UNIVERSITÉ PARIS 1 – PANTHÉON - SORBONNE –

Documents de Travail du
Centre d'Economie de la Sorbonne

CES Working Papers

Who benefits the most from peer effects within ethnic group ? Empirical evidence on the South African Labour Market

Gaëlle FERRANT, Yannick BOURQUIN

2011.44

Maison des Sciences Économiques, 106-112 boulevard de L'Hôpital, 75647 Paris Cedex 13
http://centredeconomiesorbonne.univ-paris1.fr/bandeau-haut/documents-de-travail/
ISSN : 1955-611X

Who benefits the most from peer effects within ethnic group? Empirical evidence on the South African Labour Market. *

Gaëlle Ferrant† and Yannick Bourquin‡

July 2011

*We thank Bernard Fortin, Catherine Sofer, Robert Gary-Bobo, Jean-Louis Arcand, Andrew Clark and Victor Lavy for their time and very useful advice. We also thank participants at the ABCDE World Bank conference, DIAL and EUDN (2011) Workshops, the ESPE, JMA and CSAE conferences and the SIMA seminar for comments.

†Centre d'Economie de la Sorbonne-Université Paris I Panthéon Sorbonne, 106-112 Bd de l'Hôpital, 75647 Paris Cedex 13, France (email: gaelle.ferrant@univ-paris1.fr).

‡Centre d'Economie de la Sorbonne-Université Paris I Panthéon Sorbonne, 106-112 Bd de l'Hôpital, 75647 Paris Cedex 13, France (email:yannick.bourquin@malix.univ-paris1.fr).

Abstract

This paper provides evidence that local social interactions within ethnic groups may explain the puzzling variations in labour-market outcomes across individuals. Peer effects work first by creating pressure on labor-market participation, second, by conveying information about job opportunities and by raising wages. These effects differ through a selection effect: gender and ethnic groups who are less integrated in the labour market benefit more from peer effect. Finally, networks exhibit decreasing returns. The problems of endogeneity and simultaneity of local peer effects are addressed by using (i) data aggregated at the province level, (ii) the distribution of the sex of the peers' siblings as an instrumental variable and (iii) a quasi-panel data approach relying on the Hausman-Taylor estimator. The importance of social interactions in the labour market suggests that a social multiplier exists and our estimates show that any labour-market shock is magnified with an elasticity of 0.5.

JEL classification: J15, J16, O18, Z13

Keywords: Peer effects, Development Economics, Labour, South Africa

Résumé

Les interactions sociales locales au sein des groupes ethniques peuvent nous aider à comprendre les variations inexpliquées des comportements individuels sur le marché du travail. Les effets des pairs se présentent de différentes manières : tout d'abord comme une norme sociale en termes de participation. Deuxièmement, les effets de pairs agissent comme un vecteur d'information améliorant l'accès à l'emploi et augmentant les salaires. Ces effets diffèrent selon le genre et l'origine ethnique des individus à travers un effet de sélection : les groupes ethniques les moins intégrées économiquement bénéficient plus des effets de pairs. Enfin les rendements sont décroissants. Les problèmes d'endogénéité et de simultanéité des effets de pairs locaux sont traités en utilisant (i) les données agrégées au niveau de la province, (ii) la répartition du sexe des enfants des pairs comme instrument et (iii) une approche en quasi-panel fondée sur l'estimateur d'Hausman-Taylor. La présence d'interactions sociales sur le marché du travail implique l'existence d'un multiplicateur social : tout choc sur le marché du travail est amplifié avec une élasticité de 0.5.

Mots clés : Effets de pairs, Economie du développement, Marché du travail, Afrique du Sud

1 Introduction

Although South Africa is an upper middle-income country, social indicators suggest that living standards there are closer to those in low-income countries. The country's long history of segregation and discrimination is often appealed to as an explanation for this gap between economic status and social development (Bhorat *et al.* (2004)). In post-apartheid South Africa, the ethnic group to which individuals belong plays a key role in labour-market behavior and outcomes (Keswell & Poswell (2002), Kingdon & Knight (2004) and Cornwell & Inder (2008)).

In this context, this paper considers the following question: Do peer effects reduce the gender and ethnic gap by improving the activity, the employment rate and wages of the least integrated group? To deal with this question this paper analyses whether peer behavior (i.e. behavior of neighbors speaking the same language) on the labour market can predict individual labour-force participation, work and wages. Finally, this paper studies the heterogeneity of this effect which is non-linear and group-specific.

Our results show that being surrounded by active peers increases the probability of participation. As the number of active peers increases, social pressure on the inactive provides a greater incentive to participate in the labor market. Furthermore, this effect is convex for men and concave for women: social pressure plays a role for women even when there are relatively few active peers, whereas the analogous effect for males is marginal. Along the same lines, being surrounded by employed peers increases both the probability of employment and wages. This effect is concave for both sexes, suggesting decreasing returns to networks. The wage effect confirms that peer effects do not reflect shifts in the labor-supply curve. The results can be interpreted in the framework of job-search theory, whereby employed peers act as a network relaying information between members. This information is used to improve the odds of finding a job. The wage effect suggests that networks are used as a matching device. A decomposition of these peer effects by ethnic group reveals a selection effect: the best-integrated groups[1] in the labor market, namely Afrikaans- and English-speaking individuals, are those who benefit the least from peer effects.

[1]Best integrated means overepresentation in the labor-force, i.e. the ratio "'% in the employed population/% in the total population"' >1. See last column of Table 2

Those with a locally well-integrated ethnic group do not rely on networks but rather on market mechanisms. This suggests that peer and network effects may help to reduce inequality.

A growing literature has focused on social interactions, networks and peer effects as explanations for the puzzling variations in labor-market outcomes across workers, time periods and areas (See for example Alesina *et al.* (2006)). The questions we address here are important in a number of dimensions of development and economic policy. If peer effects exist and are gender- and ethnic-specific, then policies for development and poverty reduction will be affected. The existing literature on social interactions has found evidence of a social multiplier (Glaeser *et al.* (2003)). This latter describes the snowball effect which amplifies labor-market shocks via social interactions. Individual decisions here depend both on the context (as predicted by standard economic models) and on peers' decisions. Labor-market inequality is a major component of overall inequality between men and women and between ethnic groups in South Africa (Bhorat *et al.* (2004)). In addition, the understanding the effect of labor-market networks is arguably critical in developing countries where market failures produce little information transmission. Especially in South Africa characterized by high levels of unemployment (29% in 2008 World Bank), worker discouragement (18% of inactive Kingdon & Knight (2000)) and relatively low absorption of the unemployed into the informal sector (Kingdon & Knight (2004)). It suggests that individuals do not rely on market mechanisms to find a job. Moreover, job-search methods are predominantly passive, with most jobs being obtained via word-of-mouth and other informal recruitment methods (Kingdon & Knight (2000)). The South African National Income Dynamics Study (2008) provides some information about the way in which individuals found their job, and confirms the importance of networks: over 42% of respondents claim to have found their current job via relatives or friends outside the household (See Appendix A).

In line with the theoretical work in Banerjee (1992) and Bikhchandani *et al.* (1992), we assume that network effects on the labor market operate via an information vector, as well as via the transmission of norms. This may first reflect social pressure: those who are among the few unemployed in their group are stigmatized (Clark (2003)), and thus have additional incentives to look for a job. Moreover, job offers may be obtained from direct and indirect acquaintances through word-of-mouth communication (de Martí & Zenou (2009)). By reducing employer uncertainty about worker productivity, networks may enhance the match quality between job candidates and the available jobs (Marsden & Gorman (2001)) and increase wages. Calvo-Armengol & Jackson (2004) build a simple model of a labor-market with networks. Agents randomly receive information about job opportunities. When they hear about a vacant job, they may either keep the job information for themselves, or pass the information along to other agents to whom they are connected. It has been shown that this suffices to generate a positive correlation between the employment statuses of the agents who are connected in the network. Following Granovetter (1974, 1995), the network is considered as a productive channel for job finding. An alternative explanation relies on expectations. In the traditional job-search literature, the decision to look for work is based on a cost-benefit analysis taking into account the cost of looking for a job, the probability of finding a job and the corresponding wage distribution. Such information is not readily available and can be inferred from the observation of peer behavior and outcomes. If a substantial proportion of the individual's peers work, she might conclude that finding a job will be easier than she had previously thought.[2]

The identification of peer influence on own labor-market decisions raises a number of serious problems, due to endogeneity, simultaneity and reflection (Manski (1993)). Social interactions simultaneously affect both the individual's decisions and those of other agents in the reference group (Sacerdote (2001)). In addition, individuals select into peer groups based, in part, on their unobservable characteristics (Hoxby (2000)). Individuals living in the same neighborhood or belonging to the same ethnic group then tend to take similar decisions partly because they share the same background and preferences. An OLS regression of individual behavior on the mean of the endogenous variable is then unlikely to be informative about causal effects. To avoid this worry, we add dummies for language group and area (Bertrand *et al.* (2000)) and use a non-linear probit model (Araujo *et al.* (2004)). Moreover, we use alternatively three identification strategies. First, the local measure of network is replaced by an aggregated measure at the province level (Bertrand *et al.* (2000));

[2]See Chapter 8 of Jackson (2008) for a survey of learning in networks.

second, the distribution of the sex of peers' siblings is used as instrumental variable (Maurin & Moschion (2009)); last, we adopt a quasi-panel approach (Graham & Hahn (2005)).

The paper proceeds as follows. Section 2 defines the peer group and its measure, and Section 3 briefly presents the data. Section 4 describes the empirical strategy and then Section 5 presents the empirical results. Last, Section 6 discusses the policy implications and concludes.

2 Modeling Peer Effects

2.1 Peer Group Definition

South Africa is an interesting case for study as it has adopted a multilingual language policy in order to ensure ease of communication. In the Constitution the eleven official languages of the Republic are Sepedi, Sesotho, Setswana, siSwati, Tshivenda, Xitsonga, Afrikaans, English, isiNdebele, isiXhosa and isiZulu. However, this constitutional provision may encourage the strongest language, namely English, which will then play a hegemonic role. Language usage tends confirm the practice of English monolingualism: English is the language of cities, commerce and banking, national government, road signs and most official documents. With respect to the labor market, English appears to be the language of communication. The vast majority of commercial and formal labor-market activities are in English, whereas less than 8% of the population claims English as their mother-tongue or their home language (Cornwell & Inder (2008)). One question is then whether this multilingual policy produces a greater impact of labor-market peers for those who speak English or for those who speak some other language.

One of the major concerns in the analysis of network effects is the identification of who belongs to which network. Some datasets do contain direct information, such as *AddHealth* in which we can identify each teenager's best friends. Nevertheless, in general researchers are very often obliged to make assumptions about plausible network membership. The most straightforward proxies for networks are peer groups. Much work appeals to this strategy implicitly by mixing up the terms "peers", "friends", "contacts" and "networks".

We here approximate networks by defining likely peer groups for each individual. These peer groups are defined by area and ethnic membership. Defining peer effects geographically is common in the literature and is justified on the ground that peer effects mainly result from local social interactions. Maurin & Moschion (2009) argue that social interactions in female labor supply pertain at the neighborhood level, and find a considerable impact of close neighbors' participation on individual labor-force participation. A number of pieces of work have found that individual outcomes are correlated with those of the individual's neighbors.[3] These local social interactions via neighbors reflect the quality of information decaying with distance (de Martí & Zenou (2009)).

Fernandez & Fogli (2005) also emphasize the role of neighborhoods, and show that ethnic groups tend to cluster in the same neighborhoods in the USA. They suggest that in neighborhoods with a greater proportion of individuals of the same ethnic group, there is a more pronounced transmission and maintenance of the norms regarding the attitudes that women should have in the labor-market. Polarization and hysteresis are therefore stronger in more segregated neighborhoods. Table 1 presents the maximum concentration rate for each language group, and reveals a considerable correlation between ethnicity and geographical location in South Africa. For each language group, there is one province with at least 31.08% (and at most 76.82%) of language peers.

[Table 1 about here]

An alternative to a geographical peer group is one based on ethnicity (Borjas (1992, 1994, 1995)): empirical analyses along these lines have uncovered correlations between individual and ethnic-group outcomes. In the South African context, Burns *et al.* (2010), Hofmeyr (2010) suggest that social networks defined by ethnic groups matter for labor-market outcomes. Bertrand *et al.* (2000) argue that the language spoken at home is a good proxy for ethnic groups and that it make sense to define peer groups as individuals in the neighborhood who speak the same language. This argument is based on the following elements. First, there is a vast literature on homophily, showing that people tend to form ties with other from the same demo-

[3] See Jencks & Mayer (1990) for a literature review.

graphic group and, amongst other criteria, from the same ethnic group.[4] If peers are mainly drawn from the pool of individuals with the same ethnic origin, we will have a good approximation of potential contacts by focusing on the ethnic group. Second, it is a well-established fact that language is closely related to identity, and especially ethnic identity (see Fishman (2001)), so that the language spoken at home is a strong signal of ethnic membership, and more so than skin color or ethnicity. We consider language spoken at home as an active belonging, while ancestry and race are exogenous and do not suffice to determine ethnic identity (Alba (1990)). Ethnic group linked by ancestry may include individuals who are only loosely connected, and we assume that individuals partly reject their ethnic membership if they do not speak the associated language at home. Moreover, Hofmeyr (2010) argues that differences within race and ethnic groups prevent the sole use of geographical information to identify networks: "*a black South African who speaks IsiZulu does not necessarily understand or associate with a black South African who speaks IsiXhosa even if they live in the same area*". Finally, we expect that information will be more easily transmitted between individuals who speak the same language. We assume that two people speaking the same language at home may exchange information about job opportunities, although we thereby underestimate the true number of potential contacts.

2.2 Network Measure

To measure the range and quality of the network, we follow Bertrand *et al.* (2000). Our network variable is constructed by multiplying measures of the *quality* and *size* of the network. In what follows, i indexes individuals, h households, a areas[5] and l ethnic groups defined by language spoken at home. Network quality is defined by the proportion of peers (excluding the individual and household members) who are currently active or working[6]: $\overline{L}_{ial}$. Let L_{ial} be equal to one if individual i speaking language l and living in area a is working. Thus:

[4] See for example Currarini *et al.* (2009), McPherson *et al.* (2001), Fong & Isajiw (2000) and Baerveldt *et al.* (2004).

[5] Areas are magisterial districts divided between urban and rural areas.

[6] We use alternatively the proportion of peers who are currently active/working. Nevertheless, our preferred specification uses active peers (working peers resp.) as network quality to estimate the labor-market participation (the probability to be employed resp.).

$$\overline{L}_{ial} = \frac{1}{n_{al} - n_h} \sum_{j \neq i, j \notin h} L_{jal}$$

where n_{al} denotes the number of individuals of language group l in area a, and n_h the number of working-age individuals in household h. Individual i and household h are excluded from the calculation of the mean, as Angrist & Pischke (2008) show that spurious correlation may result when the mean of the endogenous variable is used as an explanatory variable.[7]

We then construct a measure of the relative size of the network, which we call CA for "contact availability".[8] CA_{al} is formally defined as:

$$CA_{al} = \frac{s_{al}}{s_l} + 1$$

with s_{al} being the share of group l in area a and s_l being that of group l in the whole population. This is a proxy for the relative size of the pool of potential contacts in the local area. For robustness check, we use the proportion of neighbors speaking the same language at home. Our results are insensitive to this change. This measure is preferred for its nice properties for identification issue and because small ethnic groups are not underweight (Bertrand *et al.* (2000)). We will use the natural logarithm of this measure to account for potential decreasing returns in the size of the network: the more potential contacts you have, the less likely you are to exploit the full benefits provided by each additional contact. We add one to this ratio to ensure that the logarithm is always positive, which facilitates its interpretation.[9] Our network measure is then:

$$Netw_{ial} = ln(CA_{al}) \times \overline{L}_{i,al} \tag{1}$$

[7] If the individual is not excluded from the mean, the regression of L_{ial} on $\overline{L}_{ial}$ always has a coefficient of 1 (See Angrist & Pischke (2008) for direct proof). If the household members are not excluded, peer effects could be affected by intra-household decisions which refer to other issue like substitutability and specialization.

[8] See for details Bertrand *et al.* (2000).

[9] If $CA < 1$, so that $lnCA < 0$, then the measure of network (i.e. $lnCA * \overline{L}$) *decreases* as average employment in the neighborhood increases.

3 Data

We use the 10 percent sample of the 2001 South African survey, yielding information on 1,695,464 individuals aged between 16 and 65. This sample consists of 906,238 women (53.44 percent) and 789,706 men (46.56 percent). The census provides information about language spoken at home, which we use to proxy ethnic membership.

There are four geographical levels in the census data: province, district council, municipality and magisterial districts. There are nine provinces and 367 magisterial districts in South Africa (See Appendix B). We use magisterial districts, the smallest available geographical level, to define neighborhoods. We can only work with 278 of these districts as the smallest are grouped together in the database. We also use information on area type: rural or urban. Considering rural and urban areas as different geographical zones (in a given district), implies assuming that the peer groups of rural inhabitants do not include urban inhabitants.[10] Combined with the 11 language groups, we finally obtain 4,266 groups.[11]

We have three dependent variables: *labor-market participation* (P) is a dummy variable for the individual participating in the labor-market; *employment* (L) is a dummy variable for the individual being in employment;[12] and finally, *log wage per unit of time* ($\ln w$ or simply wages hereafter) is the log of annual total income divided by the number of hours worked per week.[13]

[Table 2 about here]

[10] We use alternatively the magisterial district to proxy the neighborhood instead of an interaction term between the area type (urban or rural) and the magisterial district (See Appendix D). This change reduces the magnitude of the effect, suggesting that social interactions occur mainly at the local level as the quality of information decreases with the distance.

[11] 4,266 does not equal $278 * 2 * 11$ as there are 1850 empty cells.

[12] An individual is recorded as being active in the Census if he is employed or unemployed. An individual is recorded as being employed in the Census if he responded "Yes: formal registered (non-farming)", "Yes: informal unregistered (non-farming)", "Yes: farming", "Yes: has work but was temporarily absent" to the question "In the seven days before 10 October did (the person) do any work for PAY (in cash or in kind profit or family gain, for one hour or more?" The Census attempts to apply UN and ILO standards in defining the unemployed as those who are out of work and actively seeking a job.

[13] We do not have information on the number of weeks worked per year, so we cannot compute the hourly wage, and our regressions may be misspecified if the number of weeks worked per year varies across individuals. Another limitation is that we cannot decompose income into wages and other income sources. Our estimates will be biased if non-wage income is correlated with the network. However, non-wage income is likely small relative to wages, as the average income of non-workers is less than 5% of that of workers.

Table 2 shows the sample size, and the active and employed population for each of the 11 language groups. Around 20% of our sample speaks IsiZulu, followed by IsiXhosa and Afrikaans speakers. Second, workplace behavior differs by language group. labor-market participation is over 80% for Afrikaans, English and Sesotho speakers, but under 70% for IsiXhosa speakers. Some language groups are thus over-represented[14] in the workplace (Afrikaans speakers constitute almost 15% of labor-market participants, 20% of the employed population, but only 13% of the sample) while others, such as IsiXhosa speakers, are underrepresented (14% of participants, 11% of employed population and 16% of the sample).

The proportion of workers also differs by language groups: while more than 60% of Afrikaans and English speakers work (over 50% of women and over 75% of men), the employment rate rarely exceeds 40% for IsiNdebele, Sesotho, Setswana and Siswati speakers, and is even lower for other groups (See Table 2). On average, only 43% of individuals in the sample are employed, (34% of women and 53% of men). Women never represent more than 45% of the work force whatever ethnic group we consider. The female to male ratio and the standard deviations summarize considerable variation in labor-market behavior across individuals by sex and ethnicity. The gender gap differs widely according to ethnic group: the female to male ratio varies between 0.73 (Xitsonga-speakers) and 0.87 (Sesotho-speakers) for the participation rate, and 0.42 (Xitsonga-speakers) and 0.75 (English speakers) for the employment rate. Moreover, within each gender the standard deviation of the employment rate fluctuates around 13.28.

Appendix C stresses the potential discrimination against women and some ethnic groups with some cautionary notes. We acknowledge the following limitations of any empirical work about discrimination. Indeed, Altonji & Blank (1999) argue that estimation of discrimination leads to two biases: discrimination could be overestimated, due to some unobservables, or underestimated due to pre-market discrimination (in education for instance). In Appendix C, the Heckman selection model[15] controls for characteristics at the individual and household levels, and adds dummies for area as fixed effect to minimize these biases. Results show that being

[14]Over representation means that the size of a language group in the employed population is bigger than the size in the sample. See Table 2.

[15]The probability to be employed is estimated only for the active population.

a male, speaking Afrikaans or English increases the probability to be employed. It confirms that women and some ethnic groups are less integrated in the South African labor market, as suggested by descriptive statistics.

[Table 3 about here]

Last, individual characteristics are not the same by language group (See Table 3). Most notably, Afrikaners and English speakers are more educated on average. The network means by language group for labor-market participation and employment are called Network1 and Network2, respectively. Xhosa speakers have the lowest network value for both of our variables of interest. These are followed by Zulu, Afrikaans and English-speakers for the network in labor-market participation, all of whom have values under the total mean. For the employment network, Zulu, Sepedi and Xitsonga-speakers all have low scores. On the contrary, Thsivenda and Siswati-speakers have the highest network values for labor-market participation and employment.

The inequalities in individual income between language groups are striking. Afrikaans- and English-speakers are the richest while Siswati- and Xitsonga-speakers are the poorest. This difference remains when we carry out the analysis by gender and is not explained by the number of hours worked, as the poorest work more hours that do the rich. As such, English- and Afrikaans-speakers record the highest hourly wage and Siswati- and Xitsonga-speakers the lowest.

4 Empirical strategy

4.1 Empirical framework

We first estimate the probabilities of participation and employment. We assume that these depend on a set of individual characteristics, local economic conditions and specific ethnic characteristics. We introduce networks into the models as follows:

$$Pr\left(P_{ihal}=1|X_{ihal},Y_a,Z_l\right)=F(Netw_{ial};X_{ihal};Y_a;Z_l;\epsilon_{ihal}) \tag{2}$$

$$Pr\left(L_{ihal}=1|X_{ihal},Y_a,Z_l\right)=F(Netw_{ial};X_{ihal};Y_a;Z_l;\epsilon_{ihal}) \tag{3}$$

where i indexes individuals, h households, a areas and l language groups; P_{ihal} and L_{ihal} are dummy variables for labor-force participation and employment, respectively. $Netw_{ial}$ is a network variable,[16] X_{ihal} are individual characteristics, Y_a geographical characteristics, Z_l ethnic group characteristics and ϵ_{ihal} is the error term. As the network variable is defined using the group mean we cluster errors at the group level.[17]

For the model to be empirically tractable, we adopt a linear specification:

$$Pr\left(L_{ihal}=1|X_{ihal},Y_a,Z_l\right)=\alpha Netw_{ial}+\beta X_{ihal}+\delta Y_a+\gamma Z_l+\epsilon_{ihal} \tag{4}$$

In other specifications, we allow α to vary by gender and ethnic group, and to be non-linear with respect to the number of potential contacts by using an interaction term. For example, to investigate the gender effect, we estimate the following equation:

$$Pr\left(L_{ihal}=1|X_{ihal},Y_a,Z_l\right)=\alpha_1 Netw_{ial}+\alpha_2 Netw_{ial}*Gender+\beta X_{ihal}+\delta Y_a+\gamma Z_l+\epsilon_{ihal} \tag{5}$$

To check the robustness of our results we also estimate equation (4) on subsamples defined by gender (See Appendix E) or ethnic group (unreported).

As shown by Manski (1993), the estimation of equation (4) presents a number of challenges. Manski proposes a useful terminology distinguishing *endogenous effects*, *exogenous effects* and *correlated effects*. Endogenous effects refer to the impact of peer behavior on that of the individual. This is what we want measure here, describing the effect of networks and social norms. Exogenous effects refer to the impact of the exogenous characteristics of the peers on individual behavior. If my geographical peers are highly-educated, I may be more likely to work due to human-capital spillovers. Last, correlated effects reflect that groups and networks are not formed at random, and that individuals in networks tend to share characteristics and/or face similar environments. Here it may be the case, if I am working, that I

[16]The construction of which is described in Section 2.

[17]If standard errors are not clustered by groups, the accuracy of the results is computed as if each single individual in a group is an independent observation, although this is not the case. The unit of observation is the group and clustering standard errors by group accounts for this fact.

like to live with other working people, because we share certain values.

Manski (1993) shows that "naive" regressions run into two problems. First, due to endogeneity, OLS estimation is biased, as network formation and location decisions are partly taken on the basis of variables that are unobserved by the econometrician, and which may be correlated with labor-market outcomes. For example, if talented people tend to be in contact with each other (but we do not completely observe talent), there will be a positive network correlation regarding employment status. It would however be misleading to interpret this as a peer effect, as talent likely has a positive direct impact on the probability of employment. Second, the model is plagued by identification problems as there is circular causality between the dependent variable and the covariates: networks favor employment but employment helps create networks. The consequence is that multiple values of the coefficients fit the data and the model is not fully identified.

Soetevent (2006) discusses four ways of overcoming these problems, via the use of: (i) data in which agents are assigned randomly to reference groups; (ii) data where only a fraction of agents within the group are treated; (iii) functional forms that explicitly account for inter-group differences by adding group-specific fixed effects; and (iv) instrumental variables. Given the non-random nature of our data, we rely on the last two strategies to attenuate the estimated social network bias due to contextual effects (Contreras *et al.* (2007)). Nevertheless, instead of using only the standard instrumental variable estimator (identification strategy No.2), first we instead appeal to a proxy and replace the local network measure by an aggregated measure at the province level (identification strategy No.1) and, last, we use the Hausman & Taylor (1981) instrumental variable technique (identification strategy No.3).

Moreover, Manski (1993), Araujo *et al.* (2004) argue that the non-linearity of the binary choice model allows empirical identification of the peer effect. Only individuals who cross the threshold between not working and working contribute to the likelihood: a change in the covariates does not affect individuals who are already working.

4.2 Fixed effects as controls for correlated effects

To allow for correlated effects, we add sets of dummies to our model. We first add a complete set of 278 district dummies, picking up differences in local labor-market conditions which may act as contextual effects. Dividing the country up into sufficiently small geographical areas should ensure that we control for local labor demand. Our model is identified as the peer effect is district times rural/urban times language group, so that we have variation in peer group labor-market outcomes even within districts.

In addition, a correlation between individual and group behavior may result from unobserved group characteristics which affect labor-force participation. Assume for example that some groups are discriminated against in the labor-market or in schools. They may then have a lower probability of being employed when looking for work, or be offered a lower wage so that fewer individuals will look for work. In this case, any observed correlation cannot be interpreted as a causal impact of individual behavior. To avoid this problem, we include language-group dummies. In unreported robustness checks, we also include a dummies for each language/area combination.

We last include a measure of contact availability as a control variable. We do so because location choice may be correlated with omitted variables which are themselves correlated with employment. For example, some workers may be more mobile and willing to move than others, so that they may uncover job opportunities that are far from home at the expense of leaving their friends. This might be why this control has a negative estimated coefficient.

Our empirical model can then be written as:

$$L_{ihal} = \alpha Netw_{ial} + \beta X_{ihal} + \delta_a + \gamma_l + \omega lnCA_{al} + \epsilon_{ihal} \tag{6}$$

where δ_a and γ_l are respectively area and language-group dummies (fixed effects).

4.3 Three strategies for the endogeneity problem

4.3.1 Using an aggregated measure of network

Following Bertrand *et al.* (2000), we replace the local network measure by an aggregated measure at the province level. This way of aggregating the data at the province level avoids the remaining omitted variable bias. Unobserved variables that are common to a given language group in a given area and correlated with employment status will produce what look like peer effects. The aggregated network measure corrects this problem, as it is uncorrelated with the common characteristics of the local group. Our assumption is that people take the average characteristics of networks in the province as given. They construct their networks at the magisterial district level and may even choose their location on the basis of network size or quality. Nevertheless, we assume that they do not move province in order to benefit from better networks so that there is no self-selection at the provincial level.[18]

We thus estimate:

$$L_{ihal} = \alpha Netw^{p}_{ial} + \beta X_{ihal} + \delta_a + \gamma_l + \omega lnCA_{al} + \epsilon_{ihal} \tag{7}$$

where $Netw^{p}_{ial}$ is a network measure using average employment at the province instead of the district level:

$$Netw^{p}_{(al} = \ln(CA_{al}) \times \overline{L}_{ipl}) \tag{8}$$

The pattern of network quality at the province level has to be exogenous with respect to employment status and conditional on our controls. For example, very mobile individuals may move between provinces to pick up job opportunities. If mobility is a characteristic valued by employers, we may find that people located in provinces offering better networks on average have a greater probability of employment even though this does not reflect peer effects. To avoid this, we control for contact availability in all of our specifications.

4.3.2 Using the distribution of the sex of peers' sibling as IV

We use the standard 2SLS estimator and the distribution of the sex of peers' sibling as instrumental variable. This paper assumes that South African households have son-preferences or diversity'-preferences. Angrist & Evans (1998), Maurin & Moschion (2009) show evidence on the positive correlation between the sex of the

[18] To test the robustness of our results, we use the mean of active or working individuals of the whole language group at the province level minus those at the managerial district level. The results remain significant.

oldest siblings and the final number of children of a household in US and France, respectively. Thus, the participation of an individual in the labor market L_{ihal} is influenced by the sex of his/her oldest siblings because parents having same-sex children (denote SS[19]) tend to have more children.

$$L_{ihal} = F(SS_{ihal}, X_{ihal}) \tag{9}$$

Appendix F presents the correlation coefficient between having same-sex children and the total number of children, the probability to be active, employed, the proportion of peers having same-sex children and the neighborhood respectively. It suggests that the distribution of the sex of peers' sibling is a valid instrumental variable, correlated with the total number of children, the labor-market behaviour and uncorrelated with the neighborhood choice. This means that having two girls or two boys encourage to have a third child, that decreases both the probabilities to be active and employed (Rosenzweig & Wolpin (2000)). Among household with same-sex siblings, the employment rate is about 47.50%. This is about 1 point lower than among household with different-sex siblings (46.65%). Maurin & Moschion (2009) on French data and Angrist & Evans (1998) on US data find the same stylised facts with a higher magnitude than in South Africa.

In contrast, the sex of the oldest siblings does not have any perceptible influence on neighborhood choices: no correlation between the sex of the siblings of an individual and the sex of the siblings of the other individuals living in the same close neighborhood. Nevertheless, we cannot exclude that the sex of children does not influence the neighborhood choice. Indeed, having same-sex children increases the probability to have a third child and then to need a bigger house. If the size of the dwelling is correlated with the neighborhood localization, we can suppose that household having same-sex children tend to be concentrate in some neighborhood. However, we do not observe any concentration of household having same-sex siblings in the local ethnic group (See Appendix F). Assuming no correlation between the sex of the siblings of an individual and the sex of the siblings of the other individuals living in the same close neighborhood speaking the same language, i.e.

[19] $SS = 1$ if the two oldest children of an individual have the same sex.

$E(SS_{ial}|\overline{SS}_{ial}) = 0$, our empirical strategy is:

$$L_{ihal} = \alpha \widehat{Netw}_{ial} + \beta X_{ihal} + \delta_a + \gamma_l + \omega lnCA_{al} + \epsilon_{ihal} \tag{10}$$

where

$$\widehat{Netw}_{ial} = F(\overline{SS}_{al})^{20} \tag{11}$$

4.3.3 Using a quasi-panel data approach

Our last identification strategy consists to adopt a quasi-panel data approach. This methodology introduced by Graham & Hahn (2005) allows identification of peer effects in a linear-in-means model. The number of observed local ethnic groups are treated as cross-sectional dimension and the number of sampled individuals within each local ethnic groups as time-series dimension.

Using a quasi-panel data approach justifies the use of Hausman & Taylor (1981) instrumental variable technique developed for panel data models. Those instruments generate exogenous between-group variations which create extra information and allow the social multiplier identification (Graham & Hahn (2005)).

Identification is obtained by combining the sources of instruments suggested by Hausman & Taylor (1981) with fixed effects to control for the contextual effects. The Hausman-Taylor estimator provides consistent and efficient estimates of the coefficient associated with singly exogenous individual-invariant variables, despite the absence of external instruments. This approach provides two source of instruments: (i) group means of exogenous variables and (ii) deviation from groups means of individual characteristics. Indeed, the deviations from local ethnic group means of all X_{ihal} are uncorrelated with the group-level error term by construction. This approach requires more explanatory variables without contextual effects than correlated variables for identification (Knight & Gunatilaka (2009)). Thus, the means of these variables (age, age squared, gender, marital status) and the mean deviations of all explanatory variables are used as instrumental variables.

[20]The mean is computed excluding the individual and the household members.

4.4 Heckman

We estimate a Heckman selection model to see whether our measure of networks influences wages. Only individuals who have a job have a wage, therefore we can only test for peer effects on wages in the subsample of the employed.

However, this estimation will suffer from selection bias if unobservable or omitted characteristics in the selection equation contribute to the determination of wages. The sample of individuals participating in the labor-market may thus contain individuals with specific characteristics correlated with wages which are included in the error term ϵ. Hence, the impact of observed individual characteristics X_i is mis-estimated.

We thus estimate the probability of having a job in a first stage, which yields the Inverse Mills Ratio (see Equation 12). In the second step, wages are estimated including this Inverse Mills Ratio (see Equation 13). However, identification in the second equation is based on the nonlinearity of the Mills ratio. If the variation in the individual characteristics is only small, the wage equation will exhibit considerable collinearity and therefore imprecise estimates. The first equation (selection equation) should thus contain one or more additional explanatory variables that do not figure in the second equation (V_i: the additional identifying constraints). In our model, one excluded variable is introduced into the employment equation: a dummy variable for the individual being a homeowner, insofar as ownership influences behavior on the labor-market by determining individual mobility. Oswald (1996) suggests that workers are less likely to move if they own a house.

The following system is thus estimated using the instrument:[21]

$$L = \widehat{Netw} + \beta X + \nu V_i + \delta_a + \gamma_l + \omega lnCA + \epsilon \tag{12}$$

$$lnW = \alpha\widehat{Netw} + \beta X + \delta_a + \gamma_l + \omega lnCA + \rho\lambda + \epsilon \text{ if } L = 1 \tag{13}$$

where λ is the Inverse Mills Ratio which corrects for selection bias, L stands for employment, lnW the log of wages and V_i the additional identifying constraints.

[21]Subscripts are omitted for clarity.

5 Results

5.1 Peer effects as a social norm

In our first specification, we estimate a probit model with robust clustered standard errors for labor-force participation. Table 4 presents the estimation results using the aggregate data strategy to deal with endogeneity problems. Fixed effects for language groups and areas are also included to deal with simultaneity problems. The right-hand side includes socio-demographic controls, our measure of networks and a measure of contact availability. The socio-demographic controls include a gender dummy, four education dummies, age and age-squared, the number of children at home, five dummies for marital status, and the number of individuals in the household who participate in the labor market.

The socio-demographic controls attract the expected signs: higher education and being a man with few children increase the probability of participation. Experience and age both have non-linear effects. Contact availability attracts a negative coefficient in the regressions. One interpretation is that motivated people are willing to move away from their peers in order to obtain a job. As such, this control successfully screens out some of the unobservable individual characteristics that could have biased our estimates. Finally, the number of other working household members attracts a positive and significant sign. This likely reflects omitted household variables that are correlated with employment: for example, wealth, pressure from family members or common attitudes toward work.

[Table 4 about here]

Regarding peer effects on labor-force participation, network measure attracts a positive and significant coefficient whatever the identification strategy used (See Table 4, column (1) for the first identification strategy, Table 5 for the second one and 6 for the last one). This implies that the more peers participate in the labor-market, the higher is the individual probability of participation. This can be interpreted as social pressure to participate from peers. This norm of participation describes social behavior conveyed by society and internalized by individuals in the process of their socialization through peers. In developing countries, where community laws dominate individual laws, men and women respect these constraints (Coleman (1990)).

Therefore, individuals surrounded by active peers may have an additional incentive to look for a job via social pressure. This is consistent with Clark (2003), who finds that own unemployment is more painful when surrounding peers are working. The elasticity of the participation probability with respect to our network measure is 0.29[22]

[Table 5 about here]

However, this is only a "first-order" figure that does not take into account any snowball effects. If we assume that the snowball effect acts as a geometric sequence, the additional effect on activity of a one percent exogenous increase in local activity is:

$$\frac{1}{1 - \overline{CA} \times \alpha} - 1$$

which is equal to 0.4. The reality is probably somewhere in between these two extremes. It is unrealistic to assume that there is no "snowball effect", but also to think that this social multiplier always works until the full effect is obtained. As such we can interpret these figures as upper and lower bounds.

[Table 6 about here]

Table 4 includes an interaction term with sex and provides evidence of gendered peer effects (column (2)):[23] on average, the participation peer effect is higher for women. However, it is worth going deeper here. When we distinguish the gendered peer effect by ethnicity, Afrikaans- and English-speakers stand out with a stronger peer effect for men than for women, while this is inversed in other ethnic groups.

[Table 7 about here]

The gender effect is considerable with a non-linear specifications (See Figure 1).[24] Table 7 presents the first results about non-linearity. It suggests that network have non-linear return. The size of our database allows to go beyond this simple relation

[22]The average value of CA is 1.51: the net peer effect is then computed as $0.194 \times 1.51 = 0.292$.

[23]We can test the robustness of these results using female sub-samples: this does not change the size or significance of the coefficients.

[24]These results are consistent with previous research providing some evidence of non-linearities in social interactions (Clark & Loheac (2007)).

and to analyze the network effect according to the size of the network. Thus, the network coefficient is computed here for each five-percent block of contact availability. This exercise confirms that peer effects are non-linear and gender-specific (suggested by Table 7): the coefficients are significant and fall with contact availability for women but rise for men. There is thus concave and convex social pressure for women and men, respectively, regarding labor-force participation. While the first contacts have a significant influence on women's participation, this is true only after a certain threshold for men. Thus the gender gap is considerable for the first contacts and falls with the number of available contacts. The convexity of the peer effect on participation suggests the possibility of multiple equilibria. When few men participate to the labor market, additional participation has little effect and men may remain stuck in an inactivity trap as the snowball effect fails to occur. However, as the network grows, the peer effect gets higher, enabling the possibility of a second equilibrium in which the participation rate is substantially higher. This is not true for women. Even small networks have a huge marginal effects on participation.

Finally, Figure 2 depicts the heterogeneity between ethnic groups: among women, the social pressure is lower for those who have higher network quality, while for men, it is lower for those who have a larger network.

5.2 Peer effects as a matching device

If peer effects work only via social norms, we would expect wages to fall with the number of working peers, as the labor demand curve should not move. In this section, we ask whether peer effects also improve job search and matching. We estimate therefore the impact of networks on employment and wages.

5.2.1 Employment

As in the previous section, a probit model with robust clustered standard errors was estimated for employment (see equation (7)) including fixed effects[25] and control variables. The results of our most basic specification are shown in the first column of Table 8. As in the previous section, the control variables have the expected sign. All else equal, men and those living in urban areas have a higher probability

[25]These are sets of dummies for language groups and areas.

of having a job. The number of children has a positive sign but the direction of causality is unclear. Children may provide additional incentives to work, or may reduce the reservation wage because of the need for additional resources to bring them up. Alternatively, individuals may have fewer children when they lack financial resources.

There are sizeable peer effects regarding the probability of employment: as the number of working peers increases, the individual probability of working rises with an elasticity of 0.43. The results remain whatever the strategy used (See Tables 5 and 6). This confirms that working peers are important in helping individuals to find a job (Granovetter (1974, 1995)). Since peer effects occur via networks, working peers can be considered as channels of information transmission between network members. Working peers can thus help to reduce information asymmetries and statistical discrimination, since firms base their hiring decisions on average rather than individual productivity (Arrow (1971), Phelps (1972)). Moreover, job offers can be obtained from direct and indirect acquaintances, through word-of-mouth communication (de Martí & Zenou (2009)). By reducing employer uncertainty about worker productivity, networks may enhance the matching between job-seekers and available jobs (Marsden & Gorman (2001)).

[Table 8 about here]

Second, column (2) of Table 8 displays the coefficients estimated separately for men and women. The effect is slightly (and significantly) higher for men, which suggests that networks have a higher return for men. This overall correlation hides some non-linearities, as revealed in Figure 3. The "network gap" is very large at the bottom end of the distribution, but the estimated coefficients become much smaller and closer to each other as the number of potential contacts increases.[26] Men with few contacts have high returns to networks, so that even a few contacts can make a considerable difference. This is also true for women but to a lesser extent. Table 7 suggests a decreasing return to networks for both men and women, confirmed by Figure 3. These results suggest that a small network is much better that no network

[26] Our specification does not allow us to use quantiles based on the distribution of $\overline{L}$ because of reflexivity, i.e. the quantile would depend on the endogenous variable. By doing so, we would mix together individuals with a low return to networks and those who have a low probability of employment for other reasons.

at all as the marginal effect is quite high for people with few contacts. Therefore, finding ways to help isolated individuals to build networks is a matter of public policy and may be quite efficient to fight poverty.

Finally, Figure 4 shows the different effects by language group and gender. The peer effect on the probability of working is significantly lower for English- and Afrikaans-speakers among women, and for English-speakers among men. The ethnic groups which are the most integrated[27] on the labor market thus have the smallest peer effects. English- and Afrikaans-speaking women are overrepresented on the labor force: they represent respectively 7.41% and 12.73% of the sample, but 13.91% and 20.36% of the labor force. Despite their labor-market presence, their peer effect elasticity is about 0.23 and 0.37 respectively, as against 0.41 for women on average We interpret this as a selection effect. Unemployment is explained by individual characteristics and characteristics that are common to the whole group. Conditional on a belonging to a group that is better integrated in the labor market, unemployment is associated with worse individual characteristics. Those who belong to 'good' groups, have 'worse' individual characteristics when unemployed. Therefore, in better-integrated groups networks are likely to be less effective.

5.2.2 Wages

We estimate the effect of networks on wages using a two-step Heckman model to correct for selection bias. The same specification as above is applied. We add dummy variables for employment and industry, and the distribution of the sex of peers' sibling and home ownership are used as exclusion variables.

Most of the estimated coefficients are unsurprising. As before, the effect of age is positive and concave. Wages are positively associated with education, being a man, the number of children and having other people working in the household.

Once again, peer effects are positive, large and significant, as shown in Table 9. This is important because it shows that peer effects are not restricted to the labor supply curve. There are three explanations for this result. First, it is possible that networks improve matching. If productivity is firm-specific, a large network can help individuals to maximize wages over the set of firms. Second, if networks reduce the

[27] These groups are respectively 1.9 and 1.6 times more likely to be in the labor force than the overall population.

cost of effort in job search, they may increase the reservation wage. Third, in a labor market with imperfect information about worker productivity, networks can be used to signal ability. In such a game involving one employer and one worker, a third party can put its reputation at stake and produce a signal regarding worker productivity. If the worker turns out to have low productivity, the reputation of the third party will suffer, so that the employer will trust the signal less.

[Table 9 about here]

This can be related to the finding that men benefit slightly less from peers than do women. Women workers will benefit more from peers if the variation in their productivity is larger, as the larger is this variation the more valuable are productivity signals. This is consistent with the variance in our wage measure being higher for women (1.352) than for men (1.336). This observation is robust to a decomposition by language group (see Figure 5).

The peer effect on wages is concave for both sexes (see Figure 6). The marginal value of additional peers is decreasing: having a small network instead of no network greatly increases the odds of finding a better-paid job, but having a large network instead of a small one only slightly increases the average wage.

Again, Afrikaans- and English-speaking individuals have the lowest peer effect, in contrast with their active presence on the labor-market relative to other language groups.

5.3 Robustness checks

Table 10 shows how the network coefficient varies across subsamples. This section checks whether our results are driven by sample characteristics. Since 20% of our sample are IsiZulu speakers (see Table 3), we check to see if our results are driven by this group. In row 2, the peer effects remain positive and significant for the three dependent variables (labor-market participation, Employment and Wages) when IsiZulu speakers are excluded. The gender effect remains significant and of the same sign as above.

We also exclude the overrepresented groups in the labor-market identified in Table 3, namely Afrikaans- and English-speakers, in rows 3 and 4 respectively. The peer effect stays positive and significant for the three dependent variables. The gender effect also remains, apart from when we exclude Afrikaans speakers in the wage equation. These exercises confirm the robustness of our results and suggest that having a higher *quality* or *size* of network does not affect the existence of the peer effect.

[Table 10 about here]

Whatever the sample analysed, the peer effects remain positive and significant. In row 5, we only include individuals with children: the coefficients remain positive and significant. In row 6, we restrict our sample to individuals aged between 25 and 40 so as to focus on the cohort that is the most likely to be active and to begin their career. The coefficient falls for labor-force participation, suggesting that social pressure increases with age.

In unreported robustness checks, we distinguish the network effect according to the gender and the ethnic group of an individual in sub-samples. The results remain unchanged: the peer effects on labor-force participation and wages are higher for females, whereas that on employment is higher for males (in all ethnic group); ethnic membership determines the magnitude of the peer effect. Moreover, alternative measures of contact availability ($ln(s_{al}+1)$ instead of $ln(\frac{s_{al}}{s_l}+1)$) and group definition (by area, language and gender) yield similar results.

6 Policy implications and concluding remarks

In this paper, we assess the presence of peer effects on the South African labor market. For a given individual, the number of peers who participate in the labor-market (resp. have a job) influences the decision to participate (resp. labor-market outcomes). A one percent increase in the proportion of working peers raises the probability of participation and the odds of finding a job by over 30% and has a substantial effect on wages. Social interactions seem to be an additional determinant of employment along the same lines as education and experience. These findings suggest the presence of two mechanisms on the labor market. First, the fact that participation is not independent of peers' decisions shows that social norms or social pressures are at work. Being surrounded by active peers provides social incentives to look for a job. The decision to look for a job is conditioned by others' behavior. This externality may generate multiple equilibria, some with high employment and others with low employment. This is a plausible explanation for the considerable variance of average employment between ethnic groups.

Second, networks seem to be a useful tool during job search, increasing wages and the probability of finding a job. The unique feature of networks is that they allow individuals to signal their productivity and overcome market failures tied to asymmetric information. If individual i suggests to employer j that he should recruit worker k, he puts his reputation at stake. If the workers turns out to have low productivity, the employer will no longer trust i's suggestions. This mechanism provides incentives to i to signal k's productivity as best as he can.

Networks seem to be more efficient for language groups that are not well integrated in the labor market. Peer effects are much lower for Afrikaans- and English-speaking individuals, although these groups are largely overrepresented on the labor market. This suggests that networks may be a tool to reduce economic inequalities between ethnic groups.

Another implication is that peer effects create social multipliers. Employment shocks can be amplified by a considerable amount (elasticity between 0.3 and 0.5). It is important for political purposes as the snowball effect due to social interaction within local ethnic groups amplifies the extent of labor-market policies. The effect of affirmative action focusing on the reduction of discrimination and labor-market inequality between ethnic group or gender cold be amplified by networks.

Moreover, recent research in labor economics has emphasized the key role of the matching process. We argue that networks are central to the efficiency of this process and greatly increase the rate at which vacant jobs are filled in addition to shortening unemployment spells.

Finally, spatial segregation may, in fact, play an economic role. Our results suggest that peer effects are strong within each local ethnic group, so that it is easier to build networks when surrounded by people from the same ethnic group. Under this assumption, spatial segregation is a way of maximizing information flows when labor-market institutions do not fulfill this role.

Bibliography

Alba, Richard D. 1990. *Ethnic Identity: The Transformation of White America.* Yale University Press.

Alesina, Alberto F., Glaeser, Edward L., & Sacerdote, Bruce. 2006. Work and Leisure in the U.S. and Europe: Why So Different? December, 1–100.

Altonji, G., & Blank, M. 1999. Race and Gender in the Labour Market. *Chap. 48, pages 3143–3258 of:* Ashenfelter, O., & Card, D. (eds), *Handbook of Labour Economics*, elsevier edn., vol. 3.

Angrist, Joshua D, & Evans, William N. 1998. Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *American Economic Review*, **88**(3), 450–77.

Angrist, Joshua D., & Pischke, Jörn-Steffen. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

Araujo, Caridad, de Janvry, Alain, & Sadoulet, Elisabeth. 2004. Peer Effects in Employment: Results from Mexico's Poor Rural Communities. *Department of Agricultural & Resource Economics, UC Berkeley, Working Paper Series*, Aug.

Arrow, Kenneth. 1971. The Theory of Discrimination. *Working Paper of Princeton University, Department of Economics, Industrial Relations Section.*, **403**(5).

Baerveldt, Chris, Duijn, Marijtje A. J. Van, Vermeij, Lotte, & Hemert, Dianne A. Van. 2004. Ethnic boundaries and personal choice. Assessing the influence of individual inclinations to choose intra-ethnic relationships on pupils' networks. *Social Networks*, **26**(1), 55 – 74.

Banerjee, Abhijit V. 1992. A Simple Model of Herd Behavior. *Quarterly Journal of Economics*, **107**(3), 797–817.

Bertrand, Marianne, Luttmer, Erzo F. P., & Mullainathan, Sendhil. 2000. Network Effects And Welfare Cultures. *Quarterly Journal of Economics*, **115**(3), 1019–1055.

Bhorat, Harron, Leibbrandt, Murray, Maziya, Muzi, van der Berg, Servaas, & Woolard, Ingrid. 2004. *Figthing Poverty: Labour markets inequality in South Africa*. Gary Fields and Erik Thorbecke, Cornell University.

Bikhchandani, Sushil, Hirshleifer, David, & Welch, Ivo. 1992. A Theory of Fads, Fashion, Custom, and Cultural Change in Informational Cascades. *Journal of Political Economy*, **100**(5), 992–1026.

Borjas, George J. 1992. Ethnic Capital and Intergenerational Mobility. *Quarterly Journal of Economics*, **107**(1), 123–50.

Borjas, George J. 1994. Long-Run Convergence of Ethnic Skill Differentials. *NBER Working Papers*, **4641(2)**(Feb.).

Borjas, George J. 1995. Ethnicity, Neighborhoods, and Human-Capital Externalities. *American Economic Review*, **85**(3), 365–90.

Burns, Justine, Godlonton, Susan, & Keswell, Malcolm. 2010. Social Networks, Employment and Worker Discouragement: Evidence from South Africa. *Labour Economics*, **17**(2), 336 – 344.

Calvo-Armengol, Antoni, & Jackson, Matthew O. 2004. The Effects of Social Networks on Employment and Inequality. *American Economic Review*, **94**(3), 426–454.

Clark, Andrew E. 2003. Unemployment as a Social Norm: Psychological Evidence from Panel Data. *Journal of Labor Economics*, **21**(2), 289–322.

Clark, Andrew E., & Loheac, Youenn. 2007. It wasn't me, it was them! Social influence in risky behavior by adolescents. *Journal of Health Economics*, **26**(4), 763–784.

Coleman, James S. 1990. *Foundations of Social Theory*. Harvard University Press.

Contreras, Dante, Kruger, Diana, Ochoa, Marcelo, & Zapata, Daniela. 2007. The role of social networks in employment outcomes of Bolivian women. *University of Chile, Department of Economics Working Paper*, **251(1)**.

Cornwell, Katy, & Inder, Brett. 2008. Language and Labour Markets in South Africa. *Journal of African Economies*, **17**(3), 490–525.

Currarini, Sergio, Jackson, Matthew O., & Pin, Paolo. 2009. An Economic Model of Friendship: Homophily, Minorities, and Segregation. *Econometrica*, **77**(4), 1003–1045.

de Martí, Joan, & Zenou, Yves. 2009. Social Networks. *IZA Discussion Papers*, **4621**.

Fernandez, Raquel, & Fogli, Alessandra. 2005. Culture: An Empirical Investigation of Beliefs, Work, and Fertility. *NBER Working Papers*, **11268**.

Fishman, Joshua A. 2001. *Handbook of Language and Ethnic Identity.* Oxford University Press.

Fong, Eric, & Isajiw, Wsevolod W. 2000. Determinants of Friendship Choices in Multiethnic Society. *Sociological Forum*, **15**(2), 249–271.

Glaeser, Edward L., Sacerdote, Bruce I., & Scheinkman, Jose A. 2003. The Social Multiplier. *Journal of the European Economic Association*, **1**(2-3), 345–353.

Graham, Bryan S., & Hahn, Jinyong. 2005. Identification and estimation of the linear-in-means model of social interactions. *Economics Letters*, **88**(1), 1–6.

Granovetter, Mark S. 1974. *Getting a Job: A Study of Contacts and Careers.* Cambridge, MA.

Granovetter, Mark S. 1995. The Economic Sociology of Firms and Entrepreneurs. *In: The Economic Sociology of Immigration*, vol. 1. A. Portes, eds., New York: Russel Sage Foundation.

Hausman, J., & Taylor, W. 1981. Panel Data and Unobservables Individual Effects. *Econometrica*, **49(6)**, 1377–1398.

Hofmeyr, Andre. 2010. Social Networks and Ethnic Niches: an Econometric Analysis of the Manufacturing Sector in South Africa. *South African Journal of Economics*, **78(1)**, 107–130.

Hoxby, Caroline. 2000. Peer Effects in the Classroom: Learning from Gender and Race Variation. *NBER Working Papers*, **7867 (1)**.

Jackson, Matthew O. 2008. *Social and economic networks*. Princeton Universty Press.

Jencks, Charles, & Mayer, S. 1990. The Social Consequences of Growing Up in a Poor Neighbourhood. *In: Inner-City Poverty in the United States*. Lynn, L. and McGeary, M. (eds), National Academy Press: Washington DC.

Keswell, M., & Poswell, L. 2002. How Important is Education for Getting Ahead in South Africa? *CSSR Working Paper, Cape Town*, **22(1)**.

Kingdon, Geeta G., & Knight, John. 2004. Unemployment in South Africa: The Nature of the Beast. *World Development*, **32**(3), 391–408.

Kingdon, Geeta G., & Knight, John B. 2000. Are searching and non-searching unemployment distinct states when unemployment is high? The case of South Africa. *CSAE Working Paper Series*, **2**.

Knight, John, & Gunatilaka, Ramani. 2009. Is Happiness Infectious? *Economics Series Working Papers*.

Manski, Charles F. 1993. Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies*, **60**(3), 531–42.

Marsden, P., & Gorman, E.H. 2001. Social Networks, Job Changes, and Recruitment. *In: Sourcebook of Labor Markets: Evolving Structure and Processes*. Ivar Berg and Arne Kalleberg. Kluwer Academic/Plenum Publishers, NY.

Maurin, Eric, & Moschion, Julie. 2009. The Social Multiplier and Labor Market Participation of Mothers. *American Economic Journal: Applied Economics*, **1**(1), 251–72.

McPherson, Miller, Lovin, Lynn S., & Cook, James M. 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, **27**(1), 415–444.

Oswald, Andrew J. 1996. A Conjecture on the Explanation for High Unemployment in the Industrialized Nations : Part I. *Warwick Economics Research Paper Series*, **475**.

Phelps, Edmund S. 1972. The Statistical Theory of Racism and Sexism. *American Economic Review*, **62**(4), 659–61.

Rosenzweig, Mark R., & Wolpin, Kenneth I. 2000. Natural "Natural Experiments" in Economics. *Journal of Economic Literature*, **38**(4), 827–874.

Sacerdote, Bruce. 2001. Peer Effects With Random Assignment: Results For Dartmouth Roommates. *Quarterly Journal of Economics*, **116**(2), 681–704.

Soetevent, Adriaan R. 2006. Empirics of the Identification of Social Interactions; An Evaluation of the Approaches and Their Results. *Journal of Economic Surveys*, **20**(2), 193–228.

7 Tables and Figures

Table 1: The maximum concentration rate in provinces by language group

language	Concentration rate	Province
IsiNdebele	49.09	Mpumalanga
IsiXhosa	58.45	Eastern Cape
IsiZulu	64.48	KwaZulu-Natal
Sepedi	55.42	Limpopo
Sesotho	44.19	Free State
Setswana	61.56	North West
SiSwati	76.82	Mpumalanga
Tshivenda	73.95	Limpopo
Xitsonga	48.87	Limpopo
Afrikaans	44.36	Western Cape
English	31.08	Gauteng

The ethnic group is significantly correlated with the province (0.5***) and the area (0.52***).

Table 2: Labor-market representation by language group.

	% in S.	% in A.	% in E.	Act. rate	Emp. rate	Repr.
Afrikaans	13.16	14.05	19.76	80.39	64.36	**1.50**
English	7.66	8.44	13.23	82.96	74.05	**1.73**
IsiNdebel	1.93	1.96	1.83	76.47	40.51	0.95
IsiXhosa	16.03	14.84	11.57	69.67	30.93	0.72
IsiZulu	20.79	20.08	15.83	72.70	32.64	0.76
Sepedi	9.83	9.46	8.76	72.49	38.20	0.89
Sesotho	9.19	9.88	9.02	80.91	42.07	0.98
Setswana	10.40	10.51	10.10	76.03	41.62	0.97
Siswati	3.07	3.07	2.89	75.27	40.39	0.94
Tshivenda	2.50	2.43	2.17	73.43	37.34	0.87
Xitsonga	5.44	5.28	4.84	73.06	38.17	0.89

S. A. and E. refers to the size of each language group in the sample, in active and employed population respectively. Repr. refers to the labor-market representation of each language group ($\%inE./\%inS.$).

Table 3: Summary Statistics

Ethnic Group	Age	No. Children	Education			No. Hours worked			Wage			Network Quality 1	Network Quality 2	Network Quantity
	All	All	All	F	M	All	F	M	All	F	M	All	All	All
Afrikaans	35.98	1.17	9.48	9.44	9.52	44.14	42.10	45.79	6.62	6.46	6.76	0.81	0.65	1.27
English	38.03	1.09	11.03	10.90	11.16	43.52	41.18	45.41	7.21	7.05	7.34	0.83	0.73	1.32
IsiNdebel	35.03	1.16	7.03	6.90	7.16	45.58	43.38	47.18	5.73	5.57	5.85	0.77	0.41	2.10
IsiXhosa	34.84	1.18	6.82	6.98	6.64	44.80	43.00	46.19	5.73	5.65	5.80	0.73	0.32	1.38
IsiZulu	34.21	1.34	6.81	6.65	7.00	45.70	44.10	46.97	5.74	5.58	5.87	0.76	0.34	1.28
Sepedi	34.67	1.25	7.62	7.36	9.91	46.50	45.46	47.22	5.79	5.57	5.95	0.71	0.36	1.73
Sesotho	35.02	1.09	7.68	7.70	7.67	46.26	44.04	47.90	5.62	5.42	5.77	0.81	0.42	1.58
Setswana	35.20	1.10	7.58	7.68	7.46	45.62	43.64	47.13	5.77	5.68	5.83	0.76	0.42	1.75
Siswati	34.06	1.35	6.81	6.62	7.03	47.27	45.39	48.48	5.55	5.34	5.67	0.75	0.41	2.59
Tshivenda	34.29	1.50	7.61	7.15	8.15	46.21	45.31	46.76	5.80	5.53	5.96	0.74	0.38	2.85
Xitsonga	33.45	1.24	6.47	6.21	6.73	47.75	46.19	48.40	5.60	5.45	5.65	0.73	0.38	1.76
All	35.03	1.21	7.73	7.64	7.83	45.14	43.19	46.61	6.15	6.01	6.25	0.77	0.44	1.51

F= Female; M= Male; Income is yearly income in logs; Hours is hours worked per week; Wage refers to the log of the ratio of yearly income to weekly hours worked; Network quantity refers to the available contact and network quality to the proportion of peers who are currently active (1) or working (2) respectively.

Table 4: Peer effects as social norms

Labor-Force Participation	(1)	(2)
Network	0.194***	0.201***
	(14.41)	(14.68)
Gender*Network		-0.021***
		(2.88)
Gender	0.164***	0.186***
	(35.41)	(23.02)
Age	0.035***	0.035***
	(64.86)	(64.82)
Age^2	-0.045***	-0.045***
	(75.03)	(75.10)
Contact availability	-0.146***	-0.146***
	(14.46)	(14.44)
HH	0.071***	0.071***
	(32.11)	(32.03)
Primary Education	0.040***	0.040***
	(23.00)	(23.07)
Secondary Education	0.133***	0.133***
	(57.33)	(57.41)
Tertiary Education	0.156***	0.156***
	(36.90)	(36.90)
No. Children	-0.006***	-0.006***
	(13.10)	(13.19)
Urban	0.006	0.006
	(1.34)	(1.34)
Observations	1695941	1695941

Columns (1) and (2) present the marginal effects. Dummies for language groups and areas are included as fixed effects. Absolute value of z-statistics in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Contact availability refers to $ln(CA_{al})$ as specified in the text. Network refers to $Netw^p_{ial} = ln(CA_{al}) \times \overline{L}_{i,pl}$ as specified in the text. HH refers to the number of working household members. Standard errors are clustered at the group level. Age^2 refers to $Age^2/100$.

Table 5: Instrument for $Netw_{ial}$: son preference or 'diversity' preference

Second Stage	Labor Market Participation	Employment
Network	0.294***	0.391***
	(3.35)	(4.72)
Gender*Network	- 0.082***	0.063***
	(24.12)	(24.12)
Observations	840016	840016
R squared	0.19	0.28
First Stage	Network1	Network2
$avr - hh - samesex$	-0.182***	-0.124***
	(24.83)	(24.05)
Constant	1.205***	0.673***
	(320.88)	(255.48)
R-squared	0.38	0.49

Dummies for language groups and areas are included as fixed effects. Absolute value of z-statistics in parentheses for the 2SLS and t-statistics for the first stage. *, **, *** significant at the 10%, 5%, 1% level. Network1 (Network2 resp.) refers to $Netw_{ial} = ln(CA_{al}) \times \overline{P}_{ial}$ ($Netw_{ial} = ln(CA_{al}) \times \overline{L}_{ial}$ resp.) as specified in the text.

Table 6: Hausman-Taylor estimator

	labor Market Participation	Employment
Network	0.660***	0.596***
	(66.60)	(22.27)
Observations	1676689	1676689
Number of group	3446	3446

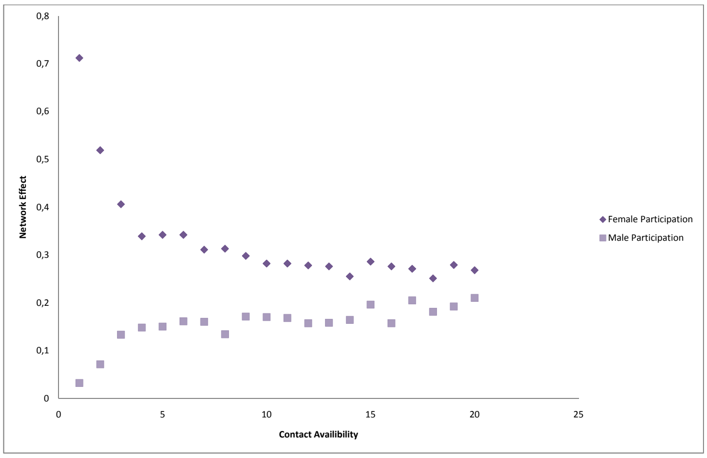
Dummies for language groups and areas are included as fixed effects. Absolute value of z-statistics in parentheses *, **, *** significant at the 10%, 5%, 1% level. Network refers to $Netw_{ial} = ln(CA_{al}) \times \overline{L}_{ial}$ as specified in the text. Variables used as instruments are: distributions of the sex of the first and the two oldest siblings of peers; the means of these instruments: age, age squared, gender, marital status; and the mean deviations of all explanatory variables.

Table 7: Non-linear peer effects

	Labour Market Participation	Employment
Network	0.184***	0.313***
	(35.10)	(39.07)
$Network^2$	-0.014***	-0.059***
	(8.82)	(14.14)
Gender*Network	0.019***	-0.039***
	(5.18)	(5.59)
$(Gender * Network)^2$	0.019***	-0.046***
	(6.33)	(10.32)
Constant	0.077***	-0.249***
	(14.90)	(42.58)
Observations	1695941	1695941
R-squared	0.19	0.26

*, **, *** significant at the 10%, 5% and 1% level respectively. Absolute value of t-statistics in parentheses.

Figure 1: The non-linearity of peer effects in labor-force participation



Contact availability is calculated by 20 five-percentile groups. Dummies for language groups and areas are included as fixed effects. All coefficients are significant at the 5% or 1% level. Net effects ($\alpha * \overline{lnCA}$) are presented. Standard errors are clustered at the group level.

Table 8: Peer Effects in Employment

Employment	(1)	(2)
Network	0.285***	0.263***
	(14.46)	(12.55)
Gender*Network		0.052***
		(3.99)
Gender	0.206***	0.175***
	(43.67)	(20.89)
Age	0.049***	0.050***
	(71.38)	(71.98)
Age^2	-0.039***	-0.039***
	(81.67)	(82.17)
Contact availability	-0.141***	-0.142***
	(16.63)	(16.71)
HH	0.069***	0.069***
	(29.47)	(29.64)
Primary Education	0.035***	0.035***
	(11.19)	(11.08)
Secondary Education	0.224***	0.223***
	(53.60)	(53.68)
Tertiary Education	0.406***	0.406***
	(38.71)	(38.72)
No. Children	0.005***	0.005***
	(6.91)	(6.72)
Urban	0.050***	0.050***
	(6.76)	(6.77)
Observations	1695941	1695941

Columns (1) and (2) present the marginal effects. Dummies for language groups and areas are included as fixed effects. Absolute value of z-statistics in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Contact availability refers to $ln(CA_{al})$ as specified in the text. Network refers to $Netw^p_{ial} = ln(CA_{al}) \times \overline{L}_{ipl}$ as specified in the text. HH refers to the number of working household members. Standard errors are clustered at the group level. Age^2 refers to $Age^2/100$.

Table 9: Peer effects in wages

Wage	(1)	(2)
Network	0.251***	0.259***
	(7.12)	(6.96)
Gender Effect		-0.013*
		(1.88)
Gender	0.511***	0.521***
	(59.23)	(33.64)
Age	0.120***	0.120***
	(55.51)	(55.73)
Age^2	-0.126***	-0.126***
	(53.83)	(53.99)
Contact Availability	-0.182***	-0.181***
	(10.63)	(10.58)
HH	0.050***	0.050***
	(13.29)	(13.26)
Primary Education	0.253***	0.253***
	(42.42)	(42.55)
Secondary Education	0.982***	0.982***
	(90.19)	(89.94)
Tertiary Education	1.654***	1.654***
	(53.80)	(53.96)
No. Children	0.007***	0.007***
	(3.67)	(3.65)
Urban	-0.235***	-0.234***
	(17.23)	(17.25)
Constant	2.962***	2.956***
	(35.69)	(35.64)
Observations	1679001	1679001

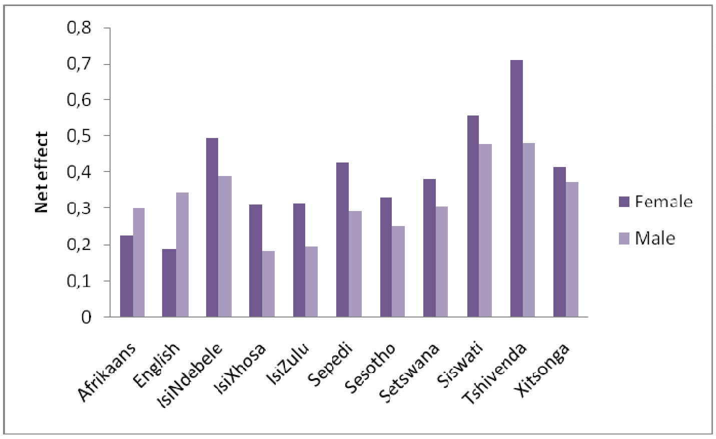
Dummies for language groups and areas are included as fixed effects. Absolute value of t-statistics in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Contact availability refers to $ln(CA_{al})$ as specified in the text. Network refers to $Netw^p_{ial} = ln(CA_{al}) \times \overline{L}_{ipl}$ as specified in the text. Wage refers to the log of the ratio of yearly income to weekly hours worked. Estimations are carried out using the Heckman selection model; the distribution of the sex of peers' sibling and a dummy variable for home ownership are included as an exclusion restrictions. Additional control variables are industry and type of job. Standard errors are clustered at the group level. Age^2 refers to $Age^2/100$.

Table 10: Robustness checks

		Labor Force Participation	Employment	Wage
1) Original sample	$Network^p_{ial}$	0.201***	0.263***	0.259***
		(14.68)	(12.55)	(23.19)
	Gender*$Network^p_{ial}$	-0.021***	0.052***	-0.013*
		(2.88)	(3.99)	(1.88)
2) IsiZulu speakers excluded	$Network^p_{ial}$	0.198***	0.258***	0.286***
		(12.70)	(11.37)	(24.51)
	Gender*$Network^p_{ial}$	-0.032***	0.037***	-0.029***
		(4.35)	(2.61)	(4.08)
2) Afrikaans speakers excluded	$Network^p_{ial}$	0.181***	0.242***	0.2111***
		(14.66)	(9.67)	(15.42)
	Gender*$Network^p_{ial}$	-0.013*	0.050***	0.004***
		(1.76)	(3.08)	(3.56)
2) English speakers excluded	$Network^p_{ial}$	0.193***	0.316***	2.777***
		(15.72)	(15.18)	(21.56)
	Gender*$Network^p_{ial}$	-0.022***	0.021*	-0.025***
		(3.67)	(1.86)	(3.09)
3) With children	$Network^p_{ial}$	0.215***	0.255***	0.148***
		(13.57)	(12.09)	(9.64)
	Gender*$Network^p_{ial}$	-0.018**	0.045***	-0.033***
		(2.00)	(3.38)	(3.65)
4) 25/40 years old	$Network^p_{ial}$	0.166***	0.265***	0.244***
		(33.60)	(42.36)	(15.35)
	Gender*$Network^p_{ial}$	-0.018***	0.020***	-0.051***
		(10.88)	(5.29)	(5.39)

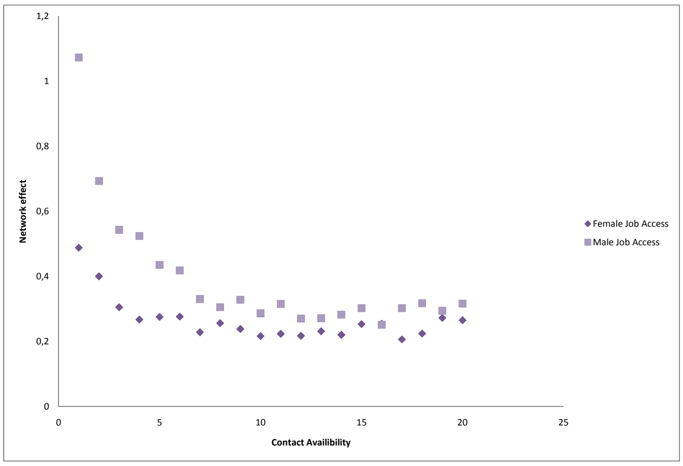
Absolute value of z and t statistics in parentheses for (a) and (b) respectively. * significant at 10%; ** significant at 5%; *** significant at 1%. Dummies for language groups and areas are included as fixed effects. Wage refers to the log of the ratio of yearly income to weekly hours worked. Standard errors are clustered at the group level. (a) Estimations are carried out using a probit model. (b) Estimations are carried out using the Heckman selection model.

Figure 2: Peer effects in labor-force participation by language and gender



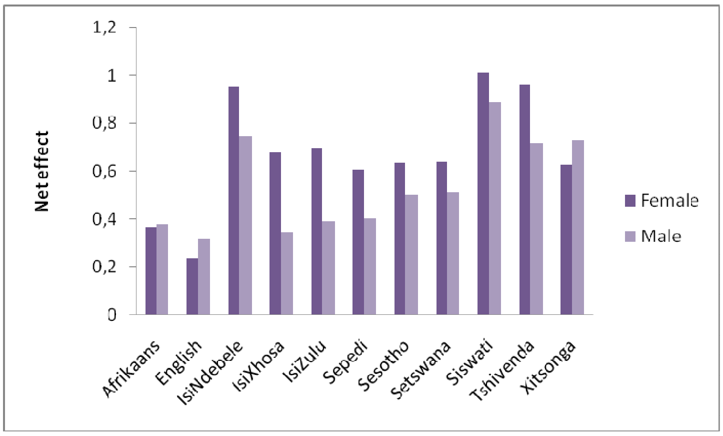
Dummies for language groups and areas are included as fixed effects. All coefficients are significant at the 5% or 1% level. Net effects ($\alpha * \overline{lnCA}$) are presented. Standard errors are clustered at the group level.

Figure 3: The non-linearity of peer effects in employment



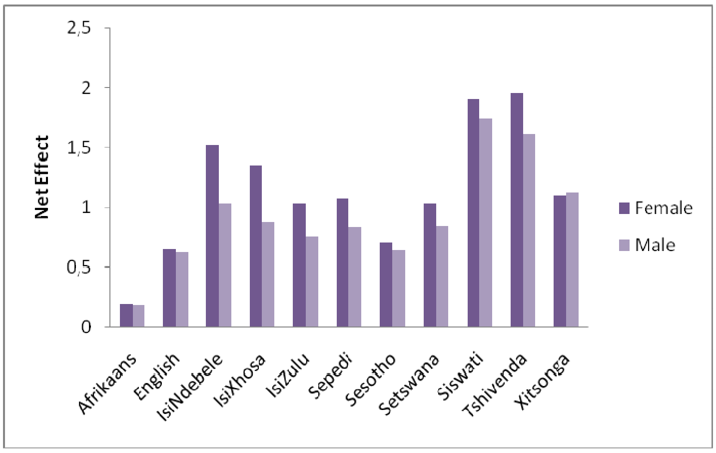
Contact availability is calculated by 20 five-percentile groups. Dummies for language groups and areas are included as fixed effects. All coefficients are significant at the 5% or 1% level. Net effects ($\alpha * \overline{lnCA}$) are presented. Standard errors are clustered at the group level.

Figure 4: Peer effects in employment by language and gender



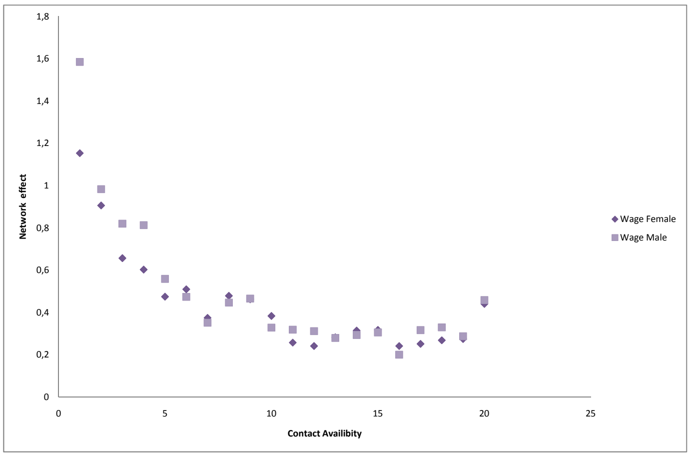
Dummies for language groups and areas are included as fixed effects. All coefficients are significant at the 5% or 1% level. Net effects ($\alpha * \overline{lnCA}$) are presented. Standard errors are clustered at the group level.

Figure 5: Peer effects in wages by language and gender



Dummies for language groups and areas are included as fixed effects. All coefficients are significant at the 5% or 1% level. Net effects ($\alpha * \overline{lnCA}$) are presented. Estimations are carried out using the Heckman selection model. Standard errors are clustered at the group level.

Figure 6: The non-linearity of peer effects in wages



Contact availability is calculated by 20 five-percentile groups. Dummies for language groups and areas are included as fixed effects. All coefficients are significant at the 5% or 1% level. Net effects ($\alpha * \overline{lnCA}$) are presented. Standard errors are clustered at the group level. Estimations are carried out using the Heckman selection model.

Appendix

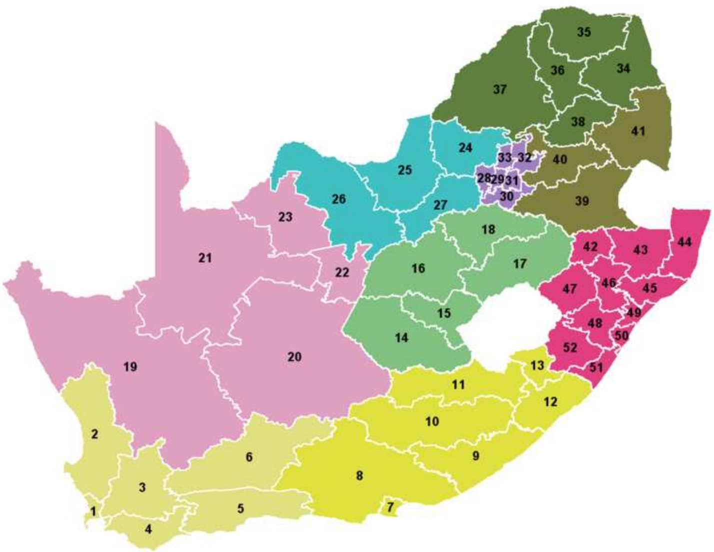
A Methods taken to find current job

Table 11: Methods taken to find current job

	%
A friend/relative (in a different household)	40.72
Saw an advert in the newspaper	11.86
I knocked on factory gates and visited	8.68
I went to a factory and waited for a job	7.85
A household member told me about the job	7.18
Saw an advert on a notice board	5.46
I asked someone who had employed me before	4.21
Through an employment agency	2.63
I waited on the side of the road	1.98
Contacted by employer	1.14
n.a	8.3

Source: South African National Income Dynamycs Study

B Districts and Provinces of South Africa



C Labor-Market Discrimination

Table 12: Labor-Market Discrimination

Heckman L1/L2	Employment	Repr.
Gender	0.083***	
	(40.74)	
English	0.013***	**1.73**
	(7.18)	
IsiNdebel	-0.108***	0.95
	(28.34)	
IsiXhosa	-0.177***	0.72
	(82.54)	
IsiZulu	-0.161***	0.76
	(74.32)	
Sepedi	-0.140***	0.89
	(57.31)	
Sesotho	-0.140***	0.98
	(59.44)	
Setswana	-0.116***	0.97
	(48.47)	
Siswati	-0.135***	0.94
	(34.68)	
Tshivenda	-0.113***	0.87
	(25.41)	
Xitsonga	-0.118***	0.89
	(42.12)	
Observations	1296909	
R-squared	0.69	

Speaking Afrikaans is omitted. Dummies for areas are included as fixed effects. Additional controls are age, age squared, education, number of children, urban dummy, household characteristics, lnCA, religion dummies, household type dummies and socio-professional category of wife or husband. Estimations are carried out using the Heckman selection model. Absolute value of t-statistics in parentheses. *, **, *** significant at the 10%, 5%, 1% level.

D Alternative to define the area of neighborhood

Table 13: Results

	Labour-Market Participation	Employment
Network	0.192***	0.245***
	(11.83)	(15.46)
Gender*Network	-0.052***	0.185***
	(11.15)	(27.30)
lnCA1	-0.214***	-0.196***
	(11.92)	(29.91)
HH	0.262***	0.181***
	(155.59)	(111.50)
Urban	-0.155***	0.021***
	(39.53)	(5.67)
Constant	-1.751***	-2.534***
	(87.75)	(129.94)
Observations	1695944	1695944
R-squared	0.21	0.36

Absolute value of z statistics in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Dummies for language groups and areas are included as fixed effects. Estimations are carried out using a probit model.

E Gender sub-sample

Table 14: Gender sub-sample results

	Labour Market Participation		Employment	
	Women	Men	Women	Men
Network	0.219***	0.114***	0.209***	0.319***
	(12.29)	(10.68)	(10.48)	(14.77)
lnCA1	-0.161***	-0.089***	-0.099***	-0.162***
	(11.96)	(11.16)	(11.72)	(17.24)
HH	0.087***	0.055***	0.068***	0.093***
	(32.69)	(37.83)	(27.01)	(37.62)
Children	-0.018***	-0.002***	-0.010***	0.012***
	(31.38)	(4.34)	(11.87)	(12.07)
Observations	906237	789704	906237	789704

Dummies for language groups and areas are included as fixed effects. Absolute value of z-statistics in parentheses *, **, *** significant at the 10%, 5%, 1% level. Network refers to $Netw_{ial} = ln(CA_{al}) \times \overline{L}_{ial}$ as specified in the text. Variables used as instruments are: distributions of the sex of the first and the two oldest siblings of peers; the means of these instruments: age, age squared, gender, marital status; and the mean deviations of all explanatory variables.

F Correlation between labor-market behaviour and the sex of oldest siblings

Table 15: Correlation

	Having same-sex children
N. of children	0.2250***
Labor-Market Participation	-0.1891***
Employment	-0.2062***
Peers having same-sex children	0.00849
Neighborhood	-0.0061

*** significant at the 1% level.