# Social Networks and Ethnic Niches: An Econometric Analysis of the Manufacturing Sector in South Africa

Andre Hofmeyr[1]

**Working Paper Number 83**

[1] School of Economics, University of Cape Town

# Social Networks and Ethnic Niches: An Econometric Analysis of the Manufacturing Sector in South Africa

Andre Hofmeyr*

June 18, 2008

**Abstract**

This paper analyses the link between social networks and ethnic occupational niches in the manufacturing sector in South Africa. To this end, it employs the methodology of Bertrand et al. (2000) to minimise the omitted variable bias induced by standard approaches investigating network effects and adopts Model's (1993) concentration index to define an ethnic niche. The results indicate that 25 percent of our sample is employed in ethnic niches in the manufacturing sector but that niche employment varies markedly by language group. With regards to the effect of social networks, increasing the quality or quantity of an individual's contacts by one standard deviation increases his probability of niche employment by 4 percent. Put differently, social networks magnify a policy shock affecting employment in ethnic niches by over 100 percent. This paper therefore highlights the importance of social networks, which channel workers into jobs that become ethnic niches, in the manufacturing sector in South Africa.

## 1 Introduction

A growing body of literature recognises the importance of social networks for labour market outcomes[1]. Networks matter because they facilitate the transmission of job related information among individuals. This paper adopts an econometric approach to analyse the impact of social networks on a particular feature of modern labour markets: ethnic occupational niches.

An ethnic occupational niche is the concentration and specialisation of members of an ethnic group in a particular occupational activity. According to theory, these niches arise because of the members' ability to supply labour through social networks and due to the special skills, experiences and other attributes they possess which employers consider relevant when hiring job applicants (see Waldinger, 1996; Elliott, 2001; Wilson, 2000). Although a large literature documents the existence of ethnic niches and the importance of social networks in channelling individuals into these occupations, few studies have investigated the impact of social networks on the probability of niche employment using large sector-wide datasets.

Social network analyses typically rely on detailed information about individuals' contacts collected through surveys. However given the limited scope of this data and the complete absence of it in many cases, scholars have devised strategies to proxy for social networks given the labour market data at their disposal. The most common method defines an individual's social network by the neighbourhood he inhabits and numerous studies have shown that individual outcomes are indeed correlated with the outcomes of the individual's neighbours[2]. Another approach uses the ethnic group of the individual to proxy for this social network (see Borjas, 1992, 1995). Analyses using this

*School of Economics, University of Cape Town

[1] For a useful review of the literature on social networks and labour market outcomes consult Ioannides and Loury (2004).

[2] Consult Jencks and Mayer (1990) for a review of the literature.

framework have also found correlations between the outcomes of the ethnic group and the outcomes of the individuals that comprise it. While suggestive of network effects, the results of these studies should be treated with caution because they are likely influenced by unobserved characteristics of individuals, neighbourhoods and ethnic groups.

To minimise the bias induced by these strategies, this paper adopts the approach of Bertrand et al. (2000) to define an individual's social network. Given that individuals tend to have homophilous social ties, we use language group to proxy for the links between individuals within a neighbourhood[3]. This defines the "quantity" dimension of our measure of social networks: individuals who live in an area with a high proportion of other individuals who speak their language will have a larger pool of potential contacts who they can rely on to find employment. However, contacts that are actually employed in a particular occupation are likely to exert a larger influence on an individual's probability of finding employment in that occupation. Thus, the number of contacts that an individual has who are employed in an ethnic niche in his area provides a measure of network "quality". We investigate whether the "quantity" and "quality" of contacts that an individual has, influences his probability of securing employment in an ethnic niche.

This paper uses the 10 percent sample of the 2001 Census survey conducted by Statistics South Africa. We use a two-part strategy to define occupations in specific areas as niche or non-niche. We then use a linear probability model to investigate whether an individual's social network influences the likelihood of niche employment. Our estimation strategy controls for many of the common omitted variable biases that have plagued previous studies. Specifically, we include language group, area and occupation fixed effects in our regression model to control for these confounds. Our results suggest that social networks have a large and significant effect on the probability of niche employment.

## 2 Social Networks and Ethnic Niches: A Review of the Literature

The importance of social networks for labour market outcomes has been increasingly recognised by the economics discipline. Social networks matter because they facilitate the transmission of job-related information between individuals. Unlike standard approaches to modelling job-search behaviour, which treats individuals like Robinson Crusoe isolates making decisions on a one-to-one basis, social network models incorporate the information spillovers and complex interactions between individuals that are prevalent in contemporary labour markets.

An empirical regularity motivating this change in emphasis, which has been identified in both the economic and sociological literature from as early as 1960, is that on average 50-60 percent of workers obtain their jobs through personal contacts (Rees, 1960; Granovetter, 1995; Holzer, 1987; Staiger, 1990; Montgomery, 1991, Topa, 2001). Furthermore, 40-50 percent of employers use their employees' social contacts to fill job openings (Holzer, 1987). Informal recruitment methods have also been identified as improving the employer-employee match: individuals recruited through personal contacts are less likely to quit (Datcher, 1983; Devine & Keifer, 1991) and have longer tenure on these jobs (Simon & Warner, 1992).

Theoretically, social networks matter for employment outcomes because of the functions that

[3] We use language as our proxy, rather than other measures like race, gender or religion, because we think it plausible that job-related information flows more quickly and effectively through individuals who speak a common language. Bakalian (1993) shows that friendship groups tend to sort across ethno-linguistic lines and Alba (1990) asserts that one's home language is a crucial determinant of ethnic identity. Furthermore we think it reasonable that individuals typically spend more time with, and thus acquire information from, individuals who live in the same locality and speak the same language. Finally, in South Africa, markers like race, religion or gender are too broad to accurately define an individual's social network because of the differences that exist between individuals within these categories. For example, 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.

personal contacts play in labour markets (Elliott, 2001). These functions are twofold. Personal contacts can provide timely information about employment opportunities that are not widely or publicly known. Furthermore, contacts can pass on information to employers about a potential employee that increases his likelihood of being hired.

While the benefits to job-seekers from using social networks are quite obvious, the benefits to employers are less so. After all, if employers can fill job vacancies through formal means then what advantages do social networks confer? Fernandez et al (2000) list five mechanisms that make hiring through social networks attractive. First, employers can enlarge the pool of applicants by drawing on the referrals of incumbents. Second, owing to the homophilous nature of incumbents' ties (Granovetter, 1995; Rees & Schults, 1970; Ullman, 1966), the employer who has already screened the quality of his employees will be more disposed to hiring referrals because they are likely to be of a similar quality. In this sense, incumbents' referrals reduce the problems associated with information asymmetries. Third, employees - for fear of damaging their reputations - will only refer qualified candidates, which reduces the costs of screening applicants. Fernandez et al. (2000) refer to these three mechanisms collectively as the "richer pool" argument: insider referrals provide a larger and better pool of applicants. The fourth mechanism that makes hiring through informal means beneficial to the employer, is that referrals are likely to provide a better match between employer and employee. Not only do incumbents pass on valuable information to employers about the individuals that they have referred, which is often difficult to obtain through formal recruitment procedures, but they also provide tacit information to candidates about the job for which they are applying. Finally, the benefits of referral hiring are present even after the job has been filled. The idea is that incumbents who referred an applicant will help the newcomer get acquainted with the organisation and even provide him with some on-the-job training, thereby boosting his productivity. In sum, there are numerous benefits to employers and employees from using social networks in the hiring process.

While a large empirical literature - from which the stylised facts presented above were drawn - regarding social networks and labour market outcomes exists, analyses of the impact of social networks on job search and employment in South Africa are somewhat limited. Most of the work that has been done defines the social network as the number of household members that are employed (see Wittenberg & Pearce, 1996; Mlatsheni and Rospabe, 1999; Schoer, 2005). Wittenberg and Pearce (1996) find that this network positively affects individuals' access to jobs, while Mlatsheni and Rospabe (1999) find that it increases the probability of youth being in wage employment. These studies, however, suffer from omitted variable bias which makes their estimates unreliable, a point recognised by Godlonton and Burns (2006). For their analysis, they instead adopt the approach of Bertrand et al (2000) and define the social network in terms of age, language group and geographical proximity. Their estimation strategy also includes geographic area and language group fixed effects to minimise the potential for omitted variable bias. Their findings suggest that social networks may increase employment probabilities by an estimated 3-9 percent.

The preceding discussion highlights the importance of social networks on labour market outcomes. Given the widespread unemployment and stark income inequality in South Africa, it is crucial that we understand the impact of these networks on individuals' employment probabilities. This paper follows the approach of Bertrand et al. (2000) to analyse the extent to which social networks in South Africa create ethnic niches in certain occupations. This is relevant to the labour market situation in this country in so far as ethnic niches promote or hinder upward social mobility.

Up to this point, the term "ethnic occupational niche" has been used glibly to refer to co-ethnic concentrated workplaces. However, there is widespread confusion in the literature regarding distinctions between the terms "ethnic economy", "ethnic enclave" and "ethnic niche", which are artefacts of different disciplines. Before discussing the role that social networks play in generating and maintaining co-ethnic concentrated workplaces, it is therefore essential that we clarify these terms so that they can be applied appropriately in the ensuing discussion.

Ethnic niches are typically associated with the concentration and specialisation of members of an ethnic group in particular industrial/occupational activities. These niches arise because of the members' ability to supply labour through social networks and due to the special skills, experiences and other attributes they possess which employers consider relevant when hiring job applicants (Wilson, 2000). Ethnic niches are an essential component of the other forms of social formation identified above.

Ethnic economy refers to the concentration of co-ethnic owners and workers in one or more related industries, the sum of which is an ethnic economy (Logan, Alba & McNulty, 1994). By this view, the ethnic economy exists partially independently from the general economy and provides an alternative source of employment. Thus ethnic niches form part of the ethnic economy.

An ethnic enclave is essentially an ethnic economy but one in which an ethnic group specialises in the production of particular goods and there is spatial concentration (locational clustering) of ethnic enterprises (Wilson, 2000).

While there are important differences between these terms, the more general one, ethnic niche, will be used henceforth to describe an employment sector in which individuals from a specific ethnic group are concentrated above a level one would expect based on their share of the total labour force of a local labour market[4] (Wilson, 1997).

Ethnic niches are formed, at least theoretically, through social networks (Morales, 2004). As Waldinger (1996) argues, to attain employment in an ethnic niche, informal social networks are particularly important. Given that individuals tend to have homophilous social ties[5], information regarding job opportunities is more likely to flow to individuals of the same ethnic group as incumbents. This creates an insider-outsider dynamic, that becomes self-reproducing: as more individuals from a particular ethnic group are employed in an organisation, information about employment opportunities is passed on to ethnically defined insiders whilst outsiders are excluded. Put simply, ethnically segregated social networks lead to ethnically segregated workplaces. This process is characterised by path dependence and can lead to occupational closure[6] (see Waldinger, 1996).

Discrimination in the general labour market is often the precursor to the formation of niches in local labour markets because in the latter, discriminatory barriers are relatively low or other ethnic groups are not overly represented (Granovetter, 1995; Sassen, 1995). This is closely linked to the notion of the job queue, first expounded by Thurow (1969, 1972, 1975). According to the theory, individuals in the labour market form part of a job queue, with the most highly qualified individuals at the top and the less qualified lower down. However, in ethnically diverse economies, like South Africa, groups of individuals are ranked according to employers' perceptions of the productivity of the group, with the human capital of an individual serving as another mediating factor. Thus, at each level of the skill hierarchy, members of the core cultural group receive preferential treatment, relegating other ethnic groups to the periphery of the labour market (Morales, 2004). Forced to accept whatever jobs are available once groups higher up in the queue have made their selections, marginalised ethnic groups form niches[7].

Faced with discrimination in the general labour market, niches can provide a "protected environment" for co-ethnics to acquire skills and experience and receive equitable compensation (Waldinger, 1996: 95). However, niches can also trap ethnic groups in exploitative relationships where their possibility for upward mobility is circumscribed (Bonacich, 1988). Granovetter (1973) argues that individuals with weak ties are more likely to receive information about employment opportunities

[4] Refer to the discussion of the concentration index in section 4 of this paper.

[5] Buhai and van der Leij (2006) refer to this phenomenon as an in-breeding bias.

[6] Occupational closure refers to the situation where access to employment opportunities in a niche is restricted for individuals from other ethnic groups because the incumbent ethnic group has managed to establish informal regulatory mechanisms and procedures to protect the niche against encroachment from other groups. The idea is that employers and employees in a niche come to an, often tacit, agreement about hiring practices and promotional rules because of the benefits to both parties from using insider referrals (discussed above). Consequently, individuals from other ethnic groups typically face more exacting job entry requirements than co-ethnic insiders and internal promotion becomes the norm. As a result, the occupational niche effectively becomes closed to other ethnic groups.

[7] While the formation of a niche is typically characterised by self-selection, niches also arise 'spontaneously' when members of a particular ethnic group are concentrated in the 'residual' jobs of the labour market (Wilson, 1997).

from outside of their strong tie social ambit. By this view, ethnic niches - and in particular, ethnic enclaves - limit individuals' ability to access information outside of their ethnic group which may hinder their ability to secure jobs in the general labour market[8]. Regardless of their long-term effects, once an ethnic group gains a foothold in a particular occupation, the process of selective recruitment, described above, then increases its representation in the occupation while excluding other groups.

Despite their tendency towards occupational closure, ethnic niches are not immutable. Once established, a niche associated with one ethnic group may become associated with another. Waldinger (1996) identifies two scenarios through which this can occur. In the succession scenario, economic expansion increases the demand for ethnic groups further down the labour queue. As individuals from these groups move into the general labour market, ethnic groups still further down the queue may form niches in the previously restricted occupations. In the leapfrogging scenario, the quality of low-ranked ethnic groups may improve (for example, educational attainment increases) which raises the demand for these groups and re-orders the labour queue, assigning other groups to the niches previously dominated by the now more prized groups. Whatever the mechanism, it is important to note that niches are not fixed entities and are prone to change over time.

The theory of ethnic niches is predominantly sociological and descriptive. Buhai and van der Leij (2006) are the first scholars to apply social network theory to dynamically model occupational segregation in the labour market. They construct a simple three-stage model of occupational segregation with two homogenous social groups applying for two different jobs. In the first stage, individuals strategically decide to acquire one of two specialised educations. In the second stage, individuals randomly form friendship ties, with the probability of forming a tie decreasing with social distance. In the final stage, individuals use their personal contacts to search for jobs. The model predicts complete occupational segregation when social networks are important for channelling workers into jobs and when these networks are formed assortatively.

Buhai and van der Leij (2006) then extend the model to incorporate differences in wages between the two groups. To do so they assume that workers choosing one of the specialised educations (who then apply for the occupation corresponding to that education) earn a higher wage than workers choosing the other specialised education. They show that when the wage differential between occupations is small, complete segregation occurs. However, when the wage differential is large, partial segregation occurs with one group specialising in one occupation and the other group having workers in both occupations. The intuition for this result is that when the wage differential is great enough, individuals whose group is specialised in the relatively low paying occupation will have an incentive to acquire the other specialised education even though his social network will not be as helpful in finding him a job in this occupation. Thus the wage differential offsets the benefit of using social networks to acquire jobs. Interestingly, when a wage differential is introduced to the model, unemployment in both groups and inequality between groups obtains. The crucial point of their analysis though is that assortatively-formed social networks channel workers into jobs which can become niches.

There is now a growing body of literature, albeit primarily sociological, that analyses the formation and persistence of ethnic niches (see Lieberson, 1980; Morawska, 1990; Model, 1993; Logan, Alba & McNulty, 1994; Model & Ladipo, 1996; Waldinger, 1996; Wilson, 1997, 2000; Morales, 2004; Elliott, 2001). For example, Lichter (2000) finds that the use of personal contacts in job-search, channels co-ethnics into ethnic niches, which often become enclaves. Light & Gold (2000) find that once an ethnic niche has been established, incumbents exert enormous influence over the hiring of job-applicants and typically direct employment to their co-ethnics. Although rich in qualitative information, most of these analyses are localised and specific to particular sectors or occupations in developed countries. This paper takes a more encompassing approach and uses econometric techniques to test for the presence of ethnic niches in the manufacturing sector of the South African

[8]This is an important line of future research but one that is paper does not address due to data limitations.

labour market.

# 3 Methodology

This paper adopts the approach of Bertrand et al. (2000) to analyse the impact of social networks on the employment of individuals in ethnic niches. Social networks affect individual behaviour primarily through two channels: information and norms (Bertrand et al., 2000). With regards to ethnic niches, social networks can facilitate the transmission of job-related information to individuals of a specific ethnic group. In addition, the social norm channel may affect individual's preferences and thereby increase the probability of individuals applying for jobs in which their ethnic group predominates. While these channels may differentially affect employment in ethnic niches, the data at our disposal prevents us from analysing them separately. Instead we consider the ethnic group's niche participation rate as a measure of their niche 'culture', whilst acknowledging that this 'culture' is shaped both by information and norms. The extent to which an ethnic group's behaviour influences the behaviour of an individual in that group, is the social network effect that this paper investigates.

To explain how we measure the effect of social networks, assume that the true model governing employment in an ethnic niche is the following:

$$\Pr(Empl_{iljk}) = Netw_{iljk}\alpha^* + X_i^*\beta^* + Y_l^*\gamma^* + Z_j^*\delta^* + W_k^*\phi^* + \varepsilon_{iljk} \tag{1}$$

where $i$ indexes individuals, $l$ indexes language groups, $j$ indexes areas and $k$ indexes occupations. $Empl_{iljk}$ is a dummy indicating employment in niche $k$[9], $Netw_{iljk}$ measures the information and social norms of the individual's contacts, $X_i^*$ are observed and unobserved personal characteristics, $Y_l^*$ are observed and unobserved language group characteristics, $Z_j^*$are observed and unobserved local area characteristics, $W_k^*$ are observed and unobserved occupation characteristics, and $\varepsilon_{iljk}$ is an error term.

The difficulty in estimating this specification is that data for the variable $Netw_{iljk}$ is seldom collected. Ideally one would want data on individuals' actual contacts and the extent of their social networks. In the absence of such data, scholars have typically used the mean characteristics of an individual's locality as a proxy for their social networks. In doing so there is an implicit assumption that individuals are randomly distributed within the neighbourhood. According to this framework one would estimate:

$$\Pr(Empl_{iljk}) = \overline{Empl}_{jk}\alpha + X_i\beta + \varepsilon_{iljk} \tag{2}$$

where $\overline{Empl}_{jk}$ represents mean neighbourhood employment in occupation $k$ and $X_i$ are observed individual characteristics. Although a large body of empirical research validates the notion that individual outcomes are strongly correlated with mean neighbourhood characteristics (see Jencks and Mayer, 1990, for a review of the literature), the problem with this approach is that it suffers from Manski's (1993) "reflection problem".

Manski (1993) identifies three channels through which individual behaviour is affected: 1) *endogenous interactions*, where the behaviour of an individual is affected by the behaviour of the group; 2) *contextual interactions*, where an individual's behaviour is affected by exogenous characteristics of the group; and 3) *correlated effects*, where individuals in the same group behave in a similar manner because they share similar characteristics or face similar institutional environments. The "reflection problem" arises because data typically does not allow one to readily distinguish between these effects, making causal inference misguided at best, vacuous at worst. The difficulty is that mean

[9]Specifically, $Empl_{iljk}$ equals one if individual$i$ from language group $l$, living in area $j$, is employed in occupation $k$ where occupation $k$ is an ethnic niche dominated by $i$'s language group. $Empl_{iljk}$ equals zero if occupation $k$ is not dominated by the language group of individual $i$. It should be noted that this measure is restricted to employed individuals. Thus, $Empl_{iljk}$ draws a distinction between those individuals employed in niches and those not employed in niches without reference to unemployed individuals.

behaviour in a group is affected by the behaviour of the individuals comprising the group. Thus one cannot determine whether group behaviour actually affects individual behaviour or whether group behaviour is simply the aggregation of each individual's behaviour.

The "reflection problem" can be viewed as the outcome of three related omitted variable biases (Bertrand et al., 2000). First, omitted personal characteristics may be associated with $\overline{Empl}_{jk}$. For example, individuals that live in areas with widespread unemployment may be less ambitious. Second, omitted neighbourhood characteristics may be correlated with $\overline{Empl}_{jk}$. For example, areas with a training centre for occupation $k$ may raise an individual's probability of being employed in this occupation and therefore raise the mean employment in occupation $k$ in the area. Furthermore, the specification suffers from a simultaneity problem in that any policy or shock which affects mean neighbourhood employment will lead to a positive estimate of $\alpha$, regardless of whether social networks are in operation. Finally, omitted occupational characteristics may be associated with $\overline{Empl}_{jk}$. For example, differences in skill requirements for particular occupations may make certain neighbourhoods (say, ones with higher average skill levels) more likely to dominate the employment of these occupations, regardless of whether social networks in fact facilitate acquisition of these jobs. Furthermore, industrial agglomeration may make the supply of certain occupations greater in particular areas as compared to others. As these biases are all likely positive, a positive estimate of $\alpha$ does not necessarily imply the existence of social networks.

Another approach, pioneered by Borjas (1992, 1995), uses ethnic groups - rather than geographical proximity - to proxy for social networks. As individuals tend to have homophilous social ties, one would expect ethnicity to be an important determinant of social networks. In addition, this approach focuses on the effect of the previous generation's outcomes on the current generation's outcomes. Therefore, the mean outcomes of the ethnic group in the previous generation is used to construct $Netw_{iljk}$. In the context of ethnic niches, one could estimate the following equation:

$$\Pr(Empl_{iljk}) = \overline{Empl}_{(-1)lk}\alpha + X_i\beta + Y_l\gamma + \varepsilon_{iljk} \tag{3}$$

where $\overline{Empl}_{(-1)lk}$is the mean employment of ethnic group $l$ in occupation $k$ in the previous generation, and $Y_l$ are observed language group characteristics.

However, this approach also suffers from three omitted variable biases. First, omitted personal characteristics may be associated with $\overline{Empl}_{(-1)lk}$. Second, omitted ethnic group characteristics may be correlated with $\overline{Empl}_{(-1)lk}$. For example, high levels of discrimination may preclude certain language groups from obtaining employment in particular occupations, thereby concentrating their employment in others. This would show up as a positive estimate of $\alpha$ without directly capturing the effect of social networks. Finally, omitted occupational characteristics may be correlated with $\overline{Empl}_{(-1)lk}$.

This paper exploits both geographic and ethnic variation to construct a measure of social networks. Specifically, we use an individual's language group and magisterial district to proxy for his social network. In addition, we include fixed effects for language groups, magisterial districts and occupations in our regression framework to minimise the problems of omitted neighbourhood, ethnic group and occupational characteristics that have plagued previous studies.

We construct $Netw_{iljk}$ using the number of people an individual interacts with as well as the knowledge and attitudes of those people with respect to ethnic niches. Our measure therefore includes both a "quantity" and "quality" dimension of contacts. Assuming individuals primarily interact with individuals of their language group, we therefore define:

$$Netw_{ljk} \approx \begin{pmatrix} density\, of\, language \\ group\, l\;\; in\, area\, j \end{pmatrix}_{lj} \times \begin{pmatrix} ethnic\, niche\, knowledge \\ and\, attitudes\, of\, others \\ from\, language\, group \\ l\, who\, live\, in\, area\, j \end{pmatrix}_{ljk}$$

The density of language group $l$ in area $j$ is a measure of contact availability, denoted by $CA_{lj}$[10]. $CA_{lj}$ is thus our "quantity" measure. The second term in the construction of our network measure suggests that we should proxy for the "quality" of an individual's contacts with the mean employment in area $j$ of language group $l$ in occupation $k$(excluding individual $i$), which we refer to as $\overline{Empl}_{(-i)ljk}$. However, using $\overline{Empl}_{(-i)ljk}$ can introduce another source of omitted variable bias because it may reflect unobserved characteristics that an individual shares with members of his language group living in his area (Manski's *correlated effect*). Consequently, we replace $\overline{Empl}_{(-i)ljk}$with the mean employment of language group $l$ in occupation $k$, $\overline{Empl}_{lk}$.

We therefore estimate:

$$\Pr(Empl_{iljk}) = \left(CA_{lj} * \overline{Empl}_{lk}\right)\alpha + X_i\beta + \gamma_l + \delta_j + \phi_k + CA_{lj}\theta + \varepsilon_{iljk} \tag{4}$$

where $\gamma_l$, $\delta_j$ and $\phi_k$ are fixed effects for language groups, magisterial districts and occupations, respectively. As discussed above, $CA_{lj}$ is a measure of the "quantity" of contacts available, whereas $\overline{Empl}_{lk}$ is a measure of the "quality" of contacts. The interaction of these terms is used to proxy for an individual's social network: they provide a measure of how social networks influence individual behaviour. A positive and significant estimate of $\alpha$ provides evidence of network effects. We also include $CA_{lj}$as a control variable but omit $\overline{Empl}_{lk}$ because the language group fixed effects $\gamma_l$incorporate it.

This approach allows us to control for many of the common omitted variable biases prevalent in studies of social networks (Bertrand et al., 2000). First, magisterial district fixed effects capture differences in local areas, such as the extent of ethnic niching. Second, language group fixed effects capture omitted differences in language groups, such as the extent of discrimination they face. Third, occupation fixed effects capture omitted differences in occupations such as varying skill requirements. Finally, by including $CA_{lj}$ as a covariate, we control for omitted personal characteristics that may be associated with $CA_{lj}$. For example, an unobserved personal characteristic such as ambition may reduce the likelihood of an individual being employed in an ethnic niche as well as living among his own language group. This would affect the estimate of $\theta$ but would not influence $\alpha$.

Despite these controls, another potential source of omitted variable bias is still present: omitted personal characteristics that are correlated with the network term $\left(CA_{lj} * \overline{Empl}_{lk}\right)$. Including $CA_{lj}$ captures fixed differences between individuals that choose to live among their own language group and those that do not. However, these differences may vary by language group: individuals in a specific language group may differentially self-select away from their own language group as compared to individuals in other language groups. For example, living away from your language group may indicate that you have managed to break into the general labour market if you are from a language group that is heavily concentrated in ethnic niches. Alternatively, selecting away from your language group, if it is one that is not disproportionately represented in ethnic niches, may signal the reverse. This would affect the estimate of $\alpha$, thereby biasing the results. To control for this problem we include a variable which indicates whether a person relocated between the 1996 and 2001 Census. Doing so should minimise the bias induced from individuals' differential self-selection away from their language groups.

# 4 Data

This paper uses the 10 percent sample of the 2001 Census survey conducted by Statistics South Africa. Since our focus is on the economically active population we exclude all individuals younger

[10] Note that this measure refers to potential contacts and not actual contacts because the data lacks social network information. It should also be noted that this measure includes both employed and unemployed individuals because one's contact availability is defined simply by the "quantity" of potential contacts. The other term in our network measure captures the "quality" dimension and thus only includes employed individuals. This is discussed more thoroughly in the next section.

than 15 and older than 65. In addition, we exclude all non-housing units[11]. We use magisterial district[12] as our geographic indicator and define ethnicity by language group[13]. The 2001 Census uses the South Africa Standard Classification of Occupations (SASCO) down to the three-digit level. We classify occupations at the two-digit level[14] to allow for enough occupational heterogeneity in the sample and to ensure that our niche measures are not meaningless[15].

As discussed previously, we measure the size of social networks by contact availability. $CA_{lj}$ is the proportion of individuals from language group $l$ in area $j$ divided by the proportion of individuals from that language group in South Africa[16]. The contact availability measure is therefore defined as:

$$\left(\frac{C_{lj}/A_j}{N_l/T}\right)$$

where $C_{lj}$ is the number of people from language group $l$ in area j, $A_j$ is the total number of people in area $j$, $N_l$ is the total number of people who belong to language group $l$, and $T$ is the total number of people in the country.

We use the full sample to construct our measure of contact availability, following Bertrand et al. (2006), but limit the sample to the manufacturing sector for our regressions[17]. This is done for two primary reasons. First, the manufacturing sector is the second largest contributor to South Africa's GDP (Statistics South Africa, 2006) and it employs 13% of South Africa's workforce. It is therefore

[11]Specifically, all prisons, residential hotels, student residences, homes for the aged, workers' hostels, tourist hotel/motels, hospital/medical facility/clinics, childcare institutions, homes for the disabled, boarding school hostels, initiation schools, convents/monasteries/religious retreats, defence force barracks, prisons/correctional institutions, community or church halls, refugee camps, and homeless shelters are excluded. This is done so that measures which rely on household size are not biased by the, often, huge number of 'household' members in these non-housing units. For example, the Census records the number of household members of a prisoner as the number of individuals in the prison.

[12]There are 354 magisterial districts in the sample which vary markedly in size from 85 to 35,334 individuals. A lower level of disaggregation was precluded by lack of information. However, the relatively high level of geographical aggregation should not pose a problem for our results because all of our measures are scaled such that small magisterial districts are not underweighted in our regressions.

[13]There are 11 official languages in South Africa. Data on all other languages spoken in the country is lumped into an "other" category, which is excluded from the analysis because more detailed information would be required to proxy for these language group's social networks.

[14]To illustrate the difference between the levels at which occupations are classified, the following example is presented. A one-digit classification of an occupation is "Elementary occupations". The two-digit level breaks this classification down further into "Sales and services elementary occupations", "Agricultural, fishery and related labourers" and "Mining, construction, manufacturing and transport labourers". The three-digit level breaks each of these categories down further. For example, the two-digit level classification of "Agricultural, fishery and related labourers" at the three-digit level is "Agricultural, fishery and related labourers NFD", "Agricultural, fishery and related labourers" and "Agricultural, fishery and related labourers NEC".

[15]Elaboration on this point is clearly required. If we were to classify occupations at the one digit level this would provide too broad a measure to classify occupations as ethnic niches because individuals are matched to jobs not broad occupational categories (Elliott, 2001). As discussed later, we limit the sample to the manufacturing sector. Consequently, classifying occupations at the three digit level would be vacuous because in numerous areas only one individual from one language group is employed in a particular occupation. This would make the concentration index for those individuals approach infinity, thereby defining a niche as the employment of one individual from one language group in one area.

[16]Bertrand et al. (2000) typically use the log of this ratio in their calculations. Their rationale for doing so is that this prevents small magisterial districts from being underweighted in the regressions. However, as the CA measure is constructed by dividing the ratio of individuals of a particular language in a particular area by the ratio of the total number of individuals in this language group in the country, the measure is scale free. Thus, we use the unlogged version of CA but test the robustness of our results to different choices of this measure in section 6.

[17]In doing so, we implicitly assume that individuals from one's language group living in one's area can provide information on employment prospects in the manufacturing sector even if they are not employed in this sector. This assumption is reasonable if social networks facilitate the transmission of job related information outside the confines of an individual's employment sector. For example, assume that person X does not know person Y. However, the father of person X (who is not employed in manufacturing) is told by person Y (who is employed in manufacturing) that the company he works for is looking to hire individuals. As a result, person X becomes aware of a job opening even though his direct social contacts are not employed in manufacturing.

broadly representative of the South African economy. Second, focussing on the manufacturing sector allows us to analyse ethnic occupational niches at a level of disaggregation that would be infeasible using the full sample due to the computational demands involved.

Our dependent variable ($Empl_{iljk}$), which we refer to as "employment in a niche", is dichotomous and equals one if individual $i$ from language group $l$, living in magisterial district $j$ is employed in occupation $k$ where occupation $k$ is dominated by the language group of individual $i$in area $j$.This implies that the individual is employed in his language group's niche. It equals zero if occupation $k$ is dominated by a language group other than individual $i$'s language group. In other words, "employment in a niche" equals one only if individual $i$'s language group is over-represented in occupation $k$ in area $j$. To bring clarity to the terms 'dominated' and 'over-represented', we define an ethnic niche using a concentration index:

$$CI_{ljk} = \frac{e_{ljk}/o_{ljk}}{e_{lj}/o_{lj}}$$

where $CI_{ljk}$ is the concentration index for language group $l$ living in area $j$ and employed in occupation k; $e_{ljk}$ is the total number of individuals from language group $l$ living in area $j$ employed in occupation $k$; $e_{lj}$ is the total number of individuals of language group $l$ living in area $j$; $o_{ljk}$ is the total number of individuals from other language groups living in area $j$ and employed in occupation $k$; and $o_{lj}$ is the total number of individuals from other language groups living in area $j$(Wilson, 2000).

Model (1993) originally developed this approach for classifying occupations as niches and argued that an ethnic niche exists if the concentration index for language group $l$living in area $j$ employed in occupation $k$ is greater than 1.5. This approach was later adopted by Waldinger (1996) and Wilson (2000), with the latter scholar adding the criterion that at least 20 workers be employed in this occupation in the sample for it to constitute a niche. The concentration index is an odds ratio which implies that, for a value of 1.5, language group $l$ is 1.5 times more likely to be concentrated in occupation $k$ than all other language groups in the area. As Wilson (2000) notes, the choice of 1.5 is arbitrary but it sets a lower limit for the extent to which an ethnic group is concentrated in an occupation.

We adopt a two-part strategy for defining an ethnic niche. First, if the concentration index for at least 20 workers in language group $l$in area $j$employed in occupation $k$is greater than 1.5 we provisionally define the language group as occupying a niche. Second, we compare the concentration indices of all language groups whose values exceed 1.5 in area $j$ and occupation $k$ and define an ethnic niche according to which group has the highest concentration index. This ensures that the language group that we classify as occupying a niche unambiguously dominates the employment of a particular occupation because it is more concentrated than all other language groups. Only once a language group living in an area and employed in an occupation passes the two-part test, is the occupation classified as an ethnic niche and the dependent variable assigned a value of one.

## 4.1 Summary Statistics

Table I presents summary statistics for the main variables used in our analysis. Of particular interest is that 25 percent of our sample is employed in ethnic niches in the manufacturing sector. Although quite high, this value falls well within the range of other studies (see Waldinger, 1996; Wilson, 2000; Elliott, 2001).

As is evident, the manufacturing sector is dominated by males in their late thirties and is highly concentrated in urban areas. We find that black South Africans are under-represented in manufacturing relative to their share of the total population whereas white, coloured and Asian/Indian South Africans are over-represented in this sector[18]. This is clearly a legacy of Apartheid-era em-

[18] Black individuals constitute 79.24 percent, coloured individuals 9.18 percent, Asian/Indian individuals 2.52 percent, and white individuals 9.06 percent of South Africa's total population.

ployment and education policies that gave preferential treatment to the population groups that are over-represented in manufacturing.

Table II presents selected summary statistics for the eleven language groups used in our analysis. We find that employment in a niche varies markedly by language group. At the one extreme, 43 percent of English speakers and 32 percent of IsiZulu speakers are employed in occupational niches. At the other, only 4 percent of Xitsonga speakers and zero IsiNdebele speakers are employed in niches.

Approximately one fifth of the individuals in our sample relocated between the 1996 and 2001 census but this too varies by language group. It is therefore important to use this variable in our regressions to control for the differential self-selection of individuals away from their language groups.

The table indicates that English and Afrikaans speakers dominate the employment of the manufacturing sector. Furthermore, these two language groups have the highest levels of education, on average, as well as the highest fraction of other adult household members employed in manufacturing[19].

Table III presents detailed information on the extent of ethnic niching of each language group in each occupation. Each cell in the table represents the fraction of individuals from language group $l$ employed in a niche in occupation $k$. We find that of all occupational categories, "Corporate managers" have the highest fraction of individuals employed in ethnic niches (49 percent). As is evident, niche employment in this occupation is dominated by English and Afrikaans speakers, with 82 percent of all English corporate managers working in ethnic niches.

The table indicates that the three occupations with the highest fraction of individuals employed in ethnic niches ("Corporate managers", "General managers" and "Other professionals"), are dominated by the employment of English and, to a lesser extent, Afrikaans speakers. This is not surprising given the employment preponderance of these language groups in manufacturing, their higher levels of education, on average, and Apartheid employment practices that reserved high-ranking positions for white South Africans[20]. Given that these occupations tend to be relatively high paying, upward income mobility for other language groups may be somewhat curtailed because of the English and Afrikaans niches which dominate these occupations[21].

On the other hand, occupations like "Drivers and mobile-plant operators" and "Mining, construction, manufacturing and transport labourers", have a far greater spread of niche employment across language groups. In these particular occupations, IsiZulu and IsiXhosa speakers have the highest fraction of individuals employed in niches, respectively, but all other, traditionally black, language groups[22] (excluding IsiNdebele speakers) are also well-represented in terms of niche employment.

Interestingly, the table reveals that no ethnic niches exist in seven of the occupational categories[23], which represents one quarter of all occupations in the manufacturing sector. In the other occupations, the extent of ethnic niching varies from a high of 49 percent to a low of 2 percent, with the fraction of individuals employed in a particular niche varying across language groups.

[19] A decision had to be made as to whether we should construct the variable "fraction of other adult household members employed" as opposed to "fraction of other adult household members employed in manufacturing". We chose the latter option to capture the direct effect on an individuals likelihood of niche employment in the manufacturing sector from having other household members employed in manufacturing rather than the more indirect effect from simply having other household members employed, regardless of the sector in which they work.

[20] Of all white South Africans employed in manufacturing, 51.2 percent speak Afrikaans and 48.3 percent speak English. Although Apartheid was based on a racial classification system, the high correlation between language and race necessarily implies that English and Afrikaans speakers were the primary beneficiaries of this policy.

[21] A more thorough analysis of ethnic occupational niches and income mobility is beyond the scope of this paper.

[22] The language groups traditionally associated with black South Africans are: IsiXhosa, IsiZulu, IsiNdebele, Sepedi, Sesotho, Setswana, Siswati, Tshivenda and Xitsonga.

[23] The seven occupations are: "Legislators and senior officials", "Life science and health professionals", "Teaching professionals", "Life science and health associate professionals", "Teaching associate professionals", "Subsistence, agricultural and fishery workers", and "Agricultural, fishery and related labourers".

# 5 Empirical Results

To estimate the specification in equation (4), we use a linear probability model instead of a logit or a probit as the latter two are computationally cumbersome in the presence of so many fixed effects. Our model includes demographic controls, fixed effects for language groups, magisterial districts and occupations, a measure of contact availability $CA_{lj}$, and the interaction of $CA_{lj}$ and the mean employment of the individual's language group in occupation $k$. Specifically, our interaction term includes the mean employment of language group $l$ in occupation $k$ taken in deviation from the mean employment of all language groups in occupation $k$: $CA_{lj}*(\overline{Empl}_{lk} - \overline{Empl}_k)$. Subtracting the mean of all language groups facilitates interpretation of the interaction term because it removes what the language group fixed effect is capturing.

Our demographic controls include three race dummies, age, age squared, years of education, years of education squared, a dummy indicating whether the individual lives in an urban or rural area, marital status, gender, a dummy for whether the individual relocated between the 1996 and 2001 census, the fraction of other adult household members employed in manufacturing, a dummy indicating whether the individual has a disability, and finally a dummy indicating whether the individual has access to a telephone.

Table IV presents the estimates of our network coefficient as we include fixed effects for language groups, magisterial districts and occupations[24]. The decline in our network coefficient as we include these controls highlights the benefit of our estimation strategy because without them our network measure is clearly biased upward by omitted language group, magisterial district and occupation characteristics.

Table V presents our main results. The covariates generally display the expected signs. Being female and living in a rural area increases the probability of being employed in a niche, presumably because males find it easier to obtain employment in the general labour market and urban areas increase the range of employment possibilities for individuals. Being non-white lowers the probability of being employed in a niche. This finding is driven by the fact that white individuals are over-represented in ethnic niches given their share of the total population. If an individual relocated between the 1996 and 2001 census this lowers the probability of niche employment. Clearly social networks take time to form and therefore individuals that have recently relocated will find it more difficult to obtain niche employment. We find that higher years of education decrease the probability of niche employment. This squares well with the sociological literature which suggests that less-skilled individuals are more likely to be employed in ethnic niches. Finally, the fraction of other household members employed in manufacturing lowers the probability of niche employment. Although somewhat surprising, this finding supports our decision to construct the contact availability measure using data from the full census, rather than purely from the manufacturing sector. The negative and significant estimate of this coefficient suggests that it is not only those directly employed in manufacturing that are relevant to the probability of your employment in a *particular niche* in the manufacturing sector[25].

As in Table IV we find that our network measure is positive and significant, indicating the importance of social networks for niche employment. However, given that our network measure is an interaction term, interpretation of its coefficient is not straightforward. Consequently we perform two thought experiments, one of which is inspired by Bertrand et al. (1998), and a decomposition (see Bertrand et al., 2000) to facilitate interpretation of our network measure.

First, we investigate what is the differential effect of increasing contact availability by one standard deviation for individuals in language groups highly concentrated in ethnic niches relative to

[24] Our regressions include all of the demographic controls listed above as well as the noninteracted $CA_{lj}$ measure but their coefficient estimates are not reported in this table.

[25] An alternative explanation for this result is that households diversify with regards to employment. For example, if a number of family members are employed in different occupations in the manufacturing sector, then this increases the range of opportunities for individual $i$, thereby lowering his probability of niche employment. We are grateful to Justine Burns for suggesting this alternative interpretation to us.

individuals in language groups not highly concentrated in ethnic niches. Employment in an ethnic niche for language groups one standard deviation above the mean is 38 percent whereas for language groups one standard deviation below the mean is 12 percent[26]. The standard deviation of contact availability for all language groups is 0.31 – refer to Table I. Given our estimate of $\alpha$=0.473 this implies:

the effect for language groups highly concentrated in ethnic niches is: (0.31)(0.38)(0.473)=0.056

the effect for language groups not highly concentrated in ethnic niches is: (0.31)(0.12)(0.473)=0.018

The difference between the two, 0.056 – 0.018 = 0.038, captures the effect of increasing contact availability by one standard deviation on the likelihood of employment in a niche. As the thought experiment suggests, increasing contact availability by one standard deviation raises the probability of niche employment by approximately 4 percent. Thus social networks, and in particular the number of contacts an individual has, are crucial for employment in an ethnic niche.

The second thought experiment focuses on the impact of increasing the "quality" of an individual's contacts by one standard deviation for individuals in language groups with high contact availability relative to individuals in language groups with low contact availability. The contact availability for language groups one standard deviation below the mean is 0.63 whereas for groups one standard deviation above the mean is 1.25. As stated above the standard deviation of employment in a niche is 0.13 and $\alpha$=0.473. This implies:

the effect for high contact availability groups is: (0.13)(1.25)(0.473) = 0.077

the effect for low contact availability groups is: (0.13)(0.63)(0.473)=0.039

The difference between the two, 0.077 – 0.039 = 0.038, captures the effect of increasing the "quality" of an individuals contacts on the likelihood of niche employment. Thus an increase of one standard deviation of the mean employment of all language groups in ethnic niches increases the probability of an individual's employment in a niche by approximately 4 percent, once again affirming the importance of social networks for niche employment.

Finally, Bertrand et al. (2000) provide an ingenious method for interpreting the interaction term. Following their approach we add the variable $\xi$ to the specification in equation (4):

$$Empl_{iljk} = \xi + \left(CA_{lj} * \overline{Empl}_{lk}\right)\alpha + X_i\beta + \gamma_l + \delta_j + \phi_k + CA_{lj}\theta + \varepsilon_{iljk} \tag{5}$$

where $\xi$ is scaled such that a one percentage point increase in $\xi$ leads to a one percentage point increase in the probability of niche employment *in the absence of any social network effects*. Adding this variable allows us to investigate how much network effects would magnify a policy shock (represented by $\xi$) affecting niche employment[27]. The idea is that a policy which increases $\overline{Empl}_{lk}$will in turn raise each individual's probability of niche employment through the network effect, creating a feedback. Algebraically, we average both sides of equation (5) for each language group and differentiate with respect to $\xi$ to obtain:

$$\frac{d\overline{EMPL_{lk}}}{d\xi} = 1 + \overline{CA_l} * \frac{d\overline{EMPL_{lk}}}{d\xi}\alpha$$

where $\overline{CA_l}$ is the mean of $CA_{lj}$ within each language group[28]. Solving this equation provides a measure of each language group's change in niche employment in response to a policy shock. Since the direct effect of the policy change is already included, we subtract one from the solution to isolate the effect of social networks:

[26] The standard deviation of niche employment for all language groups is 0.13 – refer to Table I. Given mean employment in ethnic niches of 0.25, this yields the 0.38 and 0.12 values for groups one standard deviation above and one standard deviation below the mean respectively.

[27] An example of a policy shock that may affect niche employment is the adoption of affirmative action legislation. This is relevant to South Africa because black economic empowerment (BEE) has been a cornerstone of employment legislation since the demise of Apartheid.

[28] To construct $\overline{CA_l}$ we first calculate the mean contact availability for language group $l$ in each area $j$. We then calculate the mean contact availability for language group $l$across all magisterial districts.

$$1/(1-\alpha\overline{CA_l})\ -1 \tag{6}$$

The above expression captures the effect of *social networks* on the probability of niche employment for language group $l$ in response to a policy shock. Table VI presents the calculations of this indirect network impact for all language groups included in the analysis[29].

The table shows that the indirect network effects are large in magnitude and vary substantially across language groups, with a high of 256 percent for Afrikaans speakers to a low of 24 percent for English speakers. To give an indication of the response for the economy as a whole, in the final row of the table we present the weighted mean of the above measure calculated over all language groups[30]. We find that the indirect network impact on the probability of niche employment for the economy as a whole is 109 percent. Thus social networks magnify a policy shock affecting niche employment by over 100 percent.

The magnitude of these results is far greater than those found by Bertrand et al (2000) and Godlonton and Burns (2006). However, the nature of the question analysed in this paper is wholly different to those which the other scholars investigated. Given that the formation of ethnic niches is fundamentally a network driven process (see Morales, 2004; and Waldinger, 1996) it is not surprising that we should find a network effect on the order of magnitude that we do[31].

# 6 Specification Checks

In this section we test whether our results are sensitive to the specification of our network measure as well as the sample used in our regressions. Table VII presents the estimates of our network measure when we change its specification. As the table indicates we find positive and significant network effects in all specifications.

Row (1) displays the estimate of our network measure from Table V for ease of comparison. In row (2) we estimate our model by using the logged version of our contact availability measure: $\ln((C_{lj}/A_l)/(N_l/T))$. In row (3) we use $(C_{lj}/A_l)$ as our contact availability measure. Evidently, this specification does not weight the measure by the language group's share of the total population. Finally, in row (4) we use $(C_{lj})$ as our contact availability measure to investigate whether there are changing returns to scale (Bertrand et al., 1998). As the table shows, our results are robust to changes in the specification of our network measure.

Table VIII presents the estimates of our network measure when we change the sample used in our analysis. As discussed previously, English and IsiZulu speakers constitute the largest fraction of individuals employed in ethnic niches, raising concerns that our results are driven primarily by these two groups. In row (1) of the table we exclude English speakers and continue to find positive and significant network effects although the magnitude of the effect is greatly diminished[32]. In row (2) we also exclude IsiZulu speakers and witness a further decline in the coefficient on our network measure. Clearly, social networks are particularly important for these groups in obtaining niche employment. However they also have a significant effect on the probability of niche employment for other groups. In row (3) we limit the sample to non-whites and once again observe a decline in our network measure relative to its value estimated on the full manufacturing sample. Nevertheless, the coefficient is still positive and significant indicating that social networks play an essential role

[29] Please note that $\alpha$ is the coefficient from row 1 of Table V.

[30] We use the fraction of individuals from language group $l$ employed in manufacturing as the weight in this calculation.

[31] An alternative decomposition, using $\alpha$ obtained from regressions of equation (4) on language group subsets of the data, was also attempted. The results (not reported) also show large and significant indirect network impacts, which bolsters our confidence in the validity of our results.

[32] Levinsohn (2004) finds that the return to speaking English in South Africa rose between 1993 and 2000. This result coupled with the fact that English speakers have the highest fraction of individuals employed in ethnic niches explains, to some extent, the large drop in the coefficient on our network measure when we exclude English speakers.

for non-whites in securing niche employment. As women are more likely to be employed in niche employment, see Table V, we remove them from our sample to ensure that our results are not driven by this gender – results reported in row (4). Again we witness a decline in our network measure but find that the coefficient is still large in magnitude and significant. Finally, in row (5) we limit the sample to urban areas. The network measure increases dramatically indicating that social networks are particularly important in urban areas for securing niche employment. This is arguably because of the competitive nature of urban area labour markets which make social networks crucial for channelling workers into ethnic niches.

As Tables VII and VIII make clear, our results are robust to changes in specification of the contact availability measure as well as the sample used in our analysis[33]. This highlights the benefit of our estimation strategy as well as the importance of social networks for channelling workers into niche employment.

# 7 Conclusion

This paper adopted the approach of Bertrand et al. (2000) to explore the impact of social networks on ethnic niche employment in the manufacturing sector in South Africa. The estimation strategy allowed us to control for numerous omitted variables that have plagued previous studies of social networks using non-network data. We defined ethnic niches using a two-part strategy and found that social networks have a large positive and significant impact on the probability of niche employment. While networks are particularly important for certain language groups they unambiguously increase the likelihood of niche employment for all language groups in all areas and in all occupations. The robustness of our results to changes in specification as well as sample selection bolsters our confidence that our findings are not the product of the vagaries of econometric estimation.

# References

[1] Alba, R. 1990. *Ethnic Identity: The Transformation of White America*. Yale University Press: New Haven, CT.

[2] Bakalian, A. 1993. *American-Americans: From Being to Feeling American*. Transaction Publishers: New Brunswick, NJ.

[3] Bertrand, M., Luttmer, E., & Mullainathan, S. 2000. "Network Effects and Welfare Cultures", *Quarterly Journal of Economics*. Vol. 116 (1), 351 – 377.

[4] Bonacich, E. 1988. "The Social Costs of Immigrant Entrepreneurship." *Amerasia Journal*. Vol. 14, 119-128.

[5] Borjas, G. 1992. "Ethnic Capital and Intergenerational Mobility". *Quarterly Journal of Economics.* Vol. 107.

[6] Borjas, G. 1995. "Long-Run Convergence of Ethnic Skill Differentials: The Children and Grandchildren of the Great Migration". *Industrial and Labour Relations Review.* Vol. XLVII.

[33] We also ran regressions to test whether the network effect differed according to the age of the individuals in our sample (results not reported). We constructed five age cohorts: less than 25; 26 - 35; 36 – 45; 46 – 55; and 56 – 65. Despite a marked increase in the coefficient on our network measure for individuals younger than 25 (the coefficient increased to 0.642) it remained fairly stable for the other age cohorts. The increase in our network measure for individuals younger than 25 indicates the importance of social networks in channelling young individuals into ethnic niches presumably because their credentials and lack of experience often make it difficult for them to obtain employment in the general labour market.

[7] Buhai, S. & van der Leij, M. 2006. "A social network analysis of occupational segregation". Timbergen Institute Discussion Paper

[8] Datcher, L. 1983. "The Impact of Informal Networks on Quit Behaviour". *Review of Economic Statistics*. Vol 65(3), 491-95.

[9] Devine, T. & Kiefer, N. 1991. "Empirical Labor Economics: The Search Approach", Oxford University Press, NY.

[10] Elliott, James R. 2001 "Referral Hiring and Ethnically Homogeneous Jobs: How Prevalent Is the Connection and for Whom?" *Social Science Research.* Vol 30, 401-425

[11] Fernandez, R., Castilla, E., & Moore, P. 2000. "Social Capital at Work: Networks and Employment at a Phone Center". *American Journal of Sociology.* Vol 105(5), 1288 –1356.

[12] Godlonton, S. & Burns, J. 2006. "Social Networks, Employment and Worker Discouragement: Evidence from South Africa". *Unpublished mimeo*

[13] Granovetter, M. S. 1974. "Getting a Job: A Study of Contacts and Careers", first edition, Harvard University Press, Cambridge, MA.

[14] Granovetter, M. S. 1973. "The Strength of Weak Ties". *American Journal of Sociology.* Vol. 78: 6

[15] Granovetter, M. S. 1995. "The Economic Sociology of Firms and Entrepreneurs." In *The Economic Sociology of Immigration*, edited by A. Portes. New York: Russell Sage Foundation. 128–165.

[16] Holzer, H. 1987. "Informal job search and black youth unemployment". *American Economic Review*, Vol. 77(3).

[17] Ioannides, Y.M. and L.D. Loury. 2004. "Job Information Networks, Neighborhood Effects and Inequality", *Journal of Economic Literature*. Vol. XLII, December, 1056-1093.

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

[19] Lichter, M. 2000. "Race, Space, and the Employment of Less-Educated Immigrants and Natives". *Dissertation Abstracts International, A: The Humanities and Social Sciences.* Vol. 60(7).

[20] Lieberson, S. 1980. "A Piece of the Pie." *Blacks and White Immigrants since 1880.* Berkeley: University of California Press.

[21] Light, I. & Gold, S. 2000. *Ethnic Economies*. Academic Press: San Diego, CA.

[22] Logan, J., Alba, R. & McNulty, T. 1994. "Ethnic Economies in Metropolitan Regions: Miami and Beyond." *Social Forces* Vol. 72

[23] Manski, C. 1993. "Identification of Endogenous Social Effects: the Reflection Problem". *Review of Economic Studies*, Vol. 60(3).

[24] Mlatsheni, C. & Rospabe, S. 1999. "Why is youth unemployment so high an unequally spread in South Africa" *Unpublished mimeo.*

[25] Model, S. 1993. "The Ethnic Niche and the Structure of Opportunity: Immigrants and Minorities in New York City." In *The Historical Origins of the Underclass*, edited by M. Katz. Princeton: Princeton University Press.

[26] Model, S. & Ladipo, D. 1996. "Context and Opportunity: Minorities in London and New York." *Social Forces* Vol. 75.

[27] Montgomery, J. 1991a. "Social networks and labor-market outcomes: Toward an economic analysis". *American Economic Review*, Vol. 81(5).

[28] Morawska, F. 1990. "The Sociology and Historiography of Immigration." In *Immigration Reconsidered*, edited by V. Yans-McLaughlin, New York: Oxford University Press.

[29] Morales, M. 2004. "Ethnic niches, pathway to economic incorporation or exploitation? Labour market experiences of Latina/os". *Unpublished mimeo*

[30] Ooka, E. & Wellman, B. 2003. "Does social capital pay off more within or between ethnic groups? Analysing job searches in five Toronto ethnic groups", in Eric Fong (ed.) *Inside the Mosaic*. Toronto, University of Toronto Press

[31] Rath, J. 2002. "A Quintessential Immigrant Niche? The Non-Case of Immigrants in the Dutch Construction Industry". *Entrepreneurship and Regional Development*. Vol. 14.

[32] Rees, A. 1966. "Information networks in labor markets". *American Economic Review*, Vol. 56

[33] Rees, A., & Shultz, G. 1970. *Workers and Wages in an Urban Labor Market*. Chicago University Press: Chicago.

[34] Sassen, S. 1995. "Immigration and Local Labor Markets." in Protes, A (ed.) *The Economic Sociology of Immigration: Essays on Networks, Ethnicity, and Entrepreneurship*. New York: Russell Sage Foundation.

[35] Simon, C. & Warner, J. 1992. "Matchmaker, Matchmaker: The Effect of Old Boy Networks on Job Match Quality, Earnings and Tenure". *Journal of Labor Economics*. Vol. 10(3), 306-330.

[36] Spalter-Roth, R. & Lowenthal, T. 2005. "Race, Ethnicity and the American Labour Market: What's at work?" *American Sociological Association* http://www2.asanet.org/centennial/race_ethnicity_labormarket.pdf

[37] Staiger, D. 1990. "The effect of connections on the wages and mobility of young workers". *mimeo*, MIT.

[38] Topa, G. 2001 "Social interactions, local spillovers and unemployment". *Review of Economic Studies*, Vol. 68: 2

[39] Ullman, J. 1966. "Employee referrals: Prime tool for recruiting workers". *Personnel* Vol. 43**,** 30 –35

[40] Waldinger, Roger. 1996a. *Still the Promised City? African-Americans and New Immigrants in Post- Industrial New York*. Cambridge: Harvard University.

[41] Waldinger, R. 1996b. "Ethnicity and Opportunity in the Plural City." In *Ethnic Los Angeles*, edited by R. Waldinger and M. Bozorgmehr. New York: Russell Sage Foundation.

[42] Wilson, F. 1997. "Ethnic Concentrations and Labor Market Opportunities". *CDE Working Paper No. 97-17*

[43] Wilson, F. 2000. "Metropolitan Labor Markets and Ethnic Niching: Introduction to a Research Project". *Institute for Research on Poverty* Discussion paper no. 1204-00.

[44] Wittenberg, M & Pearce , C. 1996. "Youth Unemployment: Some Perspectives from the South African Living Standards and Development Survey" in L. Chisholm et al *Out-of-School Youth Report: Policy and Provision for Out-of-School and out-of-Work Youth,* Education Policy Unit, University of Witwatersrand, 1996.

TABLE I

SUMMARY STATISTICS

| Variable | Mean | Std Deviation |
|---|---|---|
| Employment in a niche | 0.25 | 0.13 |
| Contact Availability (CA) | 0.94 | 0.31 |
| Black | 0.56 | 0.50 |
| Coloured | 0.17 | 0.38 |
| Asian/Indian | 0.08 | 0.27 |
| White | 0.19 | 0.39 |
| Male | 0.66 | 0.47 |
| Rural | 0.14 | 0.35 |
| Relocated between 1996 and 2001 Census | 0.19 | 0.39 |
| Access to a telephone | 0.98 | 0.13 |
| Disability | 0.02 | 0.15 |
| Married | 0.64 | 0.48 |
| Age | 37.43 | 10.34 |
| Years of Education | 9.67 | 4.06 |
| Fraction of other adult household members employed in manufacturing | 0.08 | 0.16 |

Notes:

Mean CA is a weighted average, where the weights are given by the fraction of individuals from each language group employed in manufacturing

TABLE II
SUMMARY STATISTICS BY LANGUAGE GROUP

| Language | Group size in full sample | Group size in manufacturing | Employment in a niche | Age | Years of education | Fraction of other adult household members employed in manufacturing | Relocated between the 1996 and 2001 census |
|---|---|---|---|---|---|---|---|
| Afrikaans | 317891 | 21538 | 0.23 | 36.27 | 10.60 | 0.10 | 0.19 |
| English | 200689 | 20146 | 0.43 | 38.97 | 11.90 | 0.11 | 0.22 |
| IsiNdebele | 36656 | 1035 | 0.00 | 37.91 | 8.11 | 0.05 | 0.18 |
| IsiXhosa | 386592 | 8683 | 0.21 | 37.32 | 8.49 | 0.05 | 0.19 |
| IsiZulu | 500289 | 18435 | 0.32 | 36.95 | 8.51 | 0.06 | 0.15 |
| Sepedi | 203221 | 5662 | 0.11 | 38.09 | 8.50 | 0.05 | 0.20 |
| Sesotho | 186128 | 6684 | 0.08 | 37.14 | 8.51 | 0.05 | 0.17 |
| Setswana | 191673 | 5560 | 0.13 | 38.41 | 8.75 | 0.06 | 0.17 |
| Siswati | 58440 | 1816 | 0.09 | 36.93 | 7.44 | 0.07 | 0.16 |
| Tshivenda | 49080 | 1138 | 0.07 | 37.50 | 8.05 | 0.05 | 0.18 |
| Xitsonga | 96821 | 2711 | 0.04 | 36.59 | 7.39 | 0.06 | 0.19 |

TABLE III
THE EXTENT OF ETHNIC NICHING BY LANGUAGE GROUP AND OCCUPATION

| Occupation | Language | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Afrikaans | English | IsiNdebele | IsiXhosa | IsiZulu | Sepedi | Sesotho | Setswana | Siswati | Tshivenda | Xitsonga | **Occupation Mean** |
| Legislators and senior officials | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.00** |
| Corporate managers | 0.20 | 0.82 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.49** |
| General managers | 0.17 | 0.75 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.44** |
| Physical, mathematical and engineering science professionals | 0.27 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.27** |
| Life science and health professionals | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.00** |
| Teaching professionals | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.00** |
| Other professionals | 0.07 | 0.80 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.46** |
| Natural and engineering science associate professionals | 0.20 | 0.39 | 0.00 | 0.00 | 0.00 | 0.00 | 0.14 | 0.00 | 0.00 | 0.00 | 0.00 | **0.19** |
| Life science and health associate professionals | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.00** |
| Teaching associate professionals | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.00** |
| Other associate professionals | 0.06 | 0.83 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.31 | 0.00 | 0.00 | 0.00 | **0.41** |
| Office clerks | 0.19 | 0.66 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.23 | 0.00 | 0.00 | 0.00 | **0.30** |
| Customer service clerks | 0.10 | 0.26 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.10** |
| Personal and protective services workers | 0.00 | 0.00 | 0.00 | 0.00 | 0.09 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.02** |
| Models, salespersons and demonstrators | 0.08 | 0.68 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.47 | 0.00 | 0.00 | **0.28** |
| Market-oriented skilled agricultural and fishery workers | 0.00 | 0.00 | 0.00 | 0.00 | 0.27 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.08** |
| Subsistence agricultural and fishery workers | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.00** |
| Extraction and building trades workers | 0.05 | 0.04 | 0.00 | 0.22 | 0.10 | 0.19 | 0.15 | 0.10 | 0.00 | 0.00 | 0.09 | **0.10** |
| Metal, machinery and related trades | 0.26 | 0.07 | 0.00 | 0.00 | 0.06 | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 | 0.12 | **0.10** |
| Handicraft, printing and related trades workers | 0.00 | 0.32 | 0.00 | 0.00 | 0.07 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.11** |
| Other craft and related trades workers | 0.16 | 0.00 | 0.00 | 0.08 | 0.61 | 0.08 | 0.03 | 0.12 | 0.35 | 0.35 | 0.00 | **0.22** |
| Stationary-plant and related operators | 0.19 | 0.00 | 0.00 | 0.20 | 0.28 | 0.00 | 0.10 | 0.00 | 0.00 | 0.00 | 0.00 | **0.14** |
| Machine operators and assemblers | 0.68 | 0.01 | 0.00 | 0.22 | 0.57 | 0.06 | 0.08 | 0.19 | 0.00 | 0.00 | 0.00 | **0.36** |
| Drivers and mobile-plant operators | 0.09 | 0.00 | 0.00 | 0.43 | 0.44 | 0.42 | 0.06 | 0.25 | 0.00 | 0.00 | 0.13 | **0.26** |
| Sales and services elementary occupations | 0.20 | 0.00 | 0.00 | 0.37 | 0.25 | 0.10 | 0.08 | 0.00 | 0.00 | 0.00 | 0.00 | **0.17** |
| Agricultural, fishery and related labourers | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | **0.00** |
| Mining, construction, manufacturing and transport labourers | 0.13 | 0.00 | 0.00 | 0.49 | 0.46 | 0.16 | 0.26 | 0.26 | 0.26 | 0.19 | 0.05 | **0.28** |
| Other | 0.12 | 0.07 | 0.00 | 0.00 | 0.08 | 0.00 | 0.12 | 0.00 | 0.00 | 0.00 | 0.00 | **0.07** |
| **Group Mean** | **0.23** | **0.43** | **0.00** | **0.21** | **0.32** | **0.11** | **0.08** | **0.13** | **0.09** | **0.07** | **0.04** | **0.25** |

TABLE IV

REGRESSION ESTIMATES OF NETWORK COEFFICIENT AS ADDITIONAL FIXED EFFECTS ARE INCLUDED

| | Network Measure |
|---|---|
| | α |
| No controls | 0.544*** |
| | [0.007] |
| + Language group fixed effects | 0.520*** |
| | [0.007] |
| + Magisterial district fixed effects | 0.508*** |
| | [0.007] |
| + Occupation fixed effects | 0.473*** |
| | [0.007] |

Notes:

Standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

TABLE V

REGRESSION ESTIMATES OF NETWORK COEFFICIENT

| | Employment in a niche |
|---|---|
| Network Measure | 0.473*** |
| | [0.007] |
| CA | 0.013*** |
| | [0.001] |
| Rural dummy | 0.025*** |
| | [0.005] |
| Male | -0.029*** |
| | [0.004] |
| Individual is Black | -0.062*** |
| | [0.011] |
| Individual is Coloured | -0.059*** |
| | [0.005] |
| Individual is Asian/Indian | -0.083*** |
| | [0.007] |
| Individual relocated between 1996 & 2001 Census | -0.012*** |
| | [0.003] |
| Has access to telephone | -0.016* |
| | [0.009] |
| Has disability | -0.006 |
| | [0.008] |
| Married | -0.006 |
| | [0.004] |
| Married Male | 0.008 |
| | [0.005] |
| Age | 0.001 |
| | [0.001] |
| Years of education | -0.001*** |
| | [0.000] |
| Age squared | 0 |
| | [0.000] |
| Years of education squared | -0.000* |
| | [0.000] |
| Fraction of other adult household members employed in manufacturing | -0.011 |
| | [0.008] |
| Constant | 0.207*** |
| | [0.044] |
| Observations | 93408 |

Notes:

Standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

TABLE VI

INDIRECT NETWORK IMPACT ON PROBABILITY OF NICHE EMPLOYMENT

| Language | Mean CA | Indirect Network Impact |
|---|---|---|
| Afrikaans | 1.52 | 2.56 |
| English | 0.41 | 0.24 |
| IsiNdebele | 0.75 | 0.55 |
| IsiXhosa | 1.20 | 1.31 |
| IsiZulu | 0.80 | 0.60 |
| Sepedi | 0.64 | 0.44 |
| Sesotho | 1.24 | 1.42 |
| Setswana | 0.80 | 0.60 |
| Siswati | 0.82 | 0.64 |
| Tshivenda | 0.72 | 0.52 |
| Xitsonga | 0.66 | 0.46 |
| Weighted Average | 0.94 | 1.09 |

TABLE VII

SPECIFICATION CHECKS

| | Employment in a niche |
|---|---|
| (1) Network Measure from Table IV | 0.473*** |
| | [0.007] |
| (2) CA measured in logs rather than levels | 3.261*** |
| | [0.032] |
| (3) CA measured as ($C_{lj}$ / $A_l$) | 9.921*** |
| | [0.091] |
| (4) CA measured as ($C_{lj}$) | 0.006*** |
| | [0.000] |

Notes:
Standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

TABLE VIII

SAMPLE SELECTION CHECKS

| | Employment in a niche |
|---|---|
| (1) Exclude English Speakers | 0.115*** |
| | [0.007] |
| (2) Exclude English and IsiZulu Speakers | 0.109*** |
| | [0.007] |
| (3) Exclude white South Africans | 0.416*** |
| | [0.007] |
| (4) Exclude women | 0.427*** |
| | [0.008] |
| (5) Exclude rural areas | 0.951*** |
| | [0.010] |

Notes:
Standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%