white in 1985	[6] diff: 01 - 85
	1996	1996	1996	2001	2001	2001
Panel A: OLS estimates						
Ethnic fractionalisation ELF	20,689	2,156	18,533	-1,301	7,043	-8,344
	(23,884)	(31,934)	(17,342)	(27,706)	(24,844)	(14,749)
R-squared	0.762	0.781	0.432	0.752	0.894	0.748
Observations	205	205	205	210	210	210
Panel B: IV estimates						
Ethnic fractionalisation ELF	146,930	225,351	-78,421	36,125	15,864	20,261
	(185,524)	(202,881)	(62,681)	(110,761)	(85,383)	(36,230)
F statistics of instruments	9.959	9.959	9.959	30.86	30.86	30.86
R-squared	0.734	0.720	0.316	0.750	0.893	0.743
Observations	205	205	205	210	210	210
Individual controls	YES	YES	YES	YES	YES	YES
District controls	YES	YES	YES	YES	YES	YES
Province FE	YES	YES	YES	YES	YES	YES

Note: This table looks at whether ethnic diversity is correlated with the number of white population in 1996 and 2001 and the emigration of the white from the district after the end of Apartheid at district-level regressions. The sample is only for the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. We control for district-level variables especially geographical features and individual-level controls aggregated at district average. We also control for province fixed effects. Ethnic diversity is measured with fractionalisation index. *** p<0.01, ** p<0.05, * p<0.1.

Table 12. Ethnic diversity and skill acquisition: second language

Dependent variable	[1] OLS	[2] IV
Panel A: Whole sample		
Overall	0.034*	0.109*
	(0.0184)	(0.0659)
F statistics of the instrument		25.67
Obs	654,116	654,116
Panel B: By education		
Edu <=7	0.0345**	0.076*
	(0.01496)	(0.0458)
F statistics of the instrument		27.12
Obs	379,257	379,257
7 < Edu <=9	0.034	0.1478
	(0.0254)	(0.092)
F statistics of the instrument		22.69
Obs	110,508	110,508
9 < Edu <=12	0.041	0.247
	(0.033)	(0.1575)
F statistics of the instrument		
Obs	151,343	151,343
Edu >12	0.1027	0.412
	(0.0658)	(0.2985)
F statistics of the instrument		11.86
Obs	13,008	13,008
Panel C: By ethnicity		
Large	0.027	0.160**
	(0.0205)	(0.077)
F statistics of the instrument		19.84
Obs	469,737	469,737
Medium	0.046	-0.068
	(0.037)	(0.2)
F statistics of the instrument		7.26
Obs	131,601	131,601
Small	-0.0175	0.011
	(0.03)	(0.088)
F statistics of the instrument		13.01
Obs	52,778	52,778
Individual controls	YES	YES
District controls	YES	YES
Province FE	YES	YES

Note: This table reports results about the effect of ethnic diversity on the acquisition of social skills (proficiency of second language as a proxy) at individual-level regressions based on 1996 census data (as there is no information on the second language in 2001 census). The sample is only for the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. We control for district-level variables especially geographical features, individual-level controls aggregated at district average and ethnicity fixed effects. We also control for province fixed effects. Ethnic diversity is measured with fractionalisation index. We look at both the whole sample and sub-samples split by educational levels and group size. *** p<0.01, ** p<0.05, * p<0.1.

Table A0. Gravity model predicting the stock of black population in white districts: PPML estimator

	Coef.	Std. Err.	t-stat
	Dependent variable: ethnic population N_{jk}		
Distance Dis_{jk}	-.0039	(.0007)	-5.17
Ethnic group fixed effects:			
Group 1	.9750	(.2139)	4.56
Group 2	.6133	(.1769)	3.47
Group 3	.1778	(.2248)	0.79
Group 4	-.4604	(.2311)	-1.99
Group 5	.2220	(.2259)	0.98
Group 6	.8940	(.1803)	4.96
Group 8	.0469	(.1833)	0.26
Group 9	-.8184	(.2776)	-2.95
Constant	9.157	(.2176)	42.08
R-squared	.092		
Observations	1989		

Note: This table reports results about the gravity model which helps estimate the stock of each ethnic group in each "white" district based on 1985 census data. The sample is for all the "white" magisterial districts which can be matched to 1996 and 2001 census. We control for homeland fixed effects and run a regression of the stock of ethnic groups on the distance between their corresponding homelands and each district using PPML models. *** p<0.01, ** p<0.05, * p<0.1.

Appendix - History and Econometrics

A1 Bantu migration and the formation of ethnic diversity from historical narratives

Below we provide a summary of the history of the Bantu migration from central and eastern Africa and the settlement of these groups in South Africa for each ethnic groups in details. Narrative evidence is summarised from Mwakikagile (2010) and Gradin (2014).

Ethnicity	Time of migration into SA	Homelands	Time of moving into white areas	Bantustan
Xhosa	Before 1400s	Today's Eastern Cape	After conflicts with the native Khoisan	Ciskei and Transkei
Zulu	16th century	Eastern part, today's Kwazulu-Natal	Early 1800s	KwaZulu
Swazi	15th and 16th centuries	Southern part of Tongaland in what is now Mozambique	17th and 18th centuries into the Pongola River	KaNgwane
Ndebele	Before 1835	Today's Northern Province, Mpumalanga and Gauteng	By 1835 towards Swaziland and Northern Transvaal	KwaNdebele, Lebowa
North Sotho	1500s	Today's Limpopo and Northwest	After the war with Boers and Ndebele	Qwawa
South Sotho	1500s	Today's Limpopo and Northwest	After the war with Boers and Ndebele	Qwawa
Tswana	1500s	Today's Limpopo and Northwest	After the war with Boers and Ndebele	Bophuthatswana
Tsonga	Before the early 1500s	Close to today's Mozambique	After conflicts with Zulu	Gazankulu
Venda	Before 800s A.D.	A mountainous area in the northern part close to Limpopo River	800s A.D. to Matopo Hills	Venda

A2 Data Source and construction of district-level variables

In this section we present data sources and the construction of our district-level control variables in detail. Emphasis has been given on those geographical measures.

Variable	Data source	Construction of variable
Panel A: from census		
Area of the district	Census 1996 and 2001 district-level shape file	Calculated from the shape file directly in ArcGIS.
Population density	Census 1996 and 2001	Calcualte the total number of black in each district in census data and divide it by area.
Proportion of the black	Census 1996 and 2001	Calculate the number of black over the whole population.
Proportion of manufacturing	Census 1996 and 2001	Calculate the number of people working in manufacturing sector over the whole employed black people.
Proportion of service	Census 1996 and 2001	Calculate the number of people working in service sector over the whole employed black people.
Urban/rural	Census 1996 and 2001	Information on whether one lives in an urban or rural settlement is explicitly in census data.

Variable	Data source	Construction of variable
Panel B: sources on geography		
Overlap: district and homeland	A map (shape file) of homeland provided by Tim Brophy and Adrian Frith.	Intersect the boundary of districts with that of homelands and see the overlap in ArcGIS.
River	Census 2001 river shape file	Overlapping shape file of districts and river and directly calculate in ArcGIS.
Road	Census 2001 major road shape file	Overlapping shape file of districts and road and directly calculate in ArcGIS.
Ruggedness	From Nunn and Puga (2012). We also tried the measure of slope from the same data source with similar results.	Same as Nunn and Puga (2012).
Density of mine	Mineral Resources Data System (MRDS) https://mrdata.usgs.gov/mrds/	Overlapping shape file of districts and mines.Calculating numbers of mines in each district and divide it by area.
Nighlight per capita	The National Oceanic and Atmospheric Administration night-time light satellite images. www.noaa.gov/stories/data-earth-at-night.	Calculating nightlight measures in a district (summation of the index over grids).Dividing it by the whole population in the district obtained from census data.
Distance: districts and homeland	A map (shape file) of homeland provided by Tim Brophy and Adrian Frith.	Calculate Euclidean between centroid of districts and the border of homelands.
Distance to closest homeland	A map (shape file) of homeland provided by Tim Brophy and Adrian Frith.	Choose the minimum value of the distance to all homelands.
Conflict	The Geo-referenced Event Dataset of the Uppsala Conflict Data Program (UCDP-GED v1.5) for 1996. The Armed Conflict Location and Event Data Project (ACLED) database for 2001.	Same as Amodio and Chiovelli (2017).

Table 13. Skill acquisition and labour market outcomes: second language

Dependent variable	[1] Overall	[2] Edu <=7	[3] 7 < Edu <=9	[4] 9 < Edu <=12	[5] Edu >12	[6] Large	[7] Medium	[8] Small
Panel A: unemployed as dependent variable, conditional on diversity								
Second official	-0.133***	-0.171***	-0.12***	-0.08***	-0.048***	-0.123***	-0.144***	-0.176***
	(0.0128)	(0.1917)	(0.0175)	(0.0084)	(0.0074)	(0.01)	(0.022)	(0.036)
Individual controls	YES	YES	YES	YES	YES	YES	YES	YES
District controls	YES	YES	YES	YES	YES	YES	YES	YES
Province FE	YES	YES	YES	YES	YES	YES	YES	YES
R-squared	0.2	0.1766	0.23	0.2592	0.1632	0.192	0.212	0.229
Obs	461,942	206,934	98.949	143,928	12,131	329,416	93,673	38,853
Panel B: wage employ as dependent variable, conditional on diversity								
Second official	0.132***	0.1689***	0.1187***	0.078***	0.049***	0.121***	0.1429***	0.1796***
	(0.013)	(0.02)	(0.0179)	(0.0086)	(0.0076)	(0.011)	(0.023)	(0.0356)
Individual controls	YES	YES	YES	YES	YES	YES	YES	YES
District controls	YES	YES	YES	YES	YES	YES	YES	YES
Province FE	YES	YES	YES	YES	YES	YES	YES	YES
R-squared	0.1994	0.1793	0.2244	0.2533	0.1650	0.1914	0.2088	0.2357
Obs	447,103	200,034	95,859	139,537	11,673	319,580	90,817	36,706

Note: This table reports results the relationship between social skill acquisition (proficiency of second language as a proxy) and employment at individual-level regressions based on 1996 census data. We control for ethnic diversity and investigate whether this language skill is positively correlated with employment chances. The sample is only for the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. We control for district-level variables especially geographical features, individual-level controls aggregated at district average and ethnicity fixed effects. We also control for province fixed effects. Ethnic diversity is measured with fractionalisation index. In Panel A we keep the whole working-age black sample while in Panel B we drop self-employed people as they are a very small proportion of the whole working-age population. *** p<0.01, ** p<0.05, * p<0.1.

Table 14. Decomposing ethnic diversity into number of groups and variance in population share

Dependent variable	[1] Whole sample	[2] Large	[3] Medium	[4] Small	[5] High edu	[6] Low edu
Panel A: 1996 census						
1/No. of groups	0.168***	0.189***	0.135	-0.476	0.099**	0.185***
	(0.0425)	(0.043)	(0.2798)	(0.532)	(0.0457)	(0.043)
N*Var(share)	0.0825**	0.101***	0.025	-0.026	0.039	0.095***
	(0.032)	(0.0326)	(0.07)	(0.092)	(0.03)	(0.034)
R-squared	0.1954	0.1884	0.2037	0.2238	0.2784	0.1837
Observations	464,130	330,792	94,256	39,082	156,877	307,253
Panel B: 2001 census						
1/No. of groups	0.29***	0.27***	-0.371	0.6636	0.313***	0.262***
	(0.066)	(0.056)	(0.754)	(1.0915)	(0.054)	(0.0724)
N*Var(share)	0.1535***	0.1555***	0.2145***	0.0273	0.173***	0.1406***
	(0.037)	(0.0313)	(0.0567)	(0.0765)	(0.031)	(0.0413)
Individual controls	YES	YES	YES	YES	YES	YES
District controls	YES	YES	YES	YES	YES	YES
Province FE	YES	YES	YES	YES	YES	YES
R-squared	0.1754	0.1672	0.1796	0.1961	0.2214	0.1727
Observations	697,369	468,704	153,041	75,624	298,235	399,134

Note: This table reports results based on the decomposition of ethnic diversity index into items relating to number of ethnic groups and group share, and how these two items are associated with unemployment rate at individual-level OLS regressions with 1996 and 2001 census data. The sample is only for the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. We control for district-level variables especially geographical features, individual-level controls aggregated at district average and ethnicity fixed effects. We also control for province fixed effects. Ethnic diversity is measured with fractionalisation index. We look at both the whole sample and sub-samples split by educational levels and sample size. *** p<0.01, ** p<0.05, * p<0.1.

Table A1. First-stage regression results: district level regressions

Dependent var.	[1] ELF 1996	[2] ELF 1996	[3] ELF 2001	[4] ELF 2001
Predicted ELF	0.0455***	1.128***	1.820***	1.520***
	(0.0168)	(0.358)	(0.104)	(0.274)
District controls	YES	YES	YES	YES
Individual controls (district average)	YES	YES	YES	YES
Province fixed effect	NO	YES	NO	YES
F-statistics of the instrument	291.5	9.959	304.4	30.86
R-squared	0.815	0.870	0.853	0.897
Observations	205	205	210	210

Note: This table reports first-stage regression results for our instrumental variable at the district-level regressions based on 1996 and 2001 census data. The sample is only for the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. We control for district-level variables especially geographical features and individual-level controls aggregated at district average. We also compare results with and without controls for province fixed effects. Ethnic diversity is measured with fractionalisation index. *** p<0.01, ** p<0.05, * p<0.1.

Table A2. Ethnic diversity and different employment status: individual level regressions

Dependent variable	[1] self employment 1996	[2] wage employment 1996	[3] self employment 2001	[4] wage employment 2001
Panel A: OLS estimates				
Ethnic fractionalisation ELF	-0.00399	0.0850**	0.00972**	0.140***
	(0.00573)	(0.0329)	(0.00464)	(0.0366)
Individual controls	YES	YES	YES	YES
District controls	YES	YES	YES	YES
Province FE	YES	YES	YES	YES
R-squared	0.014	0.173	0.010	0.159
Observations	464,130	464,130	697,369	697,369
Panel B: IV estimates				
Ethnic fractionalisation ELF	-0.00416	0.103	-0.00740	0.178**
	(0.0131)	(0.0750)	(0.0120)	(0.0825)
Individual controls	YES	YES	YES	YES
District controls	YES	YES	YES	YES
Province FE	YES	YES	YES	YES
F statistics of the instrument	24.93	24.93	35.53	35.53
R-squared	0.014	0.173	0.010	0.159
Observations	464,130	464,130	697,369	697,369

Note: This table reports results about the effect of ethnic diversity on self- and wage-employment rate at individual-level regressions based on 1996 and 2001 census data. The sample is only for the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. We control for district-level variables especially geographical features, individual-level controls aggregated at district average and ethnicity fixed effects. We also control for province fixed effects. Ethnic diversity is measured with fractionalisation index. "Self-employment" is a dummy variable which equals 1 if one is self-employed and 0 for all other working-age black population. *** p<0.01, ** p<0.05, * p<0.1.

Table A3. Ethnic diversity and employment: district level regressions

Dependent variable	[1] unemployed + inactive	[2] wage employee	[3] self employment	[4] self/wage	[5] unemployed + inactive	[6] wage employee	[7] self employment	[8] self/wage
	1996	1996	1996	1996	2001	2001	2001	2001
Panel A: OLS estimates								
Ethnic fractionalisation ELF	-0.0786***	0.0744***	0.018	0.0042	-0.118***	0.105***	0.012*	0.0276
	(0.0265)	(0.0273)	(0.0195)	(0.0074)	-0.0354	(0.0356)	(0.007)	(0.0204)
Individual controls	YES	YES	YES	YES	YES	YES	YES	YES
District controls	YES	YES	YES	YES	YES	YES	YES	YES
Province FE	YES	YES	YES	YES	YES	YES	YES	YES
R-squared	0.885	0.865	0.4567	0.5573	0.874	0.859	0.3595	0.2719
Observations	205	205	205	205	210	210	210	210
Panel B: IV estimates								
Ethnic fractionalisation ELF	-0.202**	0.189**	-0.037	0.0132	-0.146	0.196**	-0.0497*	-0.091
	(0.0906)	(0.0962)	(0.066)	(0.023)	(0.0909)	(0.0818)	(0.0299)	(0.0498)
Individual controls	YES	YES	YES	YES	YES	YES	YES	YES
District controls	YES	YES	YES	YES	YES	YES	YES	YES
Province FE	YES	YES	YES	YES	YES	YES	YES	YES
F statistics of the instrument	9.959	9.959	9.959	9.959	30.86	30.86	30.86	30.86
R-squared	0.871	0.852	0.4264	0.5533	0.874	0.852	0.1769	0.089
Observations	205	205	205	205	210	210	210	210

Note: This table reports results about the effect of ethnic diversity on employment and the allocation between self- and wage-employment at district-level regressions based on 1996 and 2001 census data. The sample is only for the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. We control for district-level variables especially geographical features and individual-level controls. We also control for province fixed effects. Ethnic diversity is measured with fractionalisation index. Dependent variables are the proportion of people in each employment status over the whole working-age black population. *** p<0.01, ** p<0.05, * p<0.1.

Table A4. Estimations based on non-linear econometric models

Dependent variable	[1] Unemployed + inactive	[2]	[3]	[4] Wage employment	[5]	[6]
	Logit	Probit	IV Probit	Logit	Probit	IV Probit
Panel A: 1996 census						
Ethnic fractionalisation ELF	-0.08**	-0.078**	-0.0782	0.084**	0.0822**	0.08
	(0.0337)	(0.033)	(0.0798)	(0.0346)	(0.0347)	(0.082)
Observations	464,130	464,130	464,130	449,200	449,200	449, 200
Panel B: 2001 census						
Ethnic fractionalisation ELF	-0.152***	-0.149***	-0.148*	0.1496***	0.1467***	0.1446*
	(0.0388)	(0.0387)	(0.088)	(0.039)	(0.039)	(0.0869)
Observations	697,369	697,369	697,369	681,529	681,529	681,529
Individual controls	YES	YES	YES	YES	YES	YES
District controls	YES	YES	YES	YES	YES	YES
Province FE	YES	YES	YES	YES	YES	YES

Note: This table reports results about the effect of ethnic diversity on employment based on non-linear econometric models in 1996 and 2001. The sample is only for the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. We control for district-level variables especially geographical features, individual-level controls aggregated at district average and ethnicity fixed effects. We also control for province fixed effects. Ethnic diversity is measured with fractionalisation index. In column 4, 5 and 6 we drop self-employed people as they are a very small proportion of the whole working-age population. *** p<0.01, ** p<0.05, * p<0.1.

Table A5.1. RQ index as a measure of ethnic diversity: first stage regressions

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
	Individual level		District level		Individual level		District level	
	1996	2001	1996	2001	1996	2001	1996	2001
Dependent var.	RQ	RQ	RQ	RQ	RQ	RQ	RQ	RQ
Predicted RQ	-3.116***	-3.024***	-3.440***	-3.164***	-1.778***	-1.624**	-1.688**	-1.556**
	(0.373)	(0.397)	(0.404)	(0.456)	(0.643)	(0.638)	(0.782)	(0.726)
District controls	YES	YES	YES	YES	YES	YES	YES	YES
Individual controls	YES	YES	YES	YES	YES	YES	YES	YES
Province fixed effect	NO	NO	NO	NO	YES	YES	YES	YES
F-statistics of the instrument	69.66	58.00	72.49	48.07	7.64	6.48	4.67	4.59
R-squared	0.681	0.688	0.673	0.682	0.7673	0.7795	0.7512	0.7761
Observations	464,130	697,369	205	210	464,130	697,369	205	210

Note: This table reports first-stage results about our instrumental variable for polarisation index based on 1996 and 2001 census data, at both district- and individual-level regressions. The sample is only for the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. We control for district-level variables especially geographical features and individual-level controls aggregated at district average. We also compare results with and without control for province fixed effects. Ethnic diversity is measured with polarisation index. *** p<0.01, ** p<0.05, * p<0.1.

Table A5.2. RQ index as a measure of ethnic diversity: individual level regressions

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
	1996	1996	2001	2001	1996	1996	2001	2001
Dependent variable	unemploy	wage employment	unemploy	wage employment	unemploy	wage employment	unemploy	wage employment
Panel A: OLS estimates								
RQ	-0.0841***	0.084***	-0.102***	0.101***	-0.008	0.011	-0.05*	0.05
	(0.0205)	(0.02)	(0.0204)	(0.02)	(0.026)	(0.026)	(0.03)	(0.03)
Individual controls	YES	YES	YES	YES	YES	YES	YES	YES
District controls	YES	YES	YES	YES	YES	YES	YES	YES
Province FE	NO	NO	NO	NO	YES	YES	YES	YES
R-squared	0.192	0.1909	0.173	0.1705	0.1949	0.1934	0.1744	0.1721
Observations	464,130	449,200	697,369	681,529	464,130	449,200	697,369	681,529
Panel B: IV estimates								
RQ	-0.125***	0.1224***	-0.104***	0.1047***	0.025	-0.128	0.148	-0.138
	(0.0375)	(0.037)	(0.0395)	(0.0396)	(0.091)	(0.093)	(0.1244)	(0.123)
Individual controls	YES	YES	YES	YES	YES	YES	YES	YES
District controls	YES	YES	YES	YES	YES	YES	YES	YES
Province FE	NO	NO	NO	NO	YES	YES	YES	YES
F statistics of the instrument	69.66	70.06	58	58.37	7.64	7.69	6.48	6.52
R-squared	0.192	0.1905	0.173	0.1705	0.1948	0.1934	0.1705	0.1685
Observations	464,130	449,200	697,369	681,529	464,130	449,200	697,369	681,529

Note: This table reports results about the effect of ethnic diversity on employment at individual-level regressions based on 1996 and 2001 census data. The sample is only for the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. We control for district-level variables especially geographical features, individual-level controls aggregated at district average and ethnicity fixed effects. We also compare results with and without control for province fixed effects. Ethnic diversity is measured with polarisation index. In column 2, 4, 6 and8 we drop self-employed people as they are a very small proportion of the whole working-age population. *** p<0.01, ** p<0.05, * p<0.1.

Table A6. Inter-ethnic marriage rate and ethnic diversity: 1996 census

	Mean	Std. Dev.	Obs
Inter-ethnic marriage			
Own generation	0.966	0.18	96,031
Parental generation	0.99	0.0966	532
Second language among married people			
Any second language	0.2356	0.424	95,580
Second English/Afrikaan	0.0888	0.284	95,580
Second ethnic language	0.147	0.354	95,580
Second language among whole sample			
Any second language	0.225	0.418	203,327
Second English/Afrikaan	0.087	0.283	203,327
Second ethnic language	0.138	0.345	203,327

Note: This table reports inter-ethnic marriage rate (i.e. marriage between different ethnic groups within the black population). Ethnicity is identified from the first language spoken by both household head and spouse for the current generation, and household head's parents for the parental generation. We also report the proportion of the black population who can speak a second language.

Table A7. Ethnic diversity and the range of occupations

Dependent	[1] Var manager	[2] Var profession	[3] Var clerk	[4] Var serve	[5] Var craft	[6] Var skill agri	[7] Var operator	[8] Var unskill
Panel A: OLS estimates								
Ethnic fragmentation index	0.932	1.406	-1.287	-1.021	-0.246	-0.729	-1.868	-3.493***
	(1.015)	(3.560)	(0.870)	(1.126)	(0.589)	(1.797)	(1.687)	(1.225)
R-squared	0.788	0.861	0.825	0.759	0.536	0.810	0.826	0.800
Obs	205	205	205	205	205	205	205	205
Panel B: IV estimates								
Ethnic fragmentation index	-4.433	-9.978	-7.992***	-10.17*	-4.572*	-22.31**	-23.09***	-9.843*
	(4.783)	(16.02)	(3.044)	(5.782)	(2.498)	(10.60)	(7.873)	(5.533)
F statistics of the instrument	9.959	9.959	9.959	9.959	9.959	9.959	9.959	9.959
R-squared	0.752	0.851	0.759	0.645	0.354	0.641	0.606	0.762
Obs	205	205	205	205	205	205	205	205
Individual controls	YES	YES	YES	YES	YES	YES	YES	YES
District controls	YES	YES	YES	YES	YES	YES	YES	YES
Province FE	YES	YES	YES	YES	YES	YES	YES	YES

Note: This table reports results about the effect of ethnic diversity on the variety of occupations among employees in 1996. The sample is only for the "white" magisterial districts which can be matched to 1985 census and whose black population accounts for more than 1% of the overall population. We control for district-level variables especially geographical features and individual-level controls aggregated at district average. We also control for province fixed effects. Ethnic diversity is measured with fractionalisation index. *** p<0.01, ** p<0.05, * p<0.1.

Chapter 4

The influence of family on children's trust formation

4.1   Introduction

The role of culture on economic choices and its effect on economic development is the subject of a lively debate in recent research. Among the cultural traits, trust towards others is one of the the most studied by social scientists (see Alesina and Giuliano, 2016 for a review).[1] Following the seminal contributions of Banfield (1958), Coleman (1988, 1990) and Putnam (1993, 2000), trust has been found to affect economic development (Knack and Keefer, 1997), innovation (Fukuyama, 1995), individual performance (Butler et al., 2016), financial development and trade (see Guiso et al., 2004, 2008c, 2009), and firm productivity (Bloom et al., 2012; La Porta et al., 1997). For a comprehensive review of the role of trust in economics,

---

[1]In this paper, we refer to "generalized" trust, which concerns our beliefs about the anonymous other. In other words, trust refers to the relations among individuals who are not bound by the kind of personal ties that bind members of the same family, or fellow workers (Algan and Cahuc, 2013).

see Algan and Cahuc (2013).[2]

Considering the important influence of trust on economic outcomes, the process of its formation is of key interest. At a broader level, theoretical work has highlighted the role of intergenerational transmission of values – such as trust – shedding light on the persistence of ethnic differences (Bisin and Verdier, 2001).[3] More recent studies have attempted to provide empirical content to the intergenerational transmission of values. In particular, using data from the German Socio-Economic Panel (SOEP), Dohmen et al. (2012) analyze the transmission of trust and risk attitudes from parents to children within a regression framework whereby children's attitudes are modelled as a function of those of parents. Their results suggest the presence of a positive intergenerational correlation.[4]

We build upon existing work and exploit longitudinal data to study how trust is transmitted from parents to their children. The cruciality of longitudinal data transpires in several steps of our analysis. First, we can disentangle the role of the direct transmission of trust from parents to children from that played by other factors of the environment shared by siblings. This distinction would not be possible with cross-sectional data.

---

[2]Effectively, Arrow (1972) states that "Virtually every commercial transaction has within itself an element of trust, certainly any transaction conducted over a period of time. It can be plausibly argued that much of the economic backwardness in the world can be explained by the lack of mutual confidence." Following his reasoning, the absence of markets or their malfunctioning, the misallocation of resources and, more generally, the differences in economic performance, could be ultimately attributed to the lack of trusting behavior.

[3]Related work has shown the importance of intergenerational transmission on fertility and work practises across cultures (Fernández and Fogli, 2009; Guiso et al., 2006; Fernández, 2008)

[4]An alternative strategy to identify the intergenerational transmission process is to focus on immigrants' attitudes. The central idea is to understand how immigrants' values – shaped by the diverse cultural and institutional background of their home countries – react and adapt to the environment in the common host country. For example, Ljunge (2014) estimates the intergenerational transmission of trust by analysing children of immigrants in 29 European countries with ancestry in 87 different nations. His approach entails performing regressions of the level of trust of second-generation immigrants on the average trust of the country of origin of parents. Remarkably, their results suggest that the transmission of trust is relevant only on the mother's side. Moschion and Tabasso (2014) use a similar approach with data on second generation immigrants in Australia and the United States to study how the mechanism of the transmission process varies in the two host countries.

Second, we delve into the assumption that individual trust is invariant over time. This conjecture has been the crucial but somewhat controversial argument on which the existing empirical analysis hinges. It is crucial because parents' and children' trust used in regression analyses are contemporaneously measured at the time of the interviews – while ideally they should be gauged at the time the transmission took place. These two measurements are equivalent only under the invariance of trust over time. It is a controversial hypothesis too since, for instance, in their U.S. longitudinal study, Poulin and Haase (2015) find that generalized trust changes with age. We exploit a three-wave panel dataset constructed from the German Socio-Economic Panel to model the dynamics of individual trust over a decade, distinguishing between its *permanent* and *transient* component. Within this framework, we test for the invariance of the permanent component of trust finding no evidence to reject it for mothers, children and fathers below the median age. Remarkably, we find that the transient component accounts for a large fraction of the variance of observed trust.

Third, building on the distinction between the two components, we argue that only permanent trust matters for the transmission process. To circumvent the unobservability of permanent trust, we show that the *unfeasible* regression of the permanent trust of children on the permanent trust of their parents is equivalent to estimating the regression of the observable counterparts of permanent trust using the lagged levels of parents' trust as instruments for the corresponding current level.

Last but not least, we show how to evaluate the relative importance of the intergenerational transmission and the residual siblings correlation to the formation of children's trust. As pointed out by Solon et al. (1991) in their study on the role of family background as a determinant of the economic status of children, the distinction between transient and permanent components is crucial.

In summary, the contribution of our study is threefold. 1) From a methodological

point of view, we provide a setup for the study of the evolution of individual trust over time – as well as of its intergenerational transmission – which can be adapted to other attitudes and cultural values; 2) We complement existing work on intergenerational transmission of trust by clarifying the conditions required to attach a structural interpretation to parameter estimates of a regression of the trust of children on the *contemporaneous* trust of their parents; 3) We separately identify the intergenerational correlation and the residual siblings correlation; 4) We derive both a simple estimator of the transmission parameters and an appropriate measure of the strength of the transmission process between generations.

The main results can be summarised as follows: i) in line with previous work, but with stronger effect, we find that it is the mother that has a substantial role on the transmission of permanent trust to children. On the other hand, the correlation between the permanent trust of fathers and their children is due to the *strong* correlation between the permanent trust of the two parents; ii) the strength of the intergenerational transmission is far from being substantial. We find that the parents' permanent trust accounts for approximately 20% of the variance of children's permanent trust; iii) on the other hand, the residual siblings correlation accounts for a large proportion of that variance, pointing to the existence of environmental factors shared by siblings which are independent on their parents' trust but are relevant to the formation of their own trust.

The remainder of the paper is organized as follows. Section 2 provides a brief review of the literature on cultural transmission. Section 3 outlines a framework where we introduce the distinction between permanent and transient trust and clarifies the conditions required to attach a structural interpretation of the regression parameters. Section 4 describes the data and the econometric model. Section 5 presents the results of the test for invariance of trust and of the estimates of intergenerational transmission parameters and residual siblings correlation. Section

6 follows with a discussion of our results and of their implication for the literature on long term persistence. Section 7 concludes.

4.2   Theoretical background

The first theoretical frameworks for the study of cultural transmission are due to Cavalli-Sforza and Feldman (1981) and Boyd and Richerson (1988), who apply models of evolutionary biology to the transmission of beliefs, preferences and norms. These works show how cultural traits can be acquired through learning and other forms of social interactions. Cultural transmission is seen as the result of the direct vertical socialization (the role played by parents), and the horizontal and oblique socializations (taking place in the society). Horizontal and oblique socializations can be described as imitation and learning behaviors, and refer mainly to the interactions with peers and the environment outside the family. Cultural transmission is different from genetic evolution, although the two can interact. The distinct effects of the cultural, environmental, and genetic factors on cognitive and non-cognitive skills of an individual is at the core of a lively debate on "nature" versus "nurture", which is the object of study of several disciplines, from behavioral genetics to social sciences (for a survey, see Sacerdote, 2011).

With the growing evidence of the persistence of ethnic and religious traits across generations, cultural transmission has recently gained new emphasis in the theoretical and empirical literature. It has been documented how migrants generally struggle to maintain specific traits of the culture of the country of origin. The cultural renaissance of several ethnic and religious communities in the U.S. (Orthodox Jews, for example) apparently endangered, is a significant case. Similarly, Africa has witnessed the persistence of tribal distinctions even after the emergence of na-

tional institutions.[5] During the last decade, Bisin and Verdier (2000, 2001) have significantly extended existing models. In particular, they introduced the parental socialization choice, which is motivated by what they call imperfect empathy. In their framework, parents are altruistic and care about children's choices, which are however evaluated using the parents' preferences. Children acquire traits through their parents' socialization choices and by learning from the social environment in which they grow up. Parents choose the optimal socialization effort taking into consideration also the environment. Specifically, parental choices depend on the distribution of the population with respect to the relevant trait. Bisin et al. (2009) further extend this model by analyzing multi-trait populations. In our empirical analysis, we endeavor to reconcile vertical and horizontal socialization within the family by quantitatively estimating and distinguishing the roles of the intergenerational correlation (deemed to capture vertical socialization) and the residual siblings correlation (which is thought to embody horizontal socialization).

4.3   Analytical framework

Let the *observable* level of trust for father i at time t be:

$$Tf_{it} = Tf_{it}^p + vf_{it} \tag{4.1}$$

where Tf_{it}^p is his *permanent* level of trust at time t and vf_{it} is a zero mean *transient* shock uncorrelated over time and unrelated to past, current and future values of the permanent trust. To fix ideas, let the evolution of the *permanent* level of trust over time be driven by the following model:

$$Tf_{it}^p = \rho Tf_{it-1}^p + (1-\rho)uf_{it} \tag{4.2}$$

---

[5]For a comprehensive review, see Bisin and Verdier (2005).

where uf_{it} is a *permanent* shock hitting Tf_{it}^p at time t. The permanent shock is uncorrelated over time and uncorrelated to past values of the permanent trust.

The intuition motivating this model is as follows. Tf_{it-1}^p is the level of *permanent* trust of father i at time $t-1$ summarizing past and current events relevant to his lasting belief on whether one can trust people. At time t he experiences the *unpredictable* shock (uf_{it}, vf_{it}). The component uf_{it} brings news relevant to the father's lasting belief, therefore updating his permanent trust according to equation (4.2). The component vf_{it} affects the *current* level of observable trust but does not bring any news relevant to the father's lasting belief. Accordingly, it does not leave any trace on the father's future belief.

With this simple framework we introduce two novel aspects. First, we distinguish between *observable* and *permanent* trust. Previous studies do not contemplate a transient component of trust, implicitly assuming that the transient shock vf_{it} is negligible. Presumably, only permanent trust is relevant for the intergenerational transmission, in that transient shocks – being uninformative about the updating process of father – are unlikely to be passed to the child. In the following, we will show that the latter is actually a testable implication of the model. Since the father's permanent trust is observable only through vf_{it}, the presence of a transient shock raises the classic measurement error problem to the purpose of the econometric identification of the intergenerational transmission.

The second aspect is that we delve into an important implicit assumption made in the existing literature, namely that trust is invariant over time. Empirically speaking, this assumption is rejected by the evidence we provide in Table 4.1. Still, it could be recast in our framework with reference to the permanent component of trust:

$$Tf_{it}^p = Tf_{it-1}^p = Tf_i^p \tag{4.3}$$

In the following, we show that this restriction has testable implications. Key to the feasibility of this test is the availability of panel data of individuals observed at least at *three* points in time.

The observable trust for mothers and children, Tm_{it} and Tc_{it}, their permanent trust, Tm_{it}^p and Tc_{it}^p, their transient and permanent shocks vm_{it}, vc_{it}, um_{it} and uc_{it}, as well as the equations linking permanent and observable trust and their dynamics are defined in the same manner as for fathers.

Turning to the transmission process, the equation relating the permanent trust of children to the *contemporaneus* permanent trust of their parents is:

$$Tc_{it}^p = \beta_0 + \beta_1 Tf_{it}^p + \beta_2 Tm_{it}^p + \epsilon_{it} \tag{4.4}$$

where the subscript t refers to the time at which the interview takes place. Note that information is collected at the same time for fathers, mothers and children. Our equation is similar to the one adopted in the existing literature on intergenerational transmission of trust, the important novelty being that in our framework we emphazise that it is the permanent trust which is passed on from parents to their children.

There are a number of issues one needs to carefully take into account in order to attach a structural interpretation to the coefficients in equation (4.4) and to consistently estimate them. Since in our dataset the permanent trust of children is observed (up to measurement errors) when they are at least 17 years old, modelling the mechanism of transmission from the early childhood to the late adolescence of children is not possible. This is a common and known limitation in this type of studies. However, a feasible and interesting alternative is to model the link between the level of permanent trust of children at age 17 – when the intergenerational transmission is presumably completed – and the trust that parents put in the process

up to that time. This is a kind of reduced form model linking inputs to the outputs and silently skipping over the circumstances inside the black box of the transmission process.

Even when recasting the problem this way, several issues persist. First, the trust that parents input in the transmission process is the one that covers the years when transmission took place, *not* the one we observe at the time we collect our survey data. This implies that to achieve a meaningful structural interpretation of the parameters in equation (4.4), one needs to explicitly account for the dynamics of the permanent trust of parents over time.

Second, to our purposes, we should ideally use the trust of children exactly at age 17 (or just above 17), i.e., right at the age by when the transmission is presumably completed. To gain statistical precision, we instead include in our sample children of *any* age (see Section 4.4.1 for details on sample selection). Hence the trust we observe for children refers to their age at the time of the interview, and not at 17. The major implication is that one needs to explicitly account for the dynamics of the permanent trust from age 17 onwards.

The third issue relates to the fact that, even though parameters in equation (4.4) deserved a structural interpretation, their identification would be problematic if the permanent trust of children and of their parents were affected by correlated permanent shocks, since this would induce the endogeneity of the permanent trust of parents. Related to this, an additional potential issue could be a reverse causal link going from the trust of children to the trust of their parents. Both these cases of endogeneity would lead to inconsistent OLS estimates.

Importantly, however, there is at least one special case when all the issues above *do not* arise. If the permanent trust of parents and children were time invariant, this means that observing them (up to measurement errors) at the time of the survey

would still provide a valid measure for both the permanent trust of parents when the transmission took place and the trust of children at age 17. Moreover, invariance of permanent trust would be sufficient to exclude the existence of permanent shocks, therefore ruling out the endogeneity of Tf_{it}^p and of Tm_{it}^p in equation (4.4). It is therefore crucial to test whether the hypothesis of trust invariance holds, before turning to the estimation of the transmission equation.

A fourth issue is that to obtain a *feasible* version of the transmission equation (4.4), one needs to replace the unobservable permanent trust of children and of their parents by their error-ridden observable counterparts:

$$Tc_{it} = \beta_0 + \beta_1 Tf_{it} + \beta_2 Tm_{it} + \epsilon_{it} + vc_{it} - \beta_1 vf_{it} - \beta_2 vm_{it} \qquad (4.5)$$

This raises the problem of how to estimate this feasible equation taking into account the endogeneity problem raised by the measurement errors in the observable trust of parents (as well as by the possible correlation between the measurement errors of parents and of their children).

Last but not least, key to the identification of the structural parameters in equation (4.4) is controlling for possible confounders which could be correlated to the trust of parents and children. To deal with this issue, we check the sensitivity of the estimates of (4.5) to the inclusion of several observables.

To quantify the *strength* of the transmission process, we follow the standard practice in the literature on intergenerational transmission and consider the fraction of the variance of Tc^p explained by (Tf^p, Tm^p). This depends both on the size of the coefficients β_1 and β_2 and on the degree of correlation between the permanent trust of parents:

$$\beta_1^2 var\{Tf_{it}^p\} + \beta_2^2 var\{Tm_{it}^p\} + 2\beta_1\beta_2 cov\{Tf_{it}^p, Tm_{it}^p\}. \qquad (4.6)$$

Distinguishing between observable and permanent trust is crucial to properly assess the extent to which children inherit trust from their parents. Even leaving aside the issue of how to estimate the coefficients β_1 and β_2 (see below Section 4.4.2), it is clear that the relevant R^2 should be evaluated with respect to the variance of Tc^p and not of Tc. Whether this distinction is important is an empirical issue to which we turn in Section 4.5.1, where we provide an estimate of the variance of the two components.

4.4 Econometrics

4.4.1 Data

Our *panel* of parents and children is drawn from the German Socio-Economic Panel (SOEP). The SOEP is a large longitudinal survey extensively used by economists and that has been the base for intergenerational studies (see, e.g., Dohmen et al., 2012). The survey was introduced in West Germany in 1984 and collected data on 12,000 households; in 1990, it was extended to include about 2,000 households from East Germany.[6] Two features of SOEP are key to our study. First, the survey "tracks" individuals, which means that those who move internally in Germany can still be followed over time, thereby reducing attrition. Second, it provides indicators to match children with their biological parents inside the panel. This feature is essential in order to construct families and observe them over time. A family is defined as the parental couple (mother and father) and their biological child(ren). Given the structure of SOEP it is not necessary for the family members to live in

---

[6]A detailed description of SOEP data can be found in Wagner et al. (2007). The panel has been assembled using PanelWhiz, see Haisken-DeNew and Hahn (2010) for details. In our analysis, we have used SOEP v31: Socio-Economic Panel (SOEP), data for years 1984-2014, version 31, SOEP, 2015, doi: 10.5684/soep.v31.

the same household in order to be observed in the panel.[7]

We included in the sample all couples who took part continuously into the survey in the waves 2003, 2008 and 2013 with at least one child of age 17 or older in 2013. Crucial to our analysis, this sample selection implies that we observe the trust of both parents in *three* time periods. The trust of children included in the sample is observed at least in 2013. For a subset of children, trust is also observable in either or both the previous waves (2003 and 2008) provided they were at least 17 and present during the survey (see Table B2 in the Appendix for details on the number of waves children are observed). The age distribution of fathers, mothers and children in 2013 is set in Figure 4.1.

The resulting panel comprises 1,652 children within 1,126 families. In the left panel of Table 4.2 we report the distribution of families by number of children included in our sample in 2013. The right panel reports the distribution of families by the *total* number of children in the same year. This total includes also children who are outside the sample (because they are still younger than 17, or because they were not originally sampled, or because of other reasons).

The key variable of our analysis is the *generalized trust.* This is recorded as the level of agreement with the statement "On the whole, one can trust people" on a four-point scale. [8] From the SOEP we derived additional variables, including gender, age, number of siblings, nationality, education and information on the place of residence when aged 15. We report summary statistics for these variables in Table B1 in the Appendix separately for mothers, fathers and children. We additionally include the average level of trust in the region, following the argument of Dohmen

---

[7]It is possible, however, that some children already left the households at the time of the first survey, and hence they are not part of the panel, despite being part of the family. Table 4.2 classifies families in terms of number of children who are part of the sample and total number of children (i.e., including those outside the sample).

[8]In our analysis we treat trust as if it were a continuous variable. In the Appendix we show that this approximation just adds a component to the measurement error on the permanent trust easily accommodated for in our model.

et al. (2012) that trust in the area of residence might affect children's trust or the transmission process.[9]

A remarkable aspect that emerges from a deeper inspection of the raw data – and not easily detectable with cross-sectional studies – is the variability of observed trust over time. Figure B1 in the Appendix shows the graphs of the difference in the level of trust for two consecutive waves, for both fathers and mothers. The graphs reveal that about 40% of parents report a different value of trust across two consecutive waves.

Additional evidence about this aspect comes from Table 4.1, where we report auto-covariance matrices of trust for the three waves forming our sample. The results are reported separately for fathers, mothers and children and for whether we include or not additional covariates in the computation of the covariances.[10] A cursory inspection of these matrices immediately reveals that the observable trust is far from stable over time, complementing what observed in Figure B1. The autocorrelation of order one is in the range 0.35 - 0.44. This is in stark contrast with the assumption – implicit in the existing literature on the transmission of trust – that trust is stable over time. We argue that the low degree of persistence observed in our data is due to the *transient* component of trust, as defined in equation (4.1). The consequent issue is that we need to establish whether the *permanent* component of trust – i.e., the one relevant for the intergenerational transmission – is invariant over time. The evidence in Table 4.1 will be the basis for our test for the invariance of permanent trust over the years developed in the next Section.

---

[9]Our definition of region follows closely Dohmen et al. (2012), and consists of the 96 policy-regions of Germany (also known as RORs).

[10]In the model with controls, covariances are calculated using residuals from a regression of trust on the full set of covariates (see Table B1). Note that the number of children reported in the Table is smaller than the total available in the sample, since only children observed in all three waves are used in the calculation of the autocovariances.

4.4.2 Specification testing and estimation

The testable implication of the invariance condition (4.3) written with reference to fathers is:

$$cov\{Tf_{i2003}, Tf_{i2008}\} = cov\{Tf_{i2003}, Tf_{i2013}\} = cov\{Tf_{i2008}, Tf_{i2013}\}. \qquad (4.7)$$

In words, if the permanent trust Tf_{it}^p does not vary over time and the variation over time of the observable trust Tf_{it} is only due to random shocks, then the covariance between the observable trust at time t and at time s equals the variance of the permanent trust for any choice of (t, s). That is, if the permanent trust is invariant over time, the three covariances in each panel of Table 4.1 should be equal (up to sampling variability).

Condition (4.7) could be violated due to different reasons. Particularly relevant to our case, it would not hold if the transient shocks were correlated at lag 1. It would also be violated if the equation driving the dynamics of Tf^p were as in equation (4.2). In both cases the covariance between observable trust in (2013, 2008) would be different from the corresponding covariance in (2013, 2003).

To implement the test, note that (4.7) is equivalent to:

$$cov\{Tf_{i2003}, Tf_{i2008} - Tf_{i2013}\} = cov\{Tf_{i2008}, Tf_{i2003} - Tf_{i2013}\} = cov\{Tf_{i2013}, Tf_{i2003} - Tf_{i2008}\} = 0.$$

$$(4.8)$$

To test the first condition, it is sufficient to perform the regression of $Tf_{i2008} - Tf_{i2013}$ on Tf_{i2003} (or the other way around) and check whether the regression coefficient is zero. The same applies for the remaining two conditions. Clearly, a panel of observations of at least length three is needed to perform this test.

On accepting the invariance condition (4.7), the decomposition of the variance of

the observable trust into its components due to the permanent trust and to the transient shock, respectively, proceeds the following way.

$$var\{vf_{it}\} = var\{Tf_{it}\} - var\{Tf_i^p\}. \tag{4.9}$$

The variance of Tf_i^p follows from condition (4.7). To estimate the parameters of the feasible transmission equation (4.5), note that Tf_{it-1} and Tm_{it-1} are valid instrumental variables for Tf_{it} and Tm_{it} provided that the transient shock is not correlated over time. Also, note that with a panel of length three the model is *overidentified* since Tf_{it-2} and Tm_{it-2} are valid instruments as well.

These settings provide the basis for an additional test of the hypothesis of no autocorrelation of the transient shock. Under the alternative hypothesis of autocorrelated shocks, the IV at time $t-1$ is plausibly more correlated to the disturbance term in equation (4.5) than the IV at time $t-2$. Therefore, the Sargan overidentification test should detect a violation of the null hypothesis. The same test is in principle useful also to detect a violation of our conjecture that transient shocks of parents' trust are irrelevant for the transmission process. If these shocks were otherwise relevant, the exclusion restriction on our candidate IV would not hold since past values of parents' observable trust would matter for current values of children's observable trust, even conditional on the current values of parents' permanent trust. Since the degree of violation of the exclusion restriction is likely to vary with the lag of the instrument, the Sargan overidentification test should detect whether the null hypothesis does not hold.

To summarize the strength of the transmission of trust from parents to children we use the conventional R^2 of equation (4.4). The variance of Tc^p explained by the regression can be calculated according to expression (4.6). The variance of Tc_{it}^p, Tf_{it}^p and Tm_{it}^p are derived as a corollary of the invariance condition in Equation (4.7).

A convenient way to recover the covariance between the permanent trust of parents is to perform a regression of Tf_{it} on Tm_{it} using Tm_{it-1} as an IV to eliminate the bias due to the measurement error. This is a consistent estimate of the regression coefficient of Tf_{it}^p on Tm_{it}^p. The next step is to rescale the estimated coefficient by $var\{Tm_{it}^p\}$ to obtain the covariance between the trust of parents.

4.4.3 Sibling correlation in trust

To investigate further on the role of the family environment in the transmission process, we exploit the variation stemming from families with more that one child in the sample – which are about 38% of our sample (see Table 4.2). The availability of siblings in the data allows to estimate a transmission equation which includes a family specific unobservable effect. This can be achieved by estimating a modified version of equation (4.4):

$$Tc_{ij}^p = \beta_0 + \beta_1 Tf_{ij}^p + \beta_2 Tm_{ij}^p + \alpha_j + \epsilon_{ij} \tag{4.10}$$

The subscript ij refers to children belonging to the same family j (we drop the time suffix to ease the exposition). Similarly to Bingley and Cappellari (2012), we identify both the parental and sibling effects by estimating intergenerational and sibling correlations within a unified framework. The residual sibling effect α_j includes parental influences not captured by the direct transmission of trust, as well as other environmental factors shared by siblings that are independent from the parents. Schools, friendship networks and other circumstances operating at the community level are examples of this residual sibling effect. Note that our framework allows us to identify the direct transmission of trust from parents to children, but not *other* channels of intergenerational transmission of trust that are independent from parents' trust.

There are two important remarks about the identification of $var\{\alpha_j\}$ and its interpretation. First, since the identification of the variance is based on the between-siblings covariance of the residuals from the feasible IV regression of Tc on Tf and Tm, $var\{\alpha_j\}$ could partially capture the correlation between the transient shocks of siblings. However, this is a testable implication. Under the null hypothesis of no correlation between the transient shocks of siblings, the covariance between the trust of one sibling in 2013 and the trust of another sibling at, say, time t, does not depend on t since it is equal to the covariance between the permanent trust of the two siblings. We implement this test in the same way as in equation (4.8).

Second, $var\{\alpha_j\}$ strictly refers to families with at least two children in the sample (about 38%). Note, however, that the overall (i.e., including out of sample) number of siblings – and thus of families with more than one child – is much larger. The second panel of Table 4.2 shows that nearly 85% of the families in our sample have more than one child, meaning that the estimate of $var\{\alpha_j\}$ is virtually representative of the majority of our sample.

4.5 Results

4.5.1 Testing for invariance of permanent trust

Table 4.3 presents the results of the test for invariance of permanent trust separately for fathers, mothers and children. The pattern of autocovariance for mothers and children provides clear evidence that observable trust is equal to a time invariant component plus a random shock. This result holds true even after stratifying by age (with the exception of few correlations statistically significant at the 10% level). On the other hand, the null hypothesis is rejected for fathers. By inspecting again Table 4.1, it is clear that this violation is driven by the difference between

$cov\{Tf_{i2013,Tf_{i2008}}\}$ and $cov\{Tf_{i2013,Tf_{i2003}}\}$. Table 4.3 further shows that the null hypothesis is rejected for relatively old fathers, but not for younger fathers. This result is unaffected by the inclusion of the full set of regressors. We will take into account the results of the invariance tests in the analysis of transmission of trust.

The most important consequence of the tests in Table 4.3 for the identification of the transmission parameters is that by age 17 (and above) – i.e., the age at which the transmission of trust is presumably completed – the permanent trust of children is *not* affected by permanent shocks (at least over the time span 2003 to 2013), irrespectively of their age. Even if, given the available data, it is not possible to directly test whether this occurs also at ages before 17, we proceed in our analysis maintaining that in Equation (4.4) there is no endogeneity of Tf^p and Tm^p attributable to permanent shocks correlated between parents and children.

Table 4.4 presents the decomposition of the variance of observable trust into the permanent trust and transient shock components. The main aspect is that for mothers and children the permanent trust accounts for approximately *one third* of the variance of the observed trust, while the remaining is attributable to the noise of the transient shock. This fraction is slightly larger for fathers, and accounts for about half of the variance of the observable trust.

4.5.2 Estimating the transmission parameters

Table 4.5 presents the results of the estimation of the feasible transmission equation (4.5). We report estimations for OLS and IV models; for the latter we estimate both regression pooling together all siblings and one with a family random effect. The instruments used are the first and second lag of trust for both fathers and mothers. The validity of our instruments is supported by the absence of autocorrelation of the transient shocks, for which we provided evidence (for mothers and younger

fathers) in the previous section. The Durbin-Wu-Hausman test strongly rejects the hypothesis of exogeneity of Tf and Tm while the Sargan overidentification test does not reject the null hypothesis adding further evidence in favor of the validity of our assumption of no autocorrelation for the transient shock.

Despite the usual loss of precision, the IV estimate for the coefficient of mothers is strongly significant and three times larger than the OLS estimate. On the other hand, the estimate for the coefficient of father is remarkably similar to the OLS (with the latter being estimated with higher precision). This pattern holds for both models with and without full controls. To corroborate this result, we replicated the regression on the subsample that contains fathers of age below the median, i.e., those for which there is no evidence of violation of the restriction of permanent trust invariance. Results are essentially unaffected, even if the effect of mothers is even stronger – see Table B4.

The evidence that accounting for measurement errors makes a major difference for the estimated coefficient for mothers while it does not matter at all for the estimated coefficient for fathers might seem puzzling. In the Appendix we show in fact that this is in line with known results in the literature on measurement errors, according to which when both regressors are affected by measurement errors, the sign of the resulting bias is a priori uncertain.

We also replicated the analysis by splitting the sample by gender of the child (Table B5 in the Appendix). Results are qualitatively similar to the baseline, although the effect of mother's trust is stronger for female children.

The key result of this analysis is that a clear hierarchy emerges in the roles of mothers and fathers with the formers being more influential in the transmission process. The order of magnitude of the estimated coefficients are in line with those found by Dohmen et al. (2012), even if our results show a sharper difference between

mothers' and fathers' roles.

As for the strength of the intergenerational transmission, we summarize it as the fraction of the variance of Tc_{it}^p explained by the permanent trust of parents. As a first step, we estimate the strength of the correlation between the permanent trust of fathers and mothers as outlined at the end of the previous section. The IV estimate of the regression of Tf^p on Tm^p is approximately equal to the correlation coefficient between the two variables and is 0.647 (s.e. 0.076) for the model without controls and 0.571 (s.e. 0.081) for the model with full controls.[11]. The R^2 pertinent to the transmission process is about 0.23 in the model without controls and drops to 0.17 in the model with full controls. In words, this means that a large fraction of the variability in the permanent trust of children is *not* attributable to the parents' permanent trust. Failing to distinguish between permanent trust and transient shocks would result in a severe underestimation of the strength of the transmission process.

Turning to the results of the random effect specifications, we notice that the pattern of estimates are similar to the IV model estimated without considering sibling correlations. The striking result, however, lies in the estimated contribution of the family-specific unobservable a_j component to the variance of Tc^p. The ratio of $var\{\alpha_j\}$ to $var\{Tc^p\}$ is *three* times larger than the contribution of the parents' permanent trust. This effect becomes slightly larger when controls are included. Taken together, family-specific characteristics – whether observable (permanent trust) or unobservable (α_j) – account for nearly 90% of the variance of children's permanent trust. [12]

Table B6 in the Appendix shows that the sibling correlation estimate is not biased by the correlation between transient shocks of siblings. Only one out of the six tests

---

[11]Recall that the variance of Tf^p is approximately equal to the variance of Tm^p - see Table 4.4.

[12]The inclusion of controls in the regression yields similar results, with only marginal changes in the relative contribution of the parents' permanent trust and of α_j.

rejects the null hypothesis, and only at the 10% significance level.

The empirical relevance of α_j in Equation (4.10) casts suspicion about a possible omitted variable bias in the estimation of the coefficients of parents' permanent trust. However, note that such bias would not affect the overall quantitative relevance about the family-specific characteristics to the transmission process.

4.6 Discussion

In Table 4.6, we present the decomposition of the variance of the trust of children in 2013. Two striking facts emerge. First, the observed variability of the children's trust is dominated by random shocks – nearly two thirds of the total variance. On the other hand, permanent trust explains the remaining one third. Note, we identify the size of these components by exploiting the longitudinal variation of trust.

Second, less than one fourth of the variance of children's *permanent* trust is attributable to the direct transmission of permanent trust from parents. We identify the size of this component by exploiting the correlation between children's and parents' trust (accounting for the attenuation bias due to transient shocks).

Approximately two thirds of the variance of children's permanent trust is attributable to residual sibling correlations. This captures characteristics of the environment – within or outside the family – which are shared by siblings. In principle, also this component might include intergenerationally-transmitted trust through channels that work independently from parents' trust.

Overall, direct transmission of trust from parents and residual sibling correlations account for just less than 90% of the variance of the permanent trust of children. Even if the evidence we provide emphasizes the major role played by the family environment in shaping children's trust, it is clear that the direct transmission from

parents plays a minor role in the persistence of trust over generations.

One challenge is how to reconcile our evidence with some results coming from the literature on long term persistence of trust. For example, Guiso et al. (2008a) show that the establishment of free cities in Center-North Italy during the medieval period generated a positive shock in the accumulation of social capital in the affected municipalities which is perceivable even nowadays. In a companion paper, the authors develop a theoretical model to show how the intergenerational transmission of trust is compatible with their empirical evidence (Guiso et al., 2008b).

A possible argument to reconcile our evidence of a weak "short run" intergenerational transmission effect with the results by Guiso et al. (2008a) comes from the literature on intergenerational mobility of income and wealth. Building on Güell et al. (2015), Barone and Mocetti (2016) argue that intergenerational mobility of earnings up to the end of the 19th century in Florence might have been much lower than what observed today. The authors put forward the idea that in less mobile societies like those prevailing in the pre-industrial era, intergenerational transmission took place thanks to a variety of social institutions and not only through the direct parent-child transmission. Additional arguments postulating the environment as a driver of the long term persistence of trust come from simple models of cultural transmission (see the review in Bisin and Verdier, 2011). In these models, if trust is not vertically transmitted, the child draws it at random from the population. Our results suggest a possible "amendment" to these frameworks: the random draw from the population is sibling-specific rather than being individual-specific, i.e., it affects in the same manner the trust of children who grew up in the same family environment.

4.7   Summary and conclusion

We study the intergenerational transmission of trust using a sample of parents and children drawn from the German Socio-Economic Panel. Our key asset is the availability of longitudinal information, which is crucial to distinguish between two different ways the family might shape children's trust: the direct transmission from parents to children and the influence exerted by a broadly defined family environment shared by siblings, such as parental influences not captured by the direct transmission of trust, as well as other local effects shared by siblings and that are independent on the parents (e.g. schools, friendship networks or other factors operating at the community level).

Longitudinal information is also essential in order to disentangle the two components of observable trust, namely permanent trust and transient shock. This distinction is vital because it is plausible to postulate that parents transmit to their children only their permanent trust, i.e., their lasting belief on whether one can trust other people, while the transient shock – being temporary by construction – is unlikely to be passed to the children. Our argument is akin to the point made by Solon et al. (1991) in their analysis of intergenerational transmission of economic status. We show that permanent trust only accounts for *one third* to *a half* of the observed cross sectional variability of trust. To the purpose of the econometric identification of the transmission parameters, the remaining part of the variability rises the classic measurement error problem.

Next, with our panel data we can test the invariance of trust over time – an important assumption which is implicitly maintained in the previous literature but that has not been proven empirically before. In particular, we show that permanent trust is invariant for mothers and for children over the time window 2003 to 2013, while this holds true only for younger fathers.

Based on the evidence that permanent trust is invariant, we model the relationship between the permanent trust of children and the contemporaneous permanent trust of their parents. The structural interpretation that we give to the parameters of this equation is that they capture the link between the trust that parents input in the transmission process (up to when their children are 17 year old) and the level of permanent trust of their children at the time the transmission is completed. The estimation of these structural parameters requires replacing the unobservable permanent trust of children and of their parents by their error-ridden observable counterpart. The importance of having longitudinal information is once again evident since we can use the lag trust of parents as a valid instrumental variable to mitigate the measurement error problem. The remarkable result that transpires is that mothers play a stronger role than fathers in the transmission process. This result is in line with previous findings (see, for instance, Dohmen et al., 2012), but the difference found in the parental roles is stronger.

Finally, exploiting the availability of families with more than one child in our sample, we estimate the variance of the unobservable family-specific effect which is thought to represent additional environmental factors shared by the siblings and relevant to their permanent trust. The variance explained by this component is *three* to *four* times larger than the variance explained by the permanent trust of parents. Taken together, the intergenerational correlation and the residual sibling correlation account for approximately 90% of the variance of the permanent trust of children. In conclusion, while the family environment in which children grew up determines most of their *permanent* trust, the direct role of intergenerational transmission is rather exiguous.

Tables and Figures
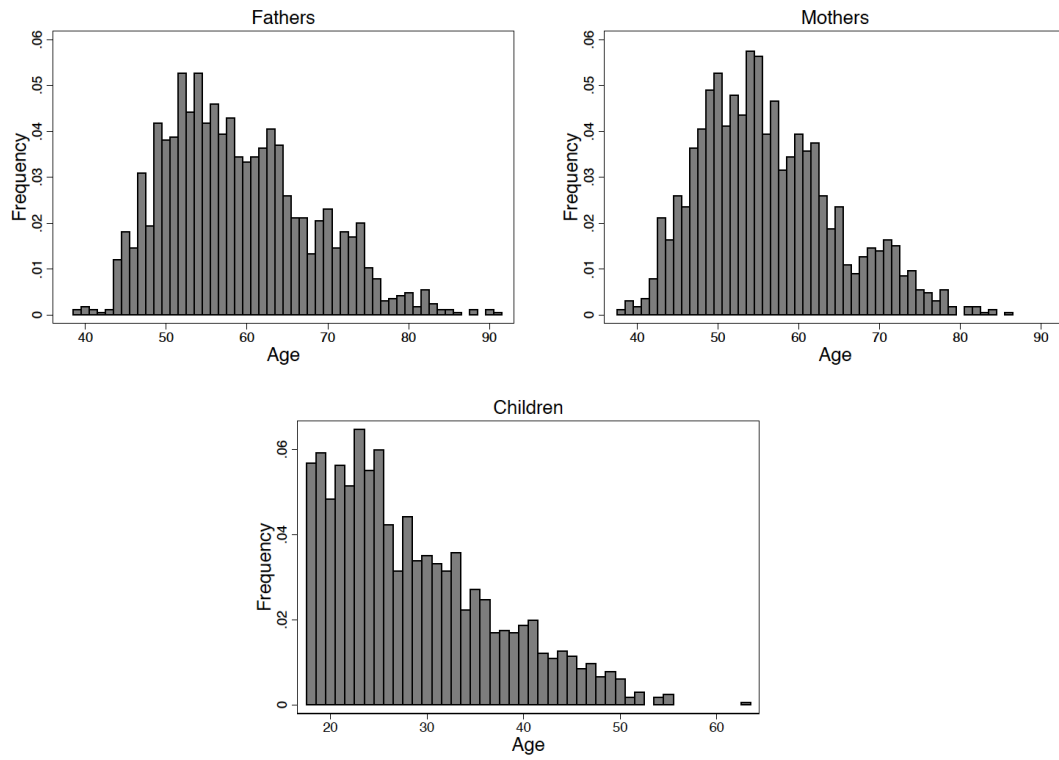
Source: SOEP waves 2003, 2008 and 2013.

Figure 4.1. Age distribution in 2013

Table 4.1. Autocovariance matrices for observable trust

	Fathers			Mothers			Children		
	No controls								
	2003	2008	2013	2003	2008	2013	2003	2008	2013
2003	0.4400	0.1688	0.1381	0.4057	0.1351	0.1358	0.4409	0.1307	0.1262
2008		0.4370	0.1698		0.4112	0.1445		0.3761	0.1334
2013			0.3908			0.3805			0.3585
	With controls								
	2003	2008	2013	2003	2008	2013	2003	2008	2013
2003	0.3483	0.1039	0.1019	0.3204	0.0917	0.1094	0.3675	0.1028	0.0672
2008		0.3682	0.1372		0.3045	0.0942		0.3471	0.1021
2013			0.3153			0.2956			0.2574
N		1126			1126			798	

Source: SOEP waves 2003, 2008 and 2013

Sample is composed by fathers, mothers and children for whom trust is observed in all three waves.

The sample of children in the table is smaller than the number used in the analyses (N=1652) since some children turn 17 after 2003 and a few others were added to SOEP in waves subsequent to 2003.

Control variables include for parents and children: age, education (No Degree or In School / Secondary School Degree / Intermediate School Degree / Technical, Upper Secondary or Other Degree), nationality (German / foreign), number of siblings, place where raised up to age 15 (unreported / small city / medium city / large city / countryside). For children, gender and the average level of trust in the ROR in 2013 are also included.

Table 4.2. Distribution of families by number of children

Number of children	Sample		Overall	
	Frequency	Percentage	Frequency	Percentage
1	699	62.08	176	15.63
2	349	30.99	576	51.15
3	59	5.24	249	22.11
4	17	1.51	89	7.90
5	2	0.18	24	2.13
6 or more	-	0	12	1.07
Total	1126	100	1126	100

Source: SOEP wave 2013.

Sample is composed by families with fathers and mothers for whom trust is observed in all three waves and with children for whom trust is observed at least in wave 2013.

The first and second column refer to the distribution of families in the sample by the number of children included in the sample reported in 2013. The third and fourth column refer to the distribution of families by the overall number of children (i.e., including also children outside the sample) reported in 2013. The number of children in each family is calculated using information on the number of siblings reported by the children in the sample.

Table 4.3. Invariance of permanent trust

Dep. Variable	Main regressor	No controls			With controls		
		Fathers	Mothers	Children	Fathers	Mothers	Children
		All age groups					
$T_{2013} - T_{2003}$	T_{2008}	0.0025	0.0227	0.0070	0.0906	0.0083	−0.0018
		(0.0371)	(0.0385)	(0.0493)	(0.0646)	(0.0623)	(0.0727)
$T_{2013} - T_{2008}$	T_{2003}	−0.0697**	0.0017	−0.0104	−0.0055	0.0555	−0.0967
		(0.0344)	(0.0365)	(0.0413)	(0.0614)	(0.0568)	(0.0597)
$T_{2008} - T_{2003}$	T_{2013}	0.0813**	0.0227	0.0201	0.1119*	−0.0516	0.1357*
		(0.0359)	(0.0374)	(0.0498)	(0.0672)	(0.0623)	(0.0758)
N		1126	1126	798	385	385	397
		Below median age					
$T_{2013} - T_{2003}$	T_{2008}	−0.0041	−0.0394	0.0218	−0.0010	−0.0374	−0.0989
		(0.0489)	(0.0535)	(0.0664)	(0.0702)	(0.0700)	(0.0883)
$T_{2013} - T_{2008}$	T_{2003}	−0.0523	−0.0158	−0.0715	0.0064	0.0714	−0.1782**
		(0.0458)	(0.0504)	(0.0546)	(0.0818)	(0.0855)	(0.0732)
$T_{2008} - T_{2003}$	T_{2013}	0.0606	−0.0265	0.1168*	−0.0093	−0.1156	0.1442
		(0.0501)	(0.0532)	(0.0661)	(0.0907)	(0.0831)	(0.0949)
N		602	584	418	226	225	230
		Above median age					
$T_{2013} - T_{2003}$	T_{2008}	0.0171	0.0994*	−0.0119	0.1701	0.0960	0.0909
		(0.0569)	(0.0542)	(0.0734)	(0.1302)	(0.1161)	(0.1231)
$T_{2013} - T_{2008}$	T_{2003}	−0.0863*	0.0247	0.0620	−0.0217	0.0545	0.0078
		(0.0516)	(0.0523)	(0.0625)	(0.1169)	(0.0925)	(0.0979)
$T_{2008} - T_{2003}$	T_{2013}	0.1041**	0.0782	−0.0832	0.1850*	0.0392	0.1061
		(0.0516)	(0.0524)	(0.0766)	(0.1094)	(0.1069)	(0.1523)
N		524	542	380	159	160	167

Source: SOEP waves 2003, 2008 and 2013.
*/**/*** indicate significance at the 0.1/0.05/0.01 level.
Sample is composed by fathers, mothers and children for whom trust is observed in all three waves.
The sample of children in the table is smaller than the number used in the analyses (N=1652) since some children turn 17 after 2003 and a few others were added to SOEP in waves subsequent to 2003.

Table 4.4. Variances of permanent trust and transient shock

	Permanent trust		Transient shock	
	No controls			
	2003	2008	2003	2008
Fathers	0.2063	0.2076	0.2338	0.2294
Mothers	0.1344	0.1437	0.2712	0.2675
Children	0.1355	0.1382	0.3055	0.2379
	With controls			
	2003	2008	2003	2008
Fathers	0.1058	0.1398	0.2424	0.2284
Mothers	0.0768	0.0789	0.2436	0.2256
Children	0.1571	0.1561	0.2104	0.1910

Source: SOEP waves 2003, 2008 and 2013
Sample is composed by fathers, mothers and children for whom trust is observed in all three waves
Permanent trust derived using Equation (4.9) under accepting the invariance condition in Equation (4.7) and the covariances from Table 4.1.

Table 4.5. Intergenerational transmission

	No controls			With controls		
	OLS	**IV**	**IV R.E.**	**OLS**	**IV**	**IV R.E.**
Father's trust	0.1014***	0.1439	0.1637	0.0801**	0.1113	0.1311
	(0.0318)	(0.1132)	(0.1272)	(0.0321)	(0.1025)	(0.1204)
Mother's trust	0.0954***	0.3275**	0.3695**	0.0690*	0.2792**	0.3201**
	(0.0334)	(0.1479)	(0.1585)	(0.0356)	(0.1346)	(0.1474)
Constant	1.5984***	1.7859***	2.0041***	0.9471	1.4777	1.8808*
	(0.1236)	(0.6501)	(0.6019)	(0.5778)	(1.0359)	(1.0387)
Partial R^2 Eq F.		0.545	0.458		0.530	0.433
Partial R^2 Eq M.		0.542	0.457		0.530	0.434
F-stat Eq F.		0.225	0.221		0.194	0.191
F-stat Eq M.		0.182	0.177		0.165	0.164
DWH χ^2		5.3382	6.0637		4.1932	4.9098
p-value Sargan		0.2308	0.4326		0.3433	0.5296
$R^2(Tf^p, Tm^p)$	0.0437	0.2342	0.2997	0.0235	0.1481	0.1979
$R^2(a_j)$			0.1038			0.1462
N families			427			427
N	953	928	928	953	928	928

Source: SOEP waves 2003, 2008 and 2013.

*/**/*** indicate significance at the 0.1/0.05/0.01 level.

Sample is composed by fathers and mothers for whom trust is observed in all three waves and children for whom trust is observed in wave 2013.

OLS: Ordinary least squares; IV: Instrumental variable; IV R.E.: Instrumental variable with random effects. In the IV models, observable trust of fathers and mothers in 2013 is instrumented by their observable trust in 2008 and 2003.

Partial R^2 refers to the Shea's partial R-squared of the first stages.

F-stat refers to the F-statistic for the joint significance of the instruments in the first stages.

p-val Sargan indicates the p-value of the Sargan test for overidentification.

DWH χ^2 refers to the Durbin-Wu-Hausman test for endogeneity.

$R^2(Tf^p, Tm^p)$ refers to the unfeasible regression for the permanent trust. See Equation (4.4).

$R^2(a_j)$ refers to the variance explained by unobservable characteristics of the family.

Table 4.6. Decomposition of observed variance of children in 2013

	No controls	With controls
Variance of observable trust in 2013	0.3585	0.2574
Variance of transient shock*	0.2379	0.1910
Variance of Tc^p	0.1206	0.0664
Intergenerational transmission	0.0276	0.0030
Household environment	0.0790	0.0000
Residual component	0.0140	0.0634

Source: SOEP waves 2003, 2008 and 2013

Sample is composed by children for whom trust is observed in 2013.

*The variance of the transient shock in 2013 is not identifiable and is thus replaced by the variance in 2008.

Components estimated using (4.9) and results from regressions in Table 4.5.

Table 4.7. Sibling correlations

	2-siblings			All siblings	First 2 siblings
Sibling's trust	1.0151*	1.0322**	0.8922**	1.2343	0.6669
	(0.5213)	(0.5183)	(0.4327)	(1.2814)	(0.5365)
Constant	−0.0437	−0.1201	0.2426	−0.5473	0.6904
	(1.2362)	(1.2179)	(1.0208)	(2.8635)	(1.2114)
Controls	N	Y	Y	Y	Y
N	1129	1129	1211	1873	1645
			First stage		
Born in first quarter	0.1146***	0.1037**	0.1140***	0.0402	0.0711*
	(0.0410)	(0.0450)	(0.0437)	(0.0401)	(0.0369)
Partial R^2	0.005	0.005	0.006	0.001	0.002
F-stat	5.225	5.259	6.750	1.009	3.740
OLS estimates	0.3033***	0.2994***	0.2729***	0.3166***	0.3171***
	(0.0420)	(0.0415)	(0.0396)	(0.0309)	(0.0337)

Source: SOEP waves 2003, 2008 and 2013.
*/**/*** indicate significance at the 0.1/0.05/0.01 level.
Sample is composed by: siblings of 2-siblings families (Col I-III); siblings from all type of families (col IV) and first 2 siblings from all type of families. Data refer to 2003 only, except in column III, which refers to 2003, 2008 and 2013.
All models are using instrumental variables.
Partial R^2 refers to the Shea's partial R-squared of the first stages.
F-stat refers to the F-statistic for the joint significance of the instruments in the first stages.

Table 4.8. Sibling correlations - first born

	2-siblings			All siblings	First 2 siblings
Sibling's trust	1.5233	1.8707	1.6254	0.0286	0.0286
	(2.5834)	(3.3939)	(2.2409)	(1.0878)	(1.0878)
Constant	−1.2530	−1.6789	−1.1544	2.2664	2.2664
	(6.1364)	(7.5357)	(4.9964)	(2.4312)	(2.4312)
Controls	N	Y	Y	Y	Y
N	544	544	547	793	793
			First stage		
Born in first quarter	0.0721	0.0384	0.0512	0.0512	0.0512
	(0.0676)	(0.0702)	(0.0698)	(0.0590)	(0.0590)
Partial R^2	0.001	0.001	0.001	0.001	0.001
F-stat	0.369	0.299	0.538	0.752	0.752
OLS estimates	0.3011***	0.3008***	0.3054***	0.3152***	0.3152***
	(0.0425)	(0.0416)	(0.0418)	(0.0339)	(0.0339)

Source: SOEP waves 2003, 2008 and 2013.
*/**/*** indicate significance at the 0.1/0.05/0.01 level.
Sample is composed by: siblings of 2-siblings families (Col I-III); siblings from all type of families (col IV) and first 2 siblings from all type of families. Data refer to 2003 only, except in column III, which refers to 2003, 2008 and 2013.
All models are using instrumental variables.
Partial R^2 refers to the Shea's partial R-squared of the first stages.
F-stat refers to the F-statistic for the joint significance of the instruments in the first stages.

Table 4.9. Sibling correlations - second born

	2-siblings			All siblings	First 2 siblings
Sibling's trust	0.8980**	0.8836**	0.7198*	1.0661	1.0661
	(0.4342)	(0.4454)	(0.3936)	(0.7145)	(0.7145)
Constant	0.2346	−0.0912	0.4829	−0.3773	−0.3773
	(1.0290)	(1.0580)	(0.9324)	(1.5775)	(1.5775)
Controls	N	Y	Y	Y	Y
N	585	585	664	852	852
			First stage		
Born in first quarter	0.1415***	0.1500**	0.1545***	0.0818*	0.0818*
	(0.0511)	(0.0589)	(0.0567)	(0.0488)	(0.0488)
Partial R^2	0.011	0.010	0.011	0.003	0.003
F-stat	7.013	6.383	7.343	2.818	2.818
OLS estimates	0.3055***	0.3062***	0.2493***	0.3246***	0.3246***
	(0.0442)	(0.0440)	(0.0411)	(0.0360)	(0.0360)

Source: SOEP waves 2003, 2008 and 2013.

*/**/*** indicate significance at the 0.1/0.05/0.01 level.

Sample is composed by: siblings of 2-siblings families (Col I-III); siblings from all type of families (col IV) and first 2 siblings from all type of families. Data refer to 2003 only, except in column III, which refers to 2003, 2008 and 2013.

All models are using instrumental variables.

Partial R^2 refers to the Shea's partial R-squared of the first stages.

F-stat refers to the F-statistic for the joint significance of the instruments in the first stages.

Table B1. Summary statistics

	Children	Fathers	Mothers
Trust: 2013	2.2912	2.368	2.3117
	(0.6172)	(0.6227)	(0.611)
Trust: 2008[+]	2.3511	2.3874	2.3372
	(0.624)	(0.654)	(0.6347)
Trust: 2003[++]	2.3679	2.3565	2.3341
	(0.6637)	(0.6535)	(0.6359)
Males*	0.5097	1	0
	(0.5001)	(0)	(0)
Age	28.7464	58.8469	56.1858
	(8.4223)	(8.91)	(8.4305)
Number of siblings	1.5145	2.0048	2.1483
	(1.0897)	(1.734)	(1.7952)
German national*	0.9643	0.9437	0.9395
	(0.1856)	(0.2306)	(0.2385)
Education: No Degree/In School*	0.1907	0.0212	0.026
	(0.393)	(0.144)	(0.1593)
Education: Secondary School Degree*	0.112	0.3493	0.2869
	(0.3154)	(0.4769)	(0.4525)
Education: Intermediate School Degree*	0.2669	0.2887	0.4201
	(0.4425)	(0.4533)	(0.4937)
Education: Technical/Upper Secondary/Other Degree*	0.4304	0.3408	0.2669
	(0.4953)	(0.4741)	(0.4425)
Place raised at 15: Unreported*	0.1659	0.1822	0.1901
	(0.3721)	(0.3861)	(0.3925)
Place raised at 15: Large city*	0.1992	0.1489	0.1562
	(0.3995)	(0.3561)	(0.3631)
Place raised at 15: Medium city*	0.2488	0.2312	0.2191
	(0.4324)	(0.4217)	(0.4138)
Place raised at 15: Small city*	0.3087	0.4195	0.4159
	(0.4621)	(0.4936)	(0.493)
Place raised at 15: Countryside*	0.0775	0.0182	0.0188
	(0.2674)	(0.1336)	(0.1357)
N		1652	

Source: SOEP waves 2003, 2008 and 2013.
Sample is composed by fathers and mothers for whom trust is observed in all three waves and children for whom trust is observed in wave 2013.
* refers to dummy variables.
[+] trust is calculated on the subsample of 1179 children for whom trust is observed in 2013 and in 2008; [++] trust is calculated on the subsample of 810 children for whom trust is observed also in 2013 and 2003.

Table B2. Structure of the panel

	Number of waves		
	1	2	3
Trust: Children, 2013	2.295	2.2926	2.2882
	(0.6526)	(0.6131)	(0.5987)
Trust: Fathers, 2013	2.2777*	2.4173	2.396
	(0.6367)	(0.5881)	(0.6265)
Trust: Fathers, 2008	2.3557	2.4427	2.3784
	(0.6751)	(0.6526)	(0.6412)
Trust: Fathers, 2003	2.3514	2.3766	2.3496
	(0.6901)	(0.6354)	(0.641)
Trust: Mothers, 2013	2.269	2.3181	2.3333
	(0.6336)	(0.5652)	(0.619)
Trust: Mothers, 2008	2.282*	2.3562	2.3596
	(0.6382)	(0.6106)	(0.6433)
Trust: Mothers, 2003	2.3297	2.3613	2.3233
	(0.6425)	(0.6322)	(0.6342)
N	461	393	798

Source: SOEP waves 2003, 2008 and 2013.
Columns header refers to the number of waves children are observed. 1 means that children are observed only in 2013; 2 means that children are observed in 2013 and in 2008 or in 2013 and in 2003; 3 means that children are observed across the three waves
* indicates that trust values are significantly different with respect to those pertinent to the subsample of children observed across the three waves (i.e. p-values of a t-test for the difference of two means are below 0.05).

Table B3. Correlation between transient shocks - parents/children

Dep. Variable	Main regressor	No controls		With controls	
		Fathers	Mothers	Fathers	Mothers
		All age groups			
$T_{2013} - T_{2003}$	Tc_{2013}	0.0516*	–0.0035	0.0381	0.0096
		(0.0294)	(0.0305)	(0.0377)	(0.0391)
$T_{2013} - T_{2008}$	Tc_{2013}	0.0148	0.0115	0.0224	–0.0091
		(0.0278)	(0.0308)	(0.0357)	(0.0401)
$T_{2008} - T_{2003}$	Tc_{2013}	0.0368	–0.0150	0.0157	0.0187
		(0.0284)	(0.0319)	(0.0362)	(0.0428)
N		1652	1652	877	877
		Below median age			
$T_{2013} - T_{2003}$	Tc_{2013}	0.0323	–0.0409	0.0245	–0.0495
		(0.0387)	(0.0382)	(0.0445)	(0.0495)
$T_{2013} - T_{2008}$	Tc_{2013}	0.0063	–0.0391	0.0229	–0.0846*
		(0.0360)	(0.0383)	(0.0428)	(0.0489)
$T_{2008} - T_{2003}$	Tc_{2013}	0.0259	–0.0018	0.0016	0.0351
		(0.0373)	(0.0404)	(0.0406)	(0.0527)
N		891	875	500	502
		Above median age			
$T_{2013} - T_{2003}$	Tc_{2013}	0.0796*	0.0446	0.0902	0.0992
		(0.0451)	(0.0492)	(0.0624)	(0.0608)
$T_{2013} - T_{2008}$	Tc_{2013}	0.0291	0.0773	0.0086	0.0738
		(0.0434)	(0.0499)	(0.0622)	(0.0718)
$trust_o b2013$				–0.0142	0.1170
				(0.0607)	(0.0724)
trustreg				0.4823	–0.1354
				(0.5820)	(0.4102)
$T_{2008} - T_{2003}$	Tc_{2013}	0.0505	–0.0327	0.0816	0.0254
		(0.0439)	(0.0513)	(0.0679)	(0.0779)
N		761	777	377	375

Source: SOEP waves 2003, 2008 and 2013.
*/**/*** indicate significance at the 0.1/0.05/0.01 level.
Sample is composed by fathers and mothers for whom trust is observed in
all three waves and children for whom trust is observed in wave 2013.

Table B4. Intergenerational transmission - fathers below median age

	No controls			With controls		
	OLS	**IV**	**IV R.E.**	**OLS**	**IV**	**IV R.E.**
Father's trust	0.1053***	0.0093	–0.0027	0.0990**	–0.0017	–0.0204
	(0.0401)	(0.1155)	(0.1197)	(0.0409)	(0.1185)	(0.1251)
Mother's trust	0.1116***	0.4715***	0.5045***	0.0859**	0.4562***	0.4978***
	(0.0409)	(0.1225)	(0.1300)	(0.0428)	(0.1241)	(0.1315)

Source: SOEP waves 2003, 2008 and 2013.
*/**/*** indicate significance at the 0.1/0.05/0.01 level.
Sample is composed by fathers and mothers for whom trust is observed in all three waves The sample
is further reduced to only fathers with age below the median and children for whom trust is observed
in wave 2013.
OLS: Ordinary least squares; IV: Instrumental variable; IV R.E.: Instrumental variable with random
effects. In the IV models, observable trust of fathers and mothers in 2013 is instrumented by their
observable trust in 2008 and 2003.

Table B5. Intergenerational transmission - by gender

	No controls			*With controls*		
	Males					
	OLS	**IV**	**IV R.E.**	**OLS**	**IV**	**IV R.E.**
Father's trust	0.1350***	0.1203	0.1269	0.1182***	0.0935	0.0935
	(0.0365)	(0.1242)	(0.1266)	(0.0388)	(0.1301)	(0.1328)
Mother's trust	0.0963**	0.3034**	0.3009**	0.0856**	0.2660**	0.2660**
	(0.0407)	(0.1306)	(0.1326)	(0.0431)	(0.1297)	(0.1325)
	Females					
	OLS	**IV**	**IV R.E.**	**OLS**	**IV**	**IV R.E.**
Father's trust	0.0761**	0.0989	0.0637	0.0414	0.0684	0.0520
	(0.0386)	(0.0906)	(0.0936)	(0.0371)	(0.0897)	(0.0936)
Mother's trust	0.1536***	0.3994***	0.4494***	0.1168***	0.3629***	0.3955***
	(0.0383)	(0.1026)	(0.1054)	(0.0383)	(0.1024)	(0.1063)

Source: SOEP waves 2003, 2008 and 2013.

*/**/*** indicate significance at the 0.1/0.05/0.01 level.

Sample is composed by fathers and mothers for whom trust is observed in all three waves and children for whom trust is observed in wave 2013.

OLS: Ordinary least squares; IV: Instrumental variable; IV R.E.: Instrumental variable with random effects. In the IV models, observable trust of fathers and mothers in 2013 is instrumented by their observable trust in 2008 and 2003.

Table B6. Correlation between transient shocks - siblings

Dep. Variable	Main regressor	*No controls*		*With controls*	
		Sibling 1	**Sibling 2**	**Sibling 1**	**Sibling 2**
$T_{2013} - T_{2003}$	Tc_{2013}	–0.1142	–0.0503	–0.0902	0.0063
		(0.1011)	(0.1086)	(0.1374)	(0.1341)
$T_{2013} - T_{2008}$	Tc_{2013}	0.0616	0.1082	0.0880	0.1576
		(0.0944)	(0.0979)	(0.1084)	(0.1163)
$T_{2008} - T_{2003}$	Tc_{2013}	–0.1758	–0.1584	–0.1781	–0.1513
		(0.1070)	(0.1131)	(0.1416)	(0.1308)
N		161	161	161	161

Source: SOEP waves 2003, 2008 and 2013.

*/**/*** indicate significance at the 0.1/0.05/0.01 level.

Sample is composed by N pairs (324 individuals) of children observed in families where there are at least two siblings in 2013 and for whom trust is observed. Pairs are formed by the two youngest siblings. Sibling 1 (2) indicates that the dependent variable refers to the youngest (second youngest) sibling and the main regressor refers to the second youngest (youngest) sibling. Controls include all covariates for both siblings and for the parents. We include only once control variables that are highly collinear between siblings (nationality, number of siblings, place of living at age of 15 and average trust in the region of residence)
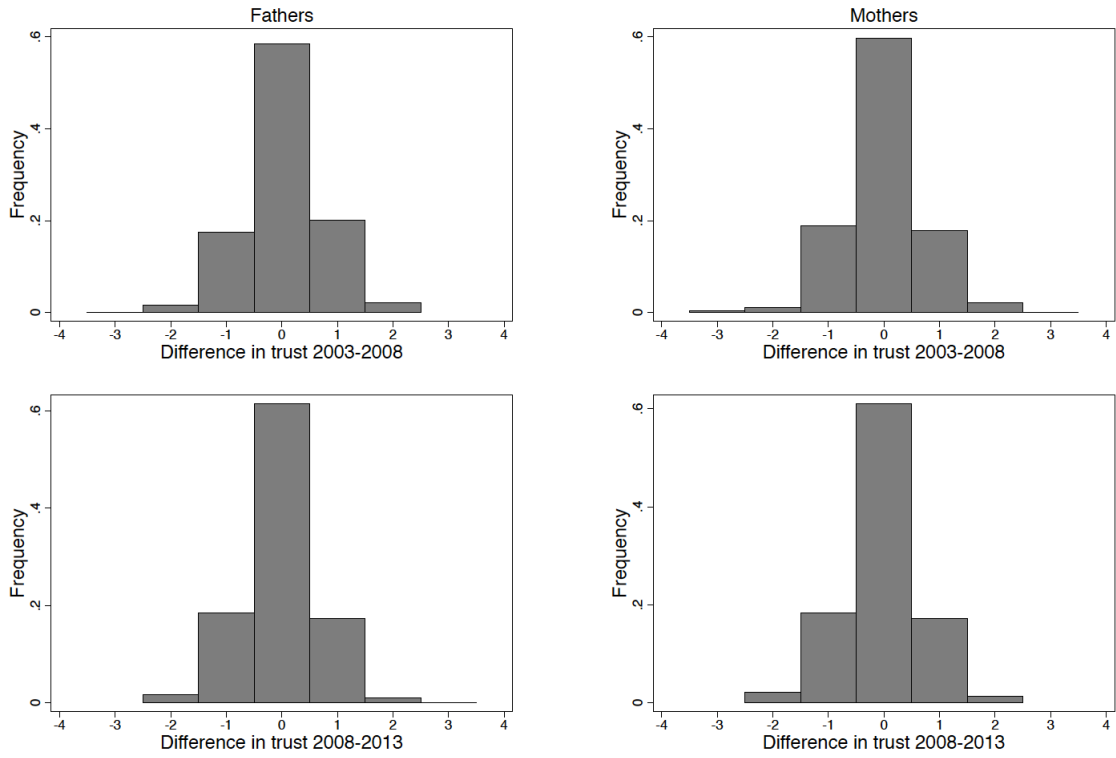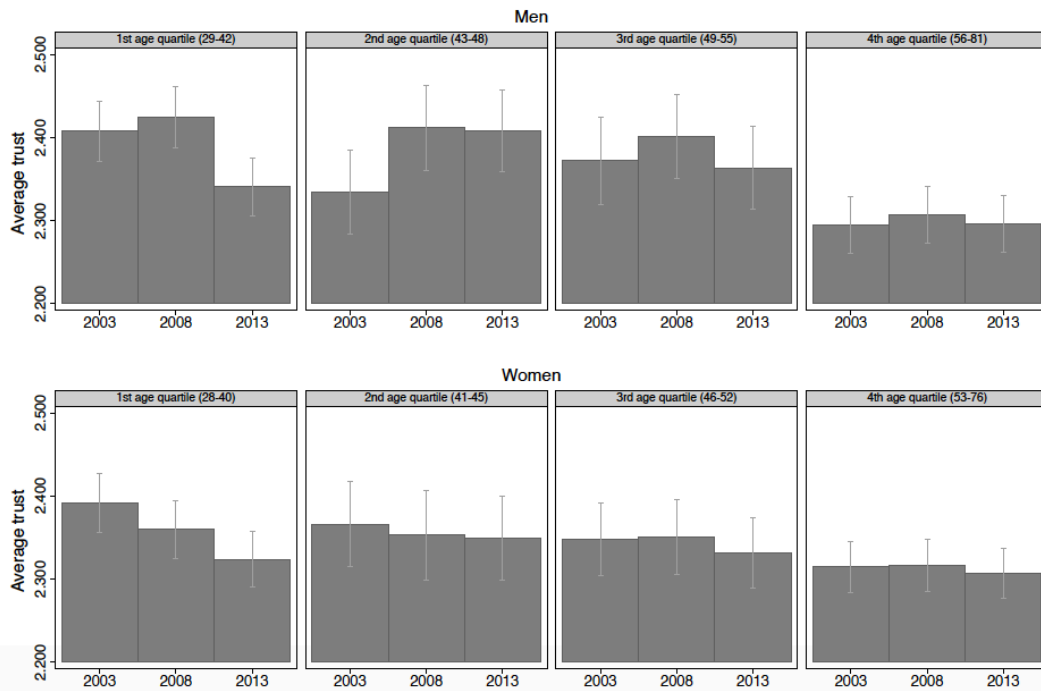
Source: SOEP waves 2003, 2008 and 2013.

Figure B1. Difference in parental trust over time
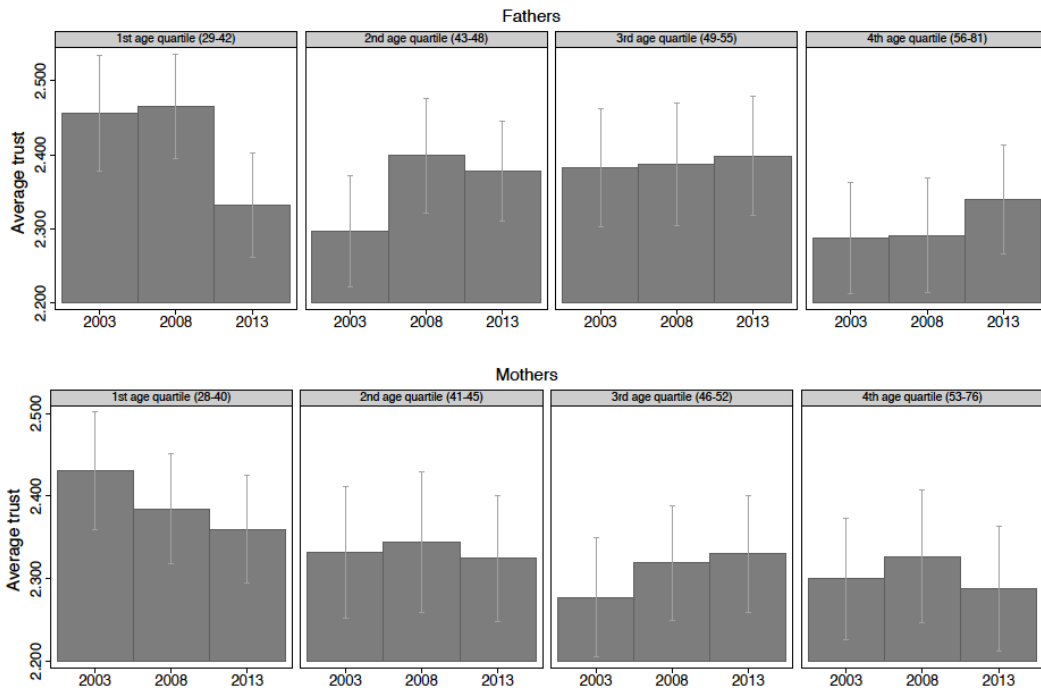
Source: SOEP waves 2003, 2008 and 2013.

Figure B2. Trust over the life cycle

Appendix - Econometrics

B1 Measurement error bias with two explanatory variables

In the transmission equation:

$$Tc_{it}^p = \beta_0 + \beta_1 Tf_{it}^p + \beta_2 Tm_{it}^p + \epsilon_{it} \qquad (4.11)$$

the permanent trust of fathers and mothers are measured with error. Results in Table 4.5 show that using lagged trust as an instrumental variable to correct for the resulting bias makes a major difference for the estimated coefficient of mothers, while it is nearly irrelevant for the estimated coefficient of fathers. This result might seem puzzling in light of the textbook notion that measurement errors on the explanatory variable imply an attenuation bias. To provide an explanation, we make use of an approximation to the OLS bias due to measurement error proposed by Theil (1961), who shows that when there are two regressors both affected by measurement errors the approximate OLS bias is:

$$bias(\beta_1) = -\frac{\beta_1 \lambda_1}{1-\rho^2} + \frac{\beta_2 \lambda_2 \rho}{1-\rho^2} \qquad (4.12)$$

$$bias(\beta_2) = -\frac{\beta_2 \lambda_2}{1-\rho^2} + \frac{\beta_1 \lambda_1 \rho}{1-\rho^2} \qquad (4.13)$$

where ρ is the correlation coefficient between the *true* regressors and $\lambda_j, j \in \{1, 2\}$ is the ratio of the variance of the measurement error to the variance of the respective observable regressor (i.e., the sum of the variances of the measurement error and of the true regressor). If ρ were equal to zero, the bias would collapse to the standard attenuation bias. In this instance, the correlation between the two explanatory

variables is large (ρ is 0.647 for the model without controls and 0.571 for the model with controls), hence the second component on the right-hand side of the equation has a positive sign, counterbalancing the standard attenuation bias, since both β_1 and β_2 are positive in our case. Deriving the values of λ_j and β_j from Tables 4.4 and 4.5 and plugging them in the equations (4.12) and (4.13), we obtain a bias for the coefficient of fathers of 0.173 and of 0.139 in the models without and with controls, while the biases for the coefficients of mothers are -0.351 and -0.313, respectively. This is in line with the difference we observe between the OLS and the IV estimates in Table 4.5, also taking into account sampling variability.

B2 Discreteness of observable trust as additional measurement error

To fix ideas, let us focus on the trust of fathers. We develop our analysis as if the observable trust takes values on a continuous scale. In reality, however, responses to the trust question are categorical, ranging between 1 to 4. A way to rationalize the problem is to think about the observable trust as a discretized version of a latent continuous score. That is, the respondent thinks about trust on a continuous scale, but the way the question is asked induces to round the score to the nearest integer in the range 1 to 4. In the following, we show that under mild conditions, the rounding generates an additional layer of measurement error. The straightforward implication is that the results of our analysis are unaffected, provided that the measurement error vf is redefined to include both the transient shock *and* the error due to rounding.

The key point to the validity of our analysis is that the measurement error must be uncorrelated to the permanent trust. Let Tf' be the unobserved continuous trust such that $Tf = rnd(Tf')$, i.e., the observable trust Tf is the rounded version of Tf'. Let vf' be the transient shock such that $Tf' = Tf^p + vf'$. The issue is then

to derive sufficient conditions for having:

$$cov(rnd(Tf') - Tf^p, Tf^p) = cov(rnd(Tf^p + vf') - Tf^p, Tf^p) = 0. \quad (4.14)$$

Note that the last condition is equivalent to:

$$\frac{cov(rnd(Tf^p + vf'), Tf^p)}{var(Tf^p)} = 1, \quad (4.15)$$

i.e., the regression coefficient of $rnd(Tf^p + vf')$ on Tf^p should be equal to 1. For this to happen it is sufficient to prove that:

$$E(rnd(Tf^p + vf')|Tf^p) = Tf^p. \quad (4.16)$$

If the transient shock vf' is *symmetrically* distributed, the result follows straightforwardly since the discrete probability distribution resulting from rounding is symmetrically distributed around Tf^p. If vf' is not symmetrically distributed the result holds approximately. We performed some simulations using an heavily asymmetric distribution (a $\chi^2(2)$) obtaining essentially the same result as in the symmetric case.

Bibliography

Abras, A., A. Hoyos, A. Narayan, and S. Tiwari (2013). Inequality of opportunities in the labor market: evidence from life in transition surveys in europe and central asia. *IZA Journal of Labor & Development 2*(1), 7.

Acemoglu, D. and D. Autor (2011). *Skills, tasks and technologies: Implications for employment and earnings*, Volume 4. Elsevier Inc.

Akerlof, G. A. (1997). Social distance and social decisions. *Econometrica 65*(5), 1005–1027.

Alesina, A., R. Baqir, and W. Easterly (1997). Public goods and ethnic divisions. *The Quarterly Journal of Economics 114*(4), 1243–1284.

Alesina, A., A. Devleeschauwer, S. Kurlat, and R. Wacziarg (2003). Fractionalization. *Journal of Economic Growth 8*, 155–194.

Alesina, A., C. Gennaioli, and S. Lovo (2016). Public goods and ethnic diversity: evidence from deforestation in Indonesia. Technical report, mimeo.

Alesina, A. and P. Giuliano (2015). Culture and institutions. *Journal of Economic Literature 53*(4), 898–944.

Alesina, A., J. Harnoss, and H. Rapoport (2015). Birthplace diversity and economic prosperity. Technical report, mimeo.

Alesina, A. and E. La Ferrara (2000). Participation in heterogeneous communities. *The quarterly journal of economics 115*(3), 847–904.

Alesina, A. and E. La Ferrara (2002). Who trusts others? *Journal of public economics 85*(2), 207–234.

Alesina, A. and E. La Ferrara (2005). Ethnic diversity and economic performance. *Journal of Economic Literature XLIII*(September), 762–800.

Alesina, A., S. Michalopoulos, and E. Papaioannou (2016). Ethnic inequality. *Journal of Political Economy 124*(2), 428–488.

Algan, Y. and P. Cahuc (2013). Trust, growth, and well-being: New evidence and policy implications. In P. Aghion and S. Durlauf (Eds.), *Handbook of Economic Growth*, Volume 2A. North Holland, Elsevier.

Algan, Y., C. Hemet, and D. D. Laitin (2016). The social effects of ethnic diversity at the local level: a natural experiment with exogenous residential allocation. *Journal of Political Economy 124*(3), 696–732.

Amodio, F. and G. Chiovelli (2017). Ethnicity and violence during democratic transitions: evidence from South Africa. *Journal of the European Economic Association*.

Anand, R., S. Kothari, and N. Kumar (2016). South Africa: labor market dynamics and inequality.

Andersson, R., J. M. Quigley, and M. Wilhelmsson (2005). Agglomeration and the spatial distribution of creativity. *Papers in Regional Science 84*(3), 445–464.

Arrow, K. J. (1972). Gift and exchanges. *Philosophy & Public Affairs 1*(4), 343–362.

Autor, D. et al. (2010). The polarization of job opportunities in the us labor market: Implications for employment and earnings. Technical report, Center for American Progress and The Hamilton Project.

Badescu, M., B. D'Hombres, and E. Villalba (2011). Returns to education in european countries. evidence from the european community statistics on income and living conditions (eu-silc).

Banerjee, A., S. Galiani, J. Levinsohn, Z. McLaren, and I. Woolard (2008). Why has unemployment risen in the New South Africa? *Economics of Transition 16*(4), 715–740.

Banfield, E. C. (1958). *The Moral Basis of a Backward Society.* Free Press; Research Center in Economic Development and Cultural Change [University of Chicago].

Barone, G. and S. Mocetti (2016). Intergenerational mobility in the very long run: Florence 1427-2011. *Temi di Discussione Banca d'Italia 1060.*

Barro, R. J. and J. W. Lee (2013). A new data set of educational attainment in the world, 1950-2010. *Journal of Development Economics 104*, 184–198.

Bartolj, T., A. Ahčan, A. Feldin, and S. Polanec (2013). Evolution of private returns to tertiary education during transition: evidence from slovenia. *Post-Communist Economies 25*(3), 407–424.

Beine, M., F. Docquier, and M. Schiff (2013). International migration, transfer of norms and home country fertility. *Canadian Journal of Economics/Revue canadienne d'économique 46*(4), 1406–1430.

Bender, K. A., S. M. Donohue, and J. S. Heywood (2005). Job satisfaction and gender segregation. *Oxford economic papers 57*(3), 479–496.

Bhorat, H. and M. Oosthuizen (2005). The post-Apartheid South African labour market.

Bingley, P. and L. Cappellari (2012). Alike in many ways: Intergenerational and sibling correlations of brothers' earnings. *IZA Discussion Paper 6987.*

Bisin, A., G. Topa, and T. Verdier (2009). Cultural transmission, socialization and the population dynamics of multiple-trait distributions. *International Journal of Economic Theory 5*(1), 139–154.

Bisin, A. and T. Verdier (2000). "Beyond the melting pot": Cultural transmission, marriage, and the evolution of ethnic and religious traits. *Quarterly Journal of Economics 115*(3), 955–988.

Bisin, A. and T. Verdier (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory 97*(2), 298–319.

Bisin, A. and T. Verdier (2005). Cultural transmission. *The New Palgrave Dictionary of Economics.*

Bisin, A. and T. Verdier (2011). The economics of cultural transmission and socialization. In M. J. Benhabib J., A. Bisin (Ed.), *Handbook of Social Economics*, pp. 339–416. North Holland, Elsevier.

Bloom, N., R. Sadun, and J. Van Reenen (2012). The organization of firms across countries. *The Quarterly Journal of Economics 127*(4), 1663–1705.

Blundell, R., L. Dearden, C. Meghir, and B. Sianesi (1999). Human capital investment: the returns from education and training to the individual, the firm and the economy. *Fiscal studies 20*(1), 1–23.

Borjas, G. J. and S. G. Bronars (1989). Consumer discrimination and self-employment. *Journal of Political Economy 97*(3), 581–605.

Bourguignon, F., F. H. Ferreira, and M. Menendez (2007). Inequality of opportunity in brazil. *Review of Income and Wealth 53*(4), 585–618.

Boyd, R. and P. J. Richerson (1988). *Culture and the evolutionary process.* University of Chicago Press.

Brock, J., R. Greenberg, B. Rambousek, and C. Wolff (2013). Inequality of opportunity at the household level. In E. T. Report (Ed.), *Economic Inclusion in Transition*. European Bank for Reconstruction and Development.

Brunello, G., E. Crivellaro, and L. Rocco (2010). Lost in transition? the returns to education acquired under communism 15 years after the fall of the berlin wall. Discussion Paper 5409, IZA.

Brunori, P., F. H. Ferreira, and V. Peragine (2013). Inequality of opportunity, income inequality, and economic mobility: Some international comparisons. In *Getting Development Right*, pp. 85–115. Springer.

Butler, J., P. Giuliano, and L. Guiso (Forthcoming 2016). The right amount of trust. *Journal of European Economic Association*.

Campante, F. R. and D. Chor (2012). Schooling, political participation, and the economy. *Review of Economics and Statistics 94*(4), 841–859.

Cavalli-Sforza, L. L. and M. W. Feldman (1981). *Cultural transmission and evolution: A quantitative approach*. Princeton University Press.

Chay, K. and K. Munshi (2015). Black networks after emancipation: evidence from reconstruction and the Great Migration.

Checchi, D., V. Peragine, and L. Serlenga (2010). Fair and unfair income inequalities in europe. Discussion Paper 5025, IZA.

Checchi, D., V. Peragine, and L. Serlenga (2015). Income inequality and opportunity inequality in europe: Recent trends and explaining factors. In *Proceedings of the 6th ECINEQ Meeting*. University of Luxembourg.

Christopher, A. J. (2001). *The atlas of changing South Africa*. London and New York: Routledge.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, S95–S120.

Coleman, J. S. (1990). *Foundations of Social Theory.* Belknap Press of Harvard University Press Cambridge, Mass.

Cutler, D. and A. Lleras-Muney (2008). Education and health: Evaluating theories and evidence. In *Making Americans Healthier: Social and Economic Policy as HealthPolicy*, New York. Russell Sage Foundation.

Dahlberg, M., K. Edmark, and H. Lundqvist (2012). Ethnic diversity and preferences for redistribution. *Journal of Political Economy 120*(1), 41–76.

Deming, D. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, 1593–1640.

Desmet, K., I. Ortuño-Ortín, and R. Wacziarg (2012). The political economy of linguistic cleavages. *Journal of development Economics 97*(2), 322–338.

Desmet, K., I. Ortuno-Ortin, and R. Wacziarg (2017). Culture, ethnicity, and diversity. *American Economic Review 107*(9), 2479–2513.

Dohmen, T. J., A. Falk, D. Huffman, and U. Sunde (2012). The intergenerational transmission of risk and trust attitudes. *Review of Economic Studies 79*(2), 645–677.

Dumont, J.-C., G. Spielvogel, and S. Widmaier (2010). International migrants in developed, emerging and developing countries. *OECD Social, Employment and Migration Working Paper No.114*.

Dustmann, C., U. Schoenberg, and J. Stuhler (2017). Labor Supply Shocks, Native Wages, and the Adjustment of Local Employment. *The Quarterly Journal of Economics 132*(1), 435–483.

Easterbrook, M. J., T. Kuppens, and A. S. Manstead (2016). The education effect: Higher educational qualifications are robustly associated with beneficial personal and socio-political outcomes. *Social Indicators Research 126*(3), 1261–1298.

Easterly, W. and R. Levine (1997). Africa's growth tragedy: policies and ethnic divisions. *The Quarterly Journal of Economics 112*(4), 1203–1250.

Eraydin, A., T. Tasan-Kok, and J. Vranken (2010). Diversity matters: immigrant entrepreneurship and contribution of different forms of social integration in economic performance of cities. *European Planning Studies 18*(4), 521–543.

Fainstein, S. S. (2005). Cities and diversity: should we want it? can we plan for it? *Urban affairs review 41*(1), 3–19.

Fernández, R. (2008). Culture and economics. Volume 2, pp. 333–40. Palgrave Macmillan Hampshire.

Fernández, R. and A. Fogli (2009). Culture: An empirical investigation of beliefs, work, and fertility. *American Economic Journal: Macroeconomics 1*(1), 146–177.

Ferreira, F. H. and J. Gignoux (2011). The measurement of inequality of opportunity: Theory and an application to latin america. *Review of Income and Wealth 57*(4), 622–657.

Ferreira, F. H., C. Lakner, M. A. Lugo, and B. Ozler (2014). Inequality of opportunity and economic growth: A cross-country analysis. Policy Research Working Paper Series 6915, The World Bank.

Ferreira, F. H. and V. Peragine (2016). Individual responsibility and equality of opportunity. In M. Adler and M. Fleurbaey (Eds.), *The Oxford Handbook of Well Being and Public Policy.* Oxford University Press.

Flemming, J. S. and J. Micklewright (2000). Income distribution, economic systems

and transition. In A. Atkinson and F. Bourguignon (Eds.), *Handbook of Income Distribution*, Volume 1, pp. 843–918. Amsterdam: Elsevier.

for Reconstruction, E. B. and Development (2016). Transition report 2016/17. Technical report, London.

Fukuyama, F. (1995). *Trust: The Social Virtues and the Creation of Prosperity.* London : Hamish Hamilton.

Glaeser, E. L., H. D. Kallal, J. A. Scheinkman, and A. Shleifer (1992). Growth in cities. *Journal of Political Economy 100*(6), 1126–1152.

Glaeser, E. L., J. Scheinkman, and A. Shleifer (1995). Economic growth in a cross-section of cities. *Journal of monetary economics 36*(1), 117–143.

Goldberg, M. S. (1982). Discrimination, nepotism, and long-run wage differentials. *The Quarterly Journal of Economics*, 307–319.

Goldin, C. and L. F. Katz (2009). *The race between education and technology.* Harvard University Press.

Gradin, C. (2014). Poverty and ethnicity among black South Africans.

Güell, M., J. V. R. Mora, and C. I. Telmer (2015). The informational content of surnames, the evolution of intergenerational mobility, and assortative mating. *The Review of Economic Studies 82*(2), 693–735.

Guiso, L., P. Sapienza, and L. Zingales (2004). The role of social capital in financial development. *American Economic Review 94*(3), 526–556.

Guiso, L., P. Sapienza, and L. Zingales (2006). Does culture affect economic outcomes? *Journal of Economic Perspectives 20*(2), 23–48.

Guiso, L., P. Sapienza, and L. Zingales (2008a). Long term persistence. *National Bureau of Economic Research Working Paper Series 14278.*

Guiso, L., P. Sapienza, and L. Zingales (2008b). Social capital as good culture. *Journal of the European Economic Association 6*(2-3), 295–320.

Guiso, L., P. Sapienza, and L. Zingales (2008c). Trusting the stock market. *Journal of Finance 63*(6), 2557–2600.

Guiso, L., P. Sapienza, and L. Zingales (2009). Cultural biases in economic exchange? *The Quarterly Journal of Economics 124*(3), 1095–1131.

Guriev, S. and E. Zhuravskaya (2009). (un) happiness in transition. *The Journal of Economic Perspectives 23*(2), 143–168.

Guy, J. and M. Thabane (1988). Technology, ethnicity and ideology: basotho miners and shaft-sinking on the South African gold mines. *Journal of Southern African Studies 14*(2), 257–278.

Haisken-DeNew, J. P. and M. Hahn (2010). Panelwhiz: Efficient data extraction of complex panel data sets? An example using the German SOEP. *Journal of Applied Social Science Studies 130*(4), 643–654.

Heintz, J. and D. Posel (2007). Revisiting informal employment and segmentation in the South African labor market.

Hjort, J. (2014). Ethnic divisions and production in firms. *The Quarterly Journal of Economics*, 1899–1946.

Hufe, P., A. Peichl, J. Roemer, and M. Ungerer. Inequality of income acquisition: the role of childhood circumstances. *Social Choice and Welfare*, 1–46.

Ivaschenko, O. (2002). Growth and inequality: evidence from transitional economies. CESifo Working Paper 746, CESifo, Munich.

Iyer, G. R. and J. M. Shapiro (1999). Ethnic entrepreneurial and marketing systems: Implications for the global economy. *Journal of International Marketing*, 83–110.

Kim, S. (1989). Labor specialization and the extent of the market. *Journal of Political Economy 97*(3), 692–705.

Kingdon, G. G. and J. Knight (2004). Unemployment in South Africa: the nature of the beast. *World Development 32*(3), 391–408.

Knack, S. and P. Keefer (1997). Does social capital have an economic payoff? A cross-country investigation. *The Quarterly Journal of Economics 112*(4), 1251–1288.

La Porta, R., F. Lopez-de Silanes, A. Shleifer, and R. Vishny (1997). Legal determinants of external finance. *Journal of Finance 52*(3), 1131–1150.

Lazear, E. P. (1999a). Culture and language. *Journal of Political Economy 107*(6), 95–124.

Lazear, E. P. (1999b). Globalisation and the market for team-mates. *Economic Journal 109*, 15–40.

Lefranc, A., N. Pistolesi, and A. Trannoy (2008). Inequality of opportunities vs. inequality of outcomes: Are western societies all alike? *Review of Income and Wealth 54*(4), 513–546.

Lefranc, A., N. Pistolesi, and A. Trannoy (2009). Equality of opportunity and luck: Definitions and testable conditions, with an application to income in france. *Journal of Public Economics 93*(11), 1189–1207.

Leibbrandt, M., I. Woolard, H. McEwen, and C. Koep (2009). Employment and inequality outcomes in South Africa.

Lewis, M. P., G. F. Simons, and C. D. Fennig (2009). *Ethnologue: languages of the world*, Volume 16. SIL international Dallas, TX.

Ljunge, M. (2014). Trust issues: Evidence on the intergenerational trust transmission among children of immigrants. *Journal of Economic Behavior & Organization 106*, 175–196.

Luongo, P. (2011). The implication of partial observability of circumstances on the measurement of inequality of opportunity. In *Inequality of Opportunity: Theory and Measurement*, pp. 23–49. Emerald Group Publishing Limited.

Magruder, J. R. (2010). Intergenerational networks, unemployment, and persistent inequality in South Africa. *American Economic Journal: Applied Economics 2*(1), 62–85.

Mariotti, M. (2012). Labour markets during apartheid in south africa. *The Economic History Review 65*(3), 1100–1122.

Marrero, G. A. and J. G. Rodríguez (2012). Inequality of opportunity in europe. *Review of Income and Wealth 58*(4), 597–621.

Marrero, G. A. and J. G. Rodríguez (2013). Inequality of opportunity and growth. *Journal of Development Economics 104*, 107–122.

Mayda, A. M. (2010). International migration: a panel data analysis of the determinants of bilateral flows. *Journal of Population Economics 23*(4), 1249–1274.

Michalopoulos, S. (2012). The origins of ethnolinguistic diversity. *American Economic Review 102*(4), 1508–1539.

Michalopoulos, S. and E. Papaioannou (2013). Pre-colonial ethnic institutions and contemporary african development. *Econometrica 81*(1), 113–152.

Miguel, E. and M. K. Gugerty (2005). Ethnic diversity, social sanctions, and public goods in Kenya. *Journal of Public Economics 89*, 2325–2368.

Milanovic, B. (1998). Income, inequality, and poverty during the transition from planned to market economy. The world bank regional and sectoral studies series, The World Bank, New York.

Milanovic, B. (1999). Explaining the increase in inequality during transition. *Economics of transition 7*(2), 299–341.

Mitra, P. and R. Yemtsov (2007). Increasing inequality in transition economies: is there more to come? In F. Bourguignon and B. Pleskovic (Eds.), *Beyond transition*, Washington, pp. 59–102. The World Bank.

Montalvo, J. G. and M. Reynal-Querol (2005). Ethnic polarization, potential conflict, and civil wars. *The American Economic Review 95*(3), 796–816.

Morgan, J. and F. Vardy (2009). Diversity in the workplace. *American Economic Review 99*(1), 472–485.

Moschion, J. and D. Tabasso (2014). Trust of second-generation immigrants: Intergenerational transmission or cultural assimilation? *IZA Journal of Migration 3*(1), 1–30.

Munshi, K. (2003). Networks in the modern economy: Mexican migrants in the U.S. labor market. *The Quarterly Journal of Economics 118*(2), 549–599.

Munshi, K. (2011). Strength in numbers: networks as a solution to occupational traps. *Review of Economic Studies 78*(3), 1069–1101.

Munshi, K. and M. Rosenzweig (2006). Traditional institutions meet the modern world: caste, gender and schooling choices in a globalizing economy. *American Economic Review 96*(4), 1225–1252.

Murphy, K. M. and R. H. Topel (2016). Human capital investment, inequality, and economic growth. *Journal of Labor Economics 34*(S2), S99–S127.

Mwakikagile, G. (2010). *South Africa as a multi-ethnic society.* Dar es Salaam, Tanzania: Continental Press.

Neumann, G. and R. Topel (1991). Employment risk, diversification, and unemployment. *The Quarterly Journal of Economics*, 1341–1365.

Neyer, G. R. (2006). Family policies and fertility in europe: Fertility policies at the intersection of gender policies, employment policies and care policies. Technical Report No. WP-2006-010, Max Planck Institute for Demographic Research - Rostock, Germany.

Niebuhr, A. (2010). Migration and innovation: Does cultural diversity matter for regional r&d activity? *Papers in Regional Science 89*(3), 563–585.

Niehues, J. and A. Peichl (2014). Upper bounds of inequality of opportunity: theory and evidence for germany and the us. *Social Choice and Welfare 43*(1), 73–99.

Nikolova, E., D. Simroth, et al. (2013). Does cultural diversity help or hinder entrepreneurs? evidence from eastern europe and central asia. Technical report, European Bank for Reconstruction and Development.

Nunn, N. and D. Puga (2012). Ruggedness: The blessing of bad geography in africa. *Review of Economics and Statistics 94*(1), 20–36.

Ortega, F. and G. Peri (2014). Openness and income: The roles of trade and migration. *Journal of International Economics 92*(2), 231–251.

Ottaviano, G. I. and G. Peri (2006). The economic value of cultural diversity: evidence from us cities. *Journal of Economic geography 6*(1), 9–44.

Pasquier-Doumer, L. (2012). Intergenerational transmission of self-employed status in the informal sector: a constrained choice or better income prospects? evidence from seven west african countries. *Journal of african economies 22*(1), 73–111.

Perugini, C. and F. Pompei (2015). *Inequalities During and After Transition in Central and Eastern Europe.* London, UK: Palgrave-MacMillan.

Posel, D. (2001). What's in a name? racial categorisations under apartheid and their afterlife. *Transformation 47*, 50–74.

Poulin, M. J. and C. M. Haase (2015). Growing to trust evidence that trust increases and sustains well-being across the life span. *Social Psychological and Personality Science 6*(6), 614–621.

Prskawetz, A., B. Mahlberg, and V. Skirbekk (2007). Firm productivity, workforce age and educational structure in austrian industries in 2001. *RL Clark, N. Ogawa und A. Mason (Hg.): Population Ageing, Intergenerational Transfers and the Macroeconomy. Northampton, MA: Edward Elgar*, 38–66.

Putnam, R. D. (1993). *Making democracy work: Civic traditions in modern Italy.* Princeton University Press.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community.* Simon & Schuster.

Putnam, R. D. (2007). E pluribus unum: diversity and community in the twenty-first century the 2006 johan skytte prize lecture. *Scandinavian political studies 30*(2), 137–174.

Ramos, X. et al. (2016). Approaches to inequality of opportunity: Principles, measures and evidence. *Journal of Economic Surveys 30*(5), 855–883.

Roemer, J. E. (1993). A pragmatic theory of responsibility for the egalitarian planner. *Philosophy & Public Affairs*, 146–166.

Roemer, J. E. (1998). *Theories of distributive justice.* Harvard University Press.

Roemer, J. E. and A. Trannoy (2015). Equality of opportunity. In *Handbook of Income Distribution*, Volume 2, Chapter 4, pp. 217–300.

Rose, S. and C. Viju (2014). Income inequality in post-communist central and eastern european countries. *Eastern Journal of European Studies 5*(1).

Sacerdote, B. (2011). Nature and nurture effects on children's outcomes: What have we learned from studies of twins and adoptees? In M. J. Benhabib J., A. Bisin (Ed.), *Handbook of Social Economics*, pp. 1–30. North Holland, Elsevier.

Santos Silva, J. and S. Tenreyro (2006). The log of gravity. *The Review of Economics and statistics 88*(4), 641–658.

Sastre, M. and A. Trannoy (2002). Shapley inequality decomposition by factor components: Some methodological issues. *Journal of Economics 9*, 51–89.

Shorrocks, A. F. (1982). Inequality decomposition by factor components. *Econometrica: Journal of the Econometric Society*, 193–211.

Shorrocks, A. F. (2013). Decomposition procedures for distributional analysis: a unified framework based on the shapley value. *Journal of Economic Inequality*, 1–28.

Sobel, R. S., N. Dutta, and S. Roy (2010). Does cultural diversity increase the rate of entrepreneurship? *The Review of Austrian Economics 23*(3), 269–286.

Solon, G., M. Corcoran, R. Gordon, and D. Laren (1991). A longitudinal analysis of sibling correlations in economic status. *The Journal of Human Resources 26*(3), 509–534.

Solt, F. (2016). The standardized world income inequality database. *Social Science Quarterly 97*(5), 1267–1281.

Sørensen, J. B. (2007). Closure and exposure: Mechanisms in the intergenerational transmission of self-employment. In *The Sociology of Entrepreneurship*, pp. 83–124. Emerald Group Publishing Limited.

Terama, E., A. Kõu, and K. Samir (2014). Early transition trends and differences of higher education attainment in the former soviet union, central and eastern european countries. *Finnish Yearbook of Population Research 49*, 105.

Theil, H. (1961). *Economic forecasts and policy.* North-Holland, Amsterdam, 2d rev. ed.

Wagner, G. G., J. R. Frick, and J. Schupp (2007). The German Socio-Economic Panel Study (SOEP): Scope, evolution and enhancements. *Schmollers Jahrbuch 127*(1), 139–169.

Wittenberg, M. (2002). Job search in South Africa: a nonparametric analysis. *South African Journal of Economics 70*(8), 1163–1197.