Labor Market Discrimination and Sorting: Evidence from South Africa

Martin Abel*†

March 22, 2017

Find latest version here.

Abstract

Using a unique data set of classified ads in South Africa, I explore whether employers discriminate against immigrants in the hiring process. I develop a quasi-experimental method to estimate discrimination exploiting variation in the applicant pool composition due to the timing of postings. Consistent with a tournament models in which immigrants are penalized, I find that both foreigners and natives benefit from being pooled with foreign job seekers. Next, I test whether discrimination affects search behavior. Controlling for location fixed effects, I find suggestive evidence for sorting: immigrants search further away and higher discrimination in the residential area is positively correlated with the decision to search in different suburbs. This additional cost to job seekers has not been explored in the discrimination literature.

*Harvard University. Email: abel@fas.harvard.edu

†Acknowledgements: this paper greatly benefited from discussions with and comments from Willa Brown, Rulof Burger, Adi Eyal, Rema Hanna, Lawrence Katz, Asim Khwaja, Sendhil Mullainathan, Devah Pager, Dorit Posel, Vimal Ranchod, Roland Rathelot, Jann Spiess and seminar participants Harvard University, University of Cape Town, Stellenbosch University and the New England Development Conference. Simeon Abel provided excellent research assistance. All errors and omissions remain fully my responsibility. I gratefully acknowledge funding from the World Bank's Gender Innovation Lab.

1 Introduction

It is notoriously difficult to study labor market discrimination. In recent years, a large literature has emerged in which researchers send fictitious resumes to estimate discrimination along many characteristics including race, gender, age, and sexual orientation (for recent reviews see Bertrand and Duflo (2016); Neumark (2016); Rich (2014)). While these correspondence studies have important advantages over observational and audit studies, they can only be used for select types of vacancies (Pager (2007)) and have several methodological limitations (Heckman (1998)).

In particular, audit studies have been applied almost exclusively in the formal sector of developed countries.[1] Yet, a minority of the global labor force works in these markets and these are precisely the settings in which employees enjoy the most comprehensive legal protection from discrimination. This study provides evidence for a very different setting: the situation of immigrants working in the informal sector in South Africa. Obtaining evidence on discrimination through audit studies is difficult in these markets as employers typically neither advertise job vacancies nor use formal hiring processes that include systematic screening of job seekers.[2] Pager (2007, p.111) therefore concludes that in this context "*any study would require in-person application procedure*". I will argue that the observational data collected in this study allows me to not only apply many of the methodological advantages of audit studies, but also address some of the key methodological limitations.

I collect a unique data set of 5,500 job seeker advertisements in the domestic work, nanny and general work sector collected from South Africa's largest classified ad website, which caters mainly to the informal sector[3]. When employers search for workers in a particular sector and region they are presented with a list of truncated summaries of job seeker profiles. Based on the information visible at the search stage employers decide whether to click on the full profile, which provides additional information and contact details of the job seeker. Importantly, the profile page reports how many people previously clicked on the profile which serves as a measure of employer demand. This setting mimics the ideal experiment in that, similar to correspondence studies, researchers observe exactly

[1]Bertrand and Duflo (2016) note, for example, that there is not a single audit study conducted in Africa.

[2]Moreover, ethical concerns have been raised about conducting audit studies that involve sending thousands of fictitious CVs in labor markets with mass unemployment as this would further impede the hiring process

[3]According to the ILO, one of the defining characteristics of the informal sector is that "*Labour relations are based mostly on casual employment, kinship or personal and social relations rather than contractual arrangements with formal guarantees.*" According to the data I collect, 80% of job seekers employed through the website are agreed on the basis of a verbal agreement rather than a written contract. It is predominantely used by private households looking to hire: among job seekers who receive a response to their advertisement, 95% report being contacted by a private household.

the same information as the potential employer at the time of her decision whether to further screen a job seeker. I find that after controlling for other covariates, stating that one is an immigrant is associated with receiving about 10-20% fewer employer visits to the profile page than job seekers who state they are South African.

This unique data set also allows me to at least partly address previous studies' limitation that the pool of applicants from which employer can hire (in addition to the fictitious applicant) is not observed. This is an important concern as the size, qualification and racial composition of the applicant pool affect the cost of discrimination. Intuitively, the fewer and relatively less qualified applicants of the preferred race, the more costly are discriminatory hiring practices.[4] For the classified ad market I investigate, we observe the full set of available candidates presented to employers in her local labor market. What is more, the nationality composition presented to employers is quasi-random as the order is determined by the time of the day that job seekers post their profile. Consistent with theoretical predictions from a tournament model in which employers discriminate against foreigners, I find that being pooled with a larger share of immigrants benefits *both* foreign and South African job seekers. Employing an estimator that soley exploits variation in the composition of the choice set is a methodological contribution that may be applied to estimate preferences in other scenarios.

Exisiting audit studies have been criticized on the ground that minority workers may be able to identify discriminating employers, which would alleviate the effects of employer discrimination (Heckman (1998). In an extension of Becker (1957), Arrow (1973) shows that the wage differential between groups can disappear if there are a sufficient number of non-discriminating employers. In equilibrium, this may result is a (partially) segregated employment without (wage) discrimination. The extent to which (wage) differentials disappear depends on how well minorities can target non-discriminating employers in their job search (Black (1995); Lang and Lehmann (2012)). On the website, I observe both where job seekers live and the location where they search for work. This data allows me to shed some light on whether minority applicants adjust their search behavior in response to discrimination. Employing a residential location fixed effect strategy, I find that immigrants search 21% further away and suggestive evidence that higher levels of discrimination faced in the area of residence induces immigrants to search for work in a different suburb. This is, to the best of my knowledge, the first

[4]A related challenge to interpreting audit studies is that results are consistent with a world in which employers' racial preferences are proportionate to the population's racial composition but in which the share of minority applicants is disproprtionately high.

evidence on geographic sorting in response to discrimination. It presents an important additional cost of labor market discrimination to job seekers, especially in a setting with high transport costs like South Africa.[5]

Last, this study hopes to contribute to the literature on the nature of discrimination.[6] Altonji and Pierret (2001, AP henceforth) try to distinguish between statistical and taste discrimination by testing how the effect of observable characteristics on wage develop over time. According to the statistical discrimination theory, observable characteristics such as education and race should explain less of wage levels over time as employers learn about harder to observe determinants of productivity.[7] Building on the AP employer learning model, I find that employer demand *diverges* between South Africans and foreigners once more information is revealed on the applicants' profile. This is consistent with a model in which employers *positively* statistically discriminate against foreigners. Results from an anonymous survey I conduct with 208 domestic workers supports this conclusion: I find that immigrants have characteristics (e.g. age, education) favored by employers.

There are other differences between South African and foreign workers which may explain the results. Employers may refrain from selecting foreigners due to uncertainty of their legal status. However, I find that among foreign job seekers, those that indicate they have a work permit receive about 8% *fewer* profile clicks. The South African government suspects that some employers prefer hiring undocumented foreigners because they have less bargaining power and are thus more exploitative (DoL 2007). Alternatively, revealing that one has a work permit may also signal a higher reservation wage to employers. Results from the survey paint a subtle picture: undocumented immigrants do not have lower reservation wages or are paid lower wages than documented immigrants or South Africans. They are, however, less likely to know about or willing to utilize the CCMA, a widely used

[5]Kerr (2015) estimates that household spend on average 11% of their income on transport cost and that modes like trains or buses impose an effective tax rate of 25-30% on hourly wages.

[6]The theory of statistical discrimination was pioneered by Phelps (1972) who reasoned that in a world of asymmetric information, employers assess the expected productivity of workers according to the average of the population with similar observable characteristics. Labor market discrimination can result when observable characteristics like race or gender are used by employers to infer information about productivity. By contrast, according to Gary Becker's model of taste-based discrimination, employers have discriminatory tastes and are thus willing to pay in order to not hire certain groups (Becker (1957)). While existing audit studies (Bertrand and Mullainathan (2003); Oreopoulos (2011)) credibly show that discrimination exists in the screening process, they do not provide conclusive evidence on the source of discrimination (Lang and Lehmann (2012)). For example, a lower return to more credentials is at odds with both the taste and statistical discrimination approach (Bertrand and Mullainathan (2003)).

[7]The authors find little support for a model of statistical discrimination by race: while the coefficient on education falls, the negative coefficient for black race persists. One concern with this interpretation is that the discriminated group may face additional forms of discrimination such as being omitted in promotion decisions due to reasons of statistical discrimination, which may explain the persistent negative effect on race. CITE EXPERIMENTAL STUDIES.

labour arbitration institution available to employees to take employers to court.

In sum, foreigners receive fewer profile clicks despite positive statistical discrimination and lower risks of being taken to court, pointing to the importance of taste discrimination. Consistent with this interpretation, I find that discrimination increases in the intimacy of the employer-worker relationship: it is highest in the nanny and lowest in the gardening sector.

One limitation of this research design is that, similar to most audit studies, we only observe the first step in the hiring process. Faced with a large number of applications they may use a simple heuristic in screening applicants and stop reading once they see a certain signal, in our case foreign nationality (Bertrand and Mullainathan (2003)). A review by Riach and Rich (2002) concludes around 90% of discrimination occurs at this first selection stage. Kuhn and Shen (2013) develop a screening cost model and show that narrow search strategies that pre-screen applicants are more commonly used in low-skilled sectors with large applicant pools. A coarse job screening strategy is thus highly relevant for the high unemployment, low-skill sectors context of this study (Bartos et al. (2016)).[8] A second limitation is that while we observe the same information as employers, we may fail to notice subtle differences between profiles that differ between natives and immigrants. I try to address this by controlling for a rich set of profile characteristics including the number of words, spelling and grammar mistakes, the time and day of the posting, and a set of variables for information such as experience, age, and qualification. I also find that results are robust to controlling for job seeker characteristics *unobservable* to employers at the screening stage.

This study adds to a growing literature on online labor markets. These markets, which are predicted to increase rapidly in the future (Horton (2010)), are particularly prone to information asymmetries since, especially at an early stage in the hiring process, employers have very limited information of the job seeker (Pager (2007)). The study is most similar to Kuhn and Shen (2013, 2014) who use data from an online labor market in Xiamen, China to test how employer call back rates differ for job seekers with varying observable characteristics like age, gender, or education. This study also contibutes to a larger literature that uses data from online markets to estimate preferences. For example, Pope and Sydnor (2011) analyze data from propsper.com, a peer-to-peer lending website in which borrowers create a loan listing with unverified personal information and pictures in order

[8] In addition, social-psychology theories of *unconscious* bias predict that discrimination is most pronounced when firms have limited information on job applicants (Arrow (1998); Pager (2007)).

to request funding. The authors find that signals about the lenders' age, race, and gender from the advertised profile significantly affect the likelihood to receive loans and the interest rates paid by borrowers.

The paper proceeds as follows. Section 2 reviews the evidence on labor market discrimination and (undocumented) immigration in the South African context. Section 3 summarizes data collected from an online labor market for domestic workers. Section 4 discusses the identification strategy, reports the main results and offers model extensions that explore the role of the applicant pool composition and market thickness. Section 5 explores mechanisms and Section 6 tests if spatial sorting of job seekers is linked to employer preferences. Section 7 concludes.

2 Background: Immigration and discrimination in the South African labor market

Racial discrimination has played an important role in South Africa's history, both before and during *apartheid*. The post-*apartheid* regime implemented legislation to improve the economic situation of previously disadvantaged groups, most notably the Employment Equity Act (Act 55 of 1998) and the Broad-based Black Economic Empowerment (BEE) Act (Act 53 of 2003). Yet, the legacy of half a century of racial segregation and discrimination under apartheid has persistent effects even 20 years after the democratic transition. Firstly, while BEE created a small class of highly successful workers, the majority of businesses are still owned by the white minority. Secondly, racial tension and discrimination exists between previously disadvantaged population groups, partly as the result of the apartheid's policies of racial division and the selective lifting of restrictions for Coloureds[9] and Asians in the 1980s (Seekings and Nattrass (2008)). Thirdly, it has been argued that apartheid's racist immigration laws, class division, de-sensitization to violence, and attitude of superiority towards the rest of Africa are partly to blame for South Africa's high degree of xenophobia that resulted in violent riots in 2008 and 2015 (Crush (2008)).

This paper focuses on the situation of immigrants in the South African labor market. At the end

[9]The term '*Coloured*' refers to a heterogeneous ethnic group with ancestry from Europe, local tribes, West Africa, Mozambique and various places in Asia including India, Indonesia, and Malaysia. Coloureds tend to have lighter skin than the black population group and make up 9% of the population in South Africa and are mainly concentrated in the Western Cape (Seekings and Nattrass (2008)).

of *apartheid*, South Africa opened up to migration. The total number of documented immigrants increased by 60% from 1994 to 2003 and the share of Africans among immigrants increased from 25% to almost 50% over this time (DOL (2007)). In 2002, amidst increasing unemployment rates, the government passed the Immigration Act to facilitate easier access for South African employers to foreign skills while limiting labor market access for unskilled and semi-skilled immigrants. Employment opportunities in the formal sector became scarcer for immigrants due to "South Africans first" legislation, most notably in the mining sector. As a result, the proportion of foreign miners fell from 51% in 1997 to 38% in 2006 (DOL (2007)). Yet, rather than deterring immigration, many believe that the new migration policy led to a shift from legal to undocumented migration and employment of immigrants in the informal rather than formal sector. Undocumented immigration in years was further spurred by the political and economic crisis in neighboring countries, most notably Zimbabwe.

It is difficult to measure the number of immigrants living in South Africa. According to the 2001 census, 688,000 foreigners lived in South Africa, yet the number of irregular immigrants is likely much higher. The government estimated the number of undocumented immigrants to be around 600,000 but they cannot rule out that the true figure might be as high as three million (DOL (2007)). The government's response to the influx of undocumented immigrants was increased deportation: more than 3 million undocumented immigrants have been deported since 1990 (97% from SADC countries including 90% from Zimbabwe and Mozambique) and the annual number of cases tripled between 2002 and 2007 (Crush (2011)).

Large immigration flows combined with high unemployment rates have fostered xenophobic attitudes among South Africans as revealed in a survey of South Africans' attitudes towards migrants and refugees (Crush (2008)).[10] Yet, the South African Department of Labor (DoL) concludes that "*although a systematic survey of employers has not been conducted, case study evidence suggests that there is widespread preference for non-South African workers.*" (DOL (2007)). If true, this poses a puzzle: Do employers prefer hiring immigrants despite widespread xenophobic attitudes? What role does the legal status and exploitability of immigrants play? And how does discrimination vary between occupations? This paper hopes to shed light on these questions using data from the domestic

[10] 84% of respondents believe that South Africa lets too many foreigners into the country, 74% favor deporting people that are not contributing economically, and two-thirds agree that rights to legal protection and police protection should never be granted to undocumented immigrants. Hostile attitudes are routed in beliefs that foreigners pose a criminal threat (48%), compete for jobs (37%), and bring diseases (29%). Xenophobic attitudes only differ slightly by employment status and tend to be somewhat stronger among whites across all income brackets (Crush 2010).

worker, nanny, and general work sector in South Africa.

3 Data

I collect data on job seekers that use the website www.gumtree.co.za, South Africa's most widely used website for classified job advertisements.[11] Gumtree allows job seekers to post 'job wanted' advertisements for free in sectors ranging from housekeeping to engineering.[12] People specify the suburb where they search for employment, include a text with information about themselves, provide their home address and the position they are looking for and may upload a picture. Employers can either post 'job offered' advertisements or search for people by sector, region and suburb. Search results provide a list of matched profiles that include a truncated version of the text (see Figure A.1). By clicking on the truncated profile of a job seeker, the employer is directed to the individual profile site that contains the full text and contact information (see Figure A.2). The profile page also has information on the number of people that previously visited this site which will serve as the main outcome variable of interest in the analysis of employer demand.

I collect advertisements posted by job seekers in the housekeeping, general work, and nanny sector between October 2012 and January 2013 in the Cape Town region.[13] Data was captured twice a day for all job seeker ads posted 12 to 24 hours before to alleviate potential selection problems as profiles of people successfully hired may be taken off the website.[14] I collect all posted information, the time of the posting and number of visits for more than 5,400 advertisements. I use only the text visible to employers on the *search result* site to encode characteristics provided by the job seeker including nationality, age and gender as well as qualification information such as whether the person has job experience or certificates, holds a driver's license, and is willing to live in the employer's house. I also manually review a random subset of job postings and quantify the number of spelling

[11] Over all provinces, Gumtree has more than 30,000 advertisements posted by employers. Comparable numbers for the next most popular online sites and newspapers are 9000 by Job mail, 1000 by Career junction, and 14,000 by JunkMail.

[12] Gumtree also gives the option to pay for priority advertisements that are posted at the top of the search results. These advertisements are excluded from the analysis as this service is mainly used by agencies that act as brokers.

[13] I focus on Cape Town as it is the region most highly on the job website. Other cities and more rural areas are unsuitable for this analysis as they have a much lower frequency of advertisement postings.

[14] Job seekers have the option of re-posting an advertisement once a month. This may potentially bias results as employers may not click on a profile that they recognize from a previous search. In practice, less than 5% of the profiles are re-posted using the identical advertisement and I exclude these from the analysis. More frequently, job seekers posts the same profile in a different suburb. This is less of a problem since employers are likely only looking for people searching in their specific suburb.

and grammar mistakes as well as flawed punctuation. The goal of this exercise is to codify all the profile information that may determine employer demand.

The main focus of this study is the domestic work sector, which provides employment to about one million people or 7% of the (employed) labor force in South Africa (Dinkelman and Ranchod (2012)). Data from the South African Labor Force Survey (LFS) confirms that domestic workers are predominately drawn from lower socio-economic classes: their average level of education (7.5 years) is lower than that of other unemployed women (8.7 years) and paid workers (11.0 years). Domestic workers tend to be older (41.5 years vs. 35.1 years of population in labor force) and more likely to be black (92% vs. national average of 77%). It is the single most important sector for women; in fact, 18% of all women with paid jobs are employed as domestic workers. A study of domestic workers in Johannesburg shows that almost half of all domestic workers are either internal or cross-border migrants (Dinat and Peberdy (2007)).

[Table 1 here]

Table 1 provides summary statistics for the gumtree domestic work sample divided by the main nationality groups. As the website does not systematically collect information of job seekers, all data reported is derived from the profile text visible in the search result. For example, the age mean is calculated from the 28.3% of Gumtree job seekers that provide information on their age. There is thus the possibility of reporting bias - an issue I will discuss below.[15]

[Figure 1 here]

For a first descriptive summary of employer demand across population groups, Figure 1 plots the distribution of profile clicks for South Africans, Malawians, Zimbabweans and people who do not state their nationality. The figure shows that the profile clicks follow a normal distribution for each population group. South Africans get more profile clicks than profiles without nationality information

[15] While this study focuses on domestic workers, I will also compare results to the nanny and general work sector (which includes gardening) to test how employer preferences vary between sectors. Table A.1 compares summary statistics across sectors. 75% and 80% of job seekers revealing their gender in the housekeeper and nanny sector, respectively, are women, whereas the general work sector is dominated by men. The share of Malawians is disproportionately high in the general work sector and Zimbabweans are overrepresented in the nanny sector. While job seeker characteristics such as age, experience, and the share holding work permits are relatively similar across sectors, I can reject a test of equal means for almost all variables.

which in turn get more clicks than profiles of Zimbabweans and Malawians. The order of employer preference seems robust to outliers; in fact, the ranking of distributions (almost) follows a pattern of statistical dominace. However, these differences may be due to differences in profile characteristics between population groups documented in Table 1. The next section offers two empirical strategies to estimate the causal effect of nationality on employer demand.

4 Identification and Results

4.1 Selection on Observables

4.1.1 Domestic Work Sector

As discussed in section 3, I observe exactly the same information as the employer at the time she makes the decision whether to screen an applicant's profile. While employment studies using observational data often suffer from the omitted variable problem, this study design guarantees that the observable data is orthogonal to the error term (assuming that I correctly codify the profile information). The first empirical strategy is thus to estimate what information revealed in the search result determines the decision of the employer to click on the full profile page. I estimate the following specification:

$$y_i = \alpha + \beta_1 SA + \beta_2 For + \gamma X_i + \phi_j + time_t + lang_i + e_i \tag{1}$$

I regress the log number of profile clicks y_i on a vector of control variables (X) which includes covariates listed in Table 1 such as gender, age, experience, and references. I also control for a set of suburb dummies ϕ_j and a vector of variables $time_t$ that control for the day and time at which the profile was posted and the number of hours for which the profile was online by the time of data collection. The covariate vector $lang_i$ controls for the number of spelling, grammar and punctuation mistakes in the profile.

This paper's main focus is on the role of nationality. I include dummies capturing if the applicants are South African (SA) or a foreigner (For). I will also estimate specifications with a set of nationality dummies to account for the fact that employers may have differential preferences for applicants from different African countries. The omitted category in these specifications is the group of people providing no information on nationality. The parameters of interest β_i thus capture the difference in profile clicks among profiles with the same observable characteristics X, posted at the same time for the same suburb but with varying nationality.

[Table 2 here]

Column 1 in Table 2 shows that profiles of South Africans get about 7.5% more clicks compared to profiles without nationality information whereas profiles of Zimbabweans and Malawians get 3.4% and 10.9% fewer clicks, respectively. The difference in coefficients for natives and foreigners are statistically significant. Once I control for language mistakes, the coefficient on the Zimbabwe dummy becomes more negative (Column 2). This reflects that Zimbabweans tend to make fewer language mistakes. Estimates are unchanged when I control for a set of suburb dummies which suggests that differences in employer demand are not due to differences in where people search for jobs (Column 3). In Column 4, I control for the full set of control variables. The nationality coefficients stay qualitatively unchanged and coefficients on most of the control variables have the expected sign which is reassuring. For example, male job seekers receive fewer profile clicks and people with more years of experience or who are willing to live with employers receive more clicks. I estimate the effect of age non-parametrically by including dummies for four age groups which roughly correspond to age quartiles. I find a monotonic decline of clicks in age indicating that employers prefer young domestic workers. Overall, it is reassuring that nationality coefficients remain very stable when controlling for demographic and socio-economic variables.[16]

[16] While we focus on the immigrant status, there are additional interesting results, e.g. on references. 31% of profiles in the domestic work sector mention a reference in the search result. I divide these profiles into three groups: ads that were posted by the employer on behalf of their current worker (*Employer posts*), ads that list the phone number of a reference (*Refer.phone)* and those that only lists having a reference (*Reference*). Results in Column 4 show that just mentioning a reference is associated with 3.5% *fewer* profile clicks whereas providing a reference without a number has no effect. By contrast, job ads posted by current employers receive about 10% more clicks. These results suggest that employers tend to be suspicious and that references need to be credible in order to be effective Abel et al. (2017). In addition, many job seekers claim on their profiles to be 'hardworking', 'reliable' and 'trustworthy'. The effect of these claims on the number of clicks is very small and statistically insignificant (results not reported), possibly because these traits are not easily verifiable and may thus be regarded as 'cheap talk'.

4.1.2 Cross-sector comparison

As described in section 3, I collect data from three sectors: domestic work (Dom), nanny (Nan) services, and general work (Gen). These professions are very different with respect to the frequency and intimacy of interaction. Domestic workers have access to the personal living space and often live with the employer. The relationship with nannies is even more intimate as they take care of the employer's children. If employers dislike being around foreigners, we would expect discrimination to be largest in the nanny sector and smallest in the general work sector. To test how employer demand for foreigners varies, I estimate equation 1 separately for each sector. Coefficients reported in Table 3 provide support for this hypothesis. The effect of being Malawian or Zimbabwean is about 50% and 200% larger in the domestic work and nanny sector, respectively, compared to the general work sector. However, this test remains inconclusive since there may be other unobservable differences like reputation of different nationality that vary across sectors.

The coefficients on adding a picture to the profile also shows interesting cross-sectoral differences. While the effect is large and highly significant in each sector, it is 30% and 40% larger for the domestic work and nanny sector, respectively, compared to the general work sector. Taken at face value, this indicates that the increased level of familiarity and trust that pictures evoke are more important in sectors with a more intimate employer-employee relationship.[17]

[Table 3 here]

4.1.3 Robustness Test: Controling for Unobservables

While I observe the exact information as employers by the time they decide whether to click on a profile, one may be concerned that I fail to codify all relevant information as job seeker texts are multi-dimensional. To address this concern, I take advantatge of the fact that I observe *more* information than the employer by the time she makes the screening decision. Specifically, I observe the residential location of 2,120 job seekers (which is revealed on the profile page). Given South Africa's history of

[17]In a seperate analysis, I asked South Africans to rate these pictures according to two metrics: i) physical attractiveness and ii) to what extent the picture reflects the image of a domestic worker. The phycial attractiveness measure was a significant positive determinant of profile clicks whereas the second measure was negatively correlated. While these results warrant further investigation they suggest that some people may review job postings for reasons unrelated to hiring workers.

spatial segregation, residential locations are correlated with socio-economic variables. By observing where job seekers live, I can therefore control for factors unobserved by the employer. The rationale for this robustness test is as follows: assume that the residential location (z_i) is correlated with the number of profile clicks after controlling for other covariates, i.e. it was previously subsumed in the error term ($e_i = v_i + z_i$). Instead of specification (1), I now control for the residential location (z_i) non-parametrically by including dummies for each 0.5x0.5 grid cell:

$$y_i = \alpha + \beta_1 SA + \beta_2 For + \gamma X_i + \phi_j + time_t + lang_i + \mathbf{z_i} + v_i \tag{2}$$

This unobservable variable explains an additional 20% of variation in profile clicks compared to specification 1. The identification concern was that $E(For\, e_i) \neq 0$ leading to biased estimates of β_2. Comparing how much coefficient β_2 changes as one reduces the error term by controlling for z_i provides an indication to what extent initial results suffered from omitted variable bias. Estimating (1) and (2) separately and testing for equal coefficients, I cannot reject that the coefficients are identical (p-value: 0.52, results not reported). This supports the validity of identification strategy (1), although it is of course still possible that $E(For\, v_i) \neq 0$.

4.2 Identification from Applicant Pool Composition

4.2.1 Framework

One common criticism of audit studies is that they cannot provide information on the effect of employer discrimination in equilibrium (Heckman (1998)). If the share of discriminating firms is small relative to the share of the minority, than the differences in callback rates observed in these studies may only have a muted effect in equilibrium. As observed by Arrow (1973), it is the attitude of the *marginal* employer that may determine first order effects of discrimination. Charles and Guryan (2008) attempt to test the effect of discrimination on the marginal by measuring regional racist sentiments with data from the General Social Survey.

In order to derive predictions on the role of the applicant pool composition in determining the cost and prevalence of discrimination, I propose a simple framework in which the employer hires the applicant with the highest expected productivity. This framework builds on tournament models developed by Rosen (1981) and Lazear and Rosen (1981) and is most closely related to Charles and Guryan (2008).[18] The contribution of this study is to provide, to my knowledge, the first direct empirical test of the role of the applicant pool composition on discrimination.

Workers can be of type s (South African) or f (foreigners) and firms derive negative utility λ from employing type f. Let's assume that there is a fixed market wage $\bar{w}_s = \bar{w}_f = \bar{w}$ and that ability of all workers follows a uniform distribution with $a \subset [\underline{a}, \bar{a}]$. Within each population type, the firm can rank applicants according to ability a. I denote the ability level of the highest ranked applicant of each type as $a^{f*}$and a^{s*}. It is straightforward to see that the probability of both South Africans and foreigners to be highest ranked, $P[a_i^s > max(a_j^{s*}, a_j^{f*} - \lambda)], \forall j \neq i$ and $P[a_i^f - \lambda > max(a_j^{s*}, a_j^{f*} - \lambda)], \forall j \neq i$ respectively, increases in the share of foreigners, denoted θ_{For}, in the applicant pool.

4.2.2 Empirical Test

This framework offers an indirect test of discrimination. If employers penalize foreigners, then *both* immigrants and South Africans would benefit from being pooled with more foreigners and fewer South Africans. I construct θ_{For} by measuring the nationality composition of the five applicants posting directly before and the five posting after a given profile in the same suburb[19] (see Figure A.1) and estimate the following model:

$$y_i = \alpha + \beta_1 For + \delta_1 \theta_{For} + \delta_2 For * \theta_{For} + \gamma X_i + \phi_j + time_t + e_i \tag{3}$$

The predictions are that both δ_1 and $(\delta_1 + \delta_2)$ are positive. It is a priori unclear if foreigners or natives benefit more from being pooled with foreigners, i.e. $\delta_2 \lessgtr 0$.

[18] Cornell and Welch (1996) argue that labor market discrimination may be the result of employers' ability to extract productivity signals more accurately for applicants of their own type. By contrast, results from the present study suggest that employers display a taste for hiring natives. However, the implications of the following simple framework do not depend on the nature of the preference for a certain group.

[19] Computing the share θ off the previous and following five profiles is based on the fact that a total of about 10 profiles are shown on a typical computer screen. Given that only a small percentage of job seekers explicitly state they are South African and that employer preference did not differ significantly between South Africans and those not stating their nationality (Table 3), I pool these two groups for the analyses (and refer to them as SA) in this section to increase precision.

To discuss the identification assumption, imagine three job seekers ordered by the time they posted on the website.

$$y_{i-1} = \alpha + \beta For_{i-1} + \gamma X_{i-1} + e_{i-1}$$

$$y_i = \alpha + \beta For_i + \gamma X_i + e_i$$

$$y_{i+1} = \alpha + \beta For_{i+1} + \gamma X_{i+1} + e_{i+1}$$

Identifying discrimination solely from the pool composition requires that the nationality status of job seekers who just posted before or after person i is orthogonal to i's error term, i.e. $E[\sum_{j \neq i} For_j e_i] = 0$. This identification is more likely to be valid than the selection on observable strategy which requires $E[For_i e_i] = 0$ as the neighboring profiles are determined by who posted a profile minutes earlier or later.[20] Regressing the share of foreigners θ_{For} on a host of covariates provides support for the validity of the identification assumption: of the 21 covariates collected, only one (ability to drive) is significant at the 5% level. This suggests that the immigrant share displayed just above and below a profile is quasi-random.

Table 7 reports results of equation (3) estimated across the three sectors. Looking at the pooled sample, we see that job seekers indeed benefit from being in a pool with more foreigners (column 1) and that South Africans benefit more, although the difference is not statistically significant (2). Being pooled only with foreigners increases the number of profile clicks for natives by about 12.2% and for immigrants by 5.6%. While the results are generally consistent across sectors, the relative gain for natives and foreigners varies (4,6,8).

[Table 7 here]

Last, it is interesting to interpret the sign of the coefficients on the foreigner main effect (β_1). This estimate measures the effect of being the only foreign job seekers that the employer observes in the local applicant pool. Across sectors this coefficient is negative and statistically significant in the nanny and housekeeper sector. This finding is inconsistent with a model of heterogeneous race preferences within the employer population. For example, if 25% of employers had preferences for

[20]While, in theory, job seekers could post their profiles at strategic times (e.g. when they see few natives), this is practically infeasible as there is a delay of about 6 hours between submitting and posting of ads.

immigrants, we would expect a foreign job seeker who is pooled only with natives to receive *more* profile clicks. Taken at face value, the finding provides evidence that employer preferences with regards to nationality are not heterogeneous.

Empirical strategy 3 offers a robustness test for results from specification 2 as preference for South Africans is a necessary condition for the composition of the applicant pool to have an effect. A second necessary condition is that employers compare applicants locally, i.e. the ones presented to them at a given time on the computer screen. One implication of this assumption is that the share of foreigners presented on a previous or following search result page should matter less than the share presented on the same screen (θ_{For}^{1-5}). I test this by estimating equation 3 with the immigrant share of the applicants posted 6 to 10 places above and below a given profile (θ_{For}^{6-10}). The coefficient drops from 0.089 to 0.024 and is not statistically significant (not reported, p-value: 0.525).

5 Mechanisms

What may explain the preference for South African job seekers? There are at least three channels: i) employers risk paying a fine if they are caught hiring an immigrant who is undocumented, ii) employers may receive disutility from interacting with foreigners, and iii) employers may believe that South Africans are more productive with respect to unobservables (Appendix C offers a simple theoretical framework). To shed light on these questions I administered an anonymous phone survey with 208 domestic workers in Cape Town who posted a profile on gumtree (for more information on survey procedures, see Appendix D).

5.1 Legal Status and Bargaining Power

Employing an immigrant entails an expected cost as employers get fined if the employment contract is monitored and the immigrant does not have a valid work permit.[21] The data allows to explicitly test for the role of the legal status as about 8% of immigrants state in the search result that they have a work permit. One straightforward prediction of this mechanism is that within the group of

[21] As per the Labour Act (Section 49), "*anyone who knowingly employs an illegal foreigner or a foreigner in violation of this Act shall be guilty of an offence and liable on conviction to a fine or to imprisonment not exceeding one year*."

immigrants, those with work permits receive more profile clicks. I test this hypothesis by including a dummy for workpermit (*permit*) in specification (1).

Column 1 in Table 3 shows that among foreigners having a work permit in significantly *negatively* correlated with the number of profile clicks in the pooled sample. Stating that one posesses a work permit reduces the number of profile clicks by about 7 percent.[22] While these findings are surprising at first glance, they are consistent with a theory that employers prefer job seekers that are exploitable. The DoL (2007) speculates about the reasons for anecdotal evidence showing lower unemployment among immigrants: "*The usual advantages of irregular employment (low wages, vulnerability, exploitative conditions) may be at the core of this preference* [for non-South African workers]".

Consistent with this exploitation hypothesis is the finding that job seekers stating that they are willing to live with employers or expressing that they need a job urgently receive more profile clicks (Table 3). An alternative plausible theory is that employers infer from reading that somebody holds a work permit that the person demands a higher wage and reading that somebody is 'urgently' looking for work may signal that she is less likely to shirk on the job. However, results from the anonymous phone survey do not support this explanation. Undocumented immigrants do not have lower reservation wages are not willing to work longer hours than documented immigrants (Table A.3).

Results in Table A.3 point to an alternative explanation for the preference of undocumented over documented immigrants: bargaining power. South Africa instituted the Commission for Conciliation, Mediation and Arbitration (CCMA), an administrative tribunal at which employers can bring cases against employer mistreatment at no charge. If disputes are not resolved and no arbitration is reached, cases are referred to labour courts. About 70% of the formal workforce fall under the juristriction of the CCMA. The CCMA is very visible in the media and widely used with about 120,000 cases per year, of which about 70% are about unfair dismissal (Bhorat et al. (2009)). Undocumented immigrants are significantly less likely to know about the CCMA (62.8% vs. 78.9%) and less willing to take their employer to court (57.7% vs. 70.1%) than documented immigrants.[23]

[22]This finding also provides evidence against the explanation that employers prefer South Africans because they are more likely to remain in the Western Cape area since, controlling for other factors, immigrants with a work permit can be expected to be more established and thus stay in South Africa for longer.

[23]It is unclear whether undocumented immigrants have access to the CCMA. The case of Discovery Health Ltd v CCMA & others sided with an immigrant whose work permit renewal process was pending.

Interestingly, this lower bargaining power does not translate into lower actual daily or hourly wages. It is also notable that undocumented and documented immigrants do not report to be treated differently by employers. And while immigrants are more likely than South Africans to report incidents of lower than agreed payments and rude behavior from employers they report positive overall treatment by employers (2.5 on a 0-3 scale), low prevalences of maltreatment (5-10%), and feel comfortable to ask for time off if sick (96%).

However differences in bargaining power cannot explain the overall preference for South Africans.They are significantly more likely to know about the CCMA (93%) and take their employer to court (95%), are more likely to hold a have a written contract (33%) and have negotiated wages (52.4%, not significant). The next section will explore whether the results can be explained by different *expectations* about the productivity of native domestic workers.

5.2 Statistical vs. Taste

To shed light on the nature of discrimination, I build on insights by Farber and Gibbons (1996) and Altonji and Pierret (2001): if firms use nationality as a proxy for productivity, then the importance of nationality should decrease as other predictors of productivity become available and are factored into employer beliefs. By contrast, if employers have the same productivity expectation of South Africans and foreigners but prefer hiring locals due to *taste*-based discrimination, we would expect the revelation of additional information to have no effect on the foreign-national gap in hiring decisions. (For a formal employer learning model see 7.)[24]

I first predict the number of clicks ($\hat{y}_i$) in the pooled sample using covariate vector X_i. The residualized numbrt of profile clicks ($y_i - \hat{y}_i$) is regressed on nationality dummies interacted with a measure of the number of relevant information (I) that the search result includes[25]:

$$y_i - \hat{y}_i = \alpha + \beta_1 SA + \beta_2 For + \delta_1 I * SA + \delta_2 I * For + \rho I + e_i \tag{4}$$

[24]Similar tests of the nature of discrimination have been conducted through audit studies, e.g. by Baert and De Pauw (2014). Neumark (2016) points out that it is unclear whether the information content that is varied is relevant to employers.

[25]All variables are categorized to be relevant that have a an absolute t-value>1 in specification (1). Relevant information includes gender, age, experience, references, pictures, workpermit, drivers license, and whether a person is willing to sleep in. The average profile has information on 2.9 of these variables. While applicants who don't provide information on their nationality reveal on average 2.6 characteristics, more information is revealed by South Africans (3.3), Zimbabweans (3.4) and Malawians (3.1).

The intuition for this test is as follows: imagine there are two binary characteristics (X_1 and X_2) that equally determine productivity. Individual A has a positive signal for X_1 and a negative for X_2 while person B does not provide any information on X_1 and X_2. While both may have the same predicted productivity, person A's profile is more informative. If employers statistically discriminate we would expect employer demand to converge as more information (I) becomes available. Given that the previous analysis found that foreigners receive fewer clicks than South Africans, we would thus expect δ_2 to be positive and δ_1 to be negative. Table 5 shows results of specification (4) estimated for the pooled sample and each sector separately.

[Table 5 here]

Results are consistent across most specifications. The interaction term of nationality and information (I) is positive for South Africans and negative for Malawians and Zimbabweans (columns 2, 4, 6, 8). While providing more information is beneficial for job seekers without nationality information and especially South Africans, immigrants do not benefit (sum ρ and δ_i): at the average level of profile information, job seekers without nationality information and South Africans receive an additional 0.57 (7.7%) and 1.35 (18.2%) clicks, respectively, while demand for immigrants is unaffected. The p-values reported show that these differences in coefficients (δ_1 and δ_2) are statistically significant at the 5% level in all but the general work specification. These estimates suggest that the number of profile clicks is *diverging* between South Africans and foreigners as more information becomes available to employers. Results are consistent with a model in which employers believe that foreigners are on average more productive than natives. However, in the aggregate this *positive* statistical discrimination is offset by the effect of taste-based discrimination.

This conclusion is supported by previous studies concluding that immigrants are *more* employable (Crush (2008)) and by evidence from the domestic worker survey: immigrants are more educated and younger than South Africans. (Section 4 found that employer prefer younger job seekers.)

6 Supply Side Response to Discrimination: Spatial Analysis

Models of discrimination and job search can be divided into discrimination with random search and discrimination with targeted search. In random search models, the discriminated groups have lower reservation wages and accept jobs with lower match quality given that they expect to receive fewer job offers. In models of targeted search, job seekers decide where to apply after observing firms' wage offers which provide information on the level of discrimination. The question whether job seekers adjust their search strategy in response to discriminatory practices of firms has important implications for how to interpret empirical results in the existing literature. In particular, Heckman (1998) points out that audit studies make the critical assumption that workers employ random search strategies. If job search is costly and workers are able to direct applications to non-discriminatory firms, this should increase their reservation wage, improve match quality and thus reduce racial wage gaps in equilibrium.

Partly due to a lack of data, there is very little evidence whether discrimination induces people to direct their job search. South Africa provides a context conducive to testing these questions. One of the legacies of apartheid is that workers tend to live far away from where jobs are located.[26] The average distance of African townships from the central business districts (CBDs) of the seven largest South African cities is 28 km. Combined with relatively high costs of public transport, this results in high search and commuting costs both monetary and time-wise.[27]

While it is widely argued that spatial segregation is a barrier to finding employment, there is little evidence on where people are looking for jobs. Does high transport cost induce them to only look for jobs in their vicinity or are they willing to accept higher commuting expenses and search in areas with more employment opportunities? Are immigrants more willing to travel further for work and, if so, is this decision linked to the level of discrimination they are facing? The online job advertisement data can shed some light on this question since job seekers have to specify the suburb where they are searching and about half provide the address or postal code. In total, I collect location data of about 3,000 job seekers in the Cape Town area.

[Figure 2 here]

[26] In an attempt to claim city centers and marginalize the black and coloured population, the apartheid regime forcefully removed large parts of the urban population to homelands or township outside urban areas; the most famous cases include Sophiatown in Johannesburg and District 6 in Cape Town. Townships suffered from poor infrastructure and provided few employment opportunities as they were located far from business and industry.

[27] Recent evidence shows that people with employment spend on average R215 (7.3% of net salary) per month on transport to and from work (SALDRU (2009)). The mean amount spent by the actively searching unemployed on transport costs related to the job search was R105.75 in the previous week. 42.5% of the unemployed refrain to local job search and report not spending anything.

Figure 2 shows a map of the Cape Town region and the location of job seekers categorized by nationality.[28] It also displays the centroid of the nine suburbs where job seekers can indicate they are searching for employment. A few facts stand out: the distribution of job seekers roughly aligns with the population density of the urban area (with the notable exception of the poorer Cape Flats area), which lends support to the claim that the website is widely used. Second, there is no clear spatial pattern of nationality. However, one may still be concerned that difference in search behavior between immigrants and natives is linked to unobservable differences that are correlated with where people live. For example, if natives live in more central parts of the city, they would mechanically have shorter commuting distances. Conversely, if natives have better access to public transport we would expect them to search further away. These concerns are particularly important in places with large informal settlements like Cape Town where the residential location of people is linked to socio-economic factors (Hellerstein and Neumark (2008)).[29]

To address these concerns and control for unobservable differences correlated with the residential location, I draw a 0.25 x 0.25 degree gridnet (approximately 2.5 x 2.5km) and employ an empirical strategy using job seeker location fixed effects following Black et al. (2013) (Figure 2). Coefficients are estimated from variation in nationality within a grid cell with standard errors clustered at the grid cell level. First, I compare how the distance between residential location and the centroid of the suburb where job seekers look for work varies between nationalities (Panel A, Table 6). Columns 1 and 2 indicate that immigrants search for work in places that are significantly further away from their residential area: target suburbs of foreigners are on average 2.2km further compared to the pooled group of South Africans and those that do not indicate their nationality. These differences are economically meaningful given the average search distance of 10.4km. When I control for location fixed effects, coefficients decrease in magnitude but remain statistically significant. One caveat with the location fixed effect strategy is that this limits the sample used to estimate effects. I find that results are robust to extending grid cells to 0.5 x 0.5 degrees which the sample used for identification (results not reported).

[Table 6 here]

[28] To guarantee anonymity of job seekers, random noise (0.5km) is applied to the location.

[29] Previous studies pointed out that employers may discriminate based on the location where job seekers are living (Bertrand and Mullainathan 2004, Rathelot 2014). The previous results cannot be explained by this as the location of the job seeker is only revealed once employers click on their profile.

One explanation is that job seekers may decide to search further away if they are faced with discrimination in their suburb. Location subgroup analysis of specification (1) suggests that levels of discrimination vary considerably by suburb - both towards immigrants in general and between immigrant groups in a given suburb. For example, in the Northern Suburbs, Malawian profiles receive 19.3% fewer clicks than those not revealing their nationality compared to 1.6% fewer clicks for Zimbabweans. Conversely, Zimbabweans get 9% fewer clicks in Somerset West compared to 2% for Malawians. This may be the result of spatial stratification within the white population and the large historic, political and cultural differences between whites from Dutch and English descent.[30] This raises the question whether discrimination is one of the factors that determine search patterns and could explain why immigrants search in places further away.

To shed light on this question I next test whether nationality-specific discrimination faced in the suburb of residence affects search behavior. I estimate the following regression:

$$y_i = \alpha + \delta\hat{\sigma}_{j,n} + \gamma X_i + time_t + e_j \tag{5}$$

y_i is an indicator variable capturing whether a workers searches for work outside her suburb of residence. Variable $\hat{\sigma}_{j,n}$ measures the suburb and nationality specific coefficients of discrimination estimated from equation 1. It is normalized so a higher value presents more discrimination and measures how many fewer percent profiles of nationality n receives in suburb j. To account for the facct that $\hat{\sigma}_{j,n}$ is a generated regressor, I compute standard errors (clustered at the grid cell level) through bootstrapping.

Results reported in Panel B provide suggestive evidence that discrimination in the home suburb induces migrants to search in other areas. The coefficient δ is positive and significant in column 6 and 7. Column 8 reports marginal effects estimated from a probit model. To interpret the magnitude of the coefficients, it helps to keep in mind that the aggregate coefficient of σ_n is around 0.1 (Section 4). Discrimination faced on the home market thus increases the probability that immigrants search

[30] English-speaking whites live predominantly in the southern suburbs while Afrikaans-speakers more frequently live in the northern suburbs. The imaginary dividing line between northern and southern suburbs is referred to by locals as the '*boerewors curtain*' in reference to the Afrikaans word for sausage. Furthermore, spatial stratification is reflected by the fact that the most popular newspaper in the "Die Burger" in the northern suburbs and "The Cape Argus" in the southern suburbs.

in different suburbs on average by about 8 percentage points which is sizable given the mean value of 28.1% among natives.

At face value, estimates provide some of the first evidence that a group adjusts job search behavior in response to discrimination faced by employers. This has important implications for how to measure the cost discrimination. In particular it would imply that estimates of discrimination observed in Section 4 would underestimate the total cost of discrimination since they are estimated in equilibrium after immigrants adjusted their search behavior. The additional cost of discrimination stems from the transport cost of searching and working in places further away. However, several important caveats should be kept in mind when interpreting these results. First, coefficients ($\hat{\sigma}_{j,n}$) are estimated from samples of only a few hundred job seekers. Secondly, location choices are endogeneous and native and migrant job seekers living in the same location may differ along unobservable characteristics correlated with job search behavior. Rathelot (2013) notes that if immigrants are penalized on the housring or labor market, the marginal immigrant to participate in the labor force should be more qualified than the marginal native. However, this should lead to *lower* commuting rates among immigrants as they have an advantage over native competitors in the same market which should reduce their inclination to search in different suburbs.

Can job seekers target suburbs with lower discrimination? To shed light on this question, I look at the subset of 1,073 commuters and measure the difference in discrimination between the home and target suburb. I find that on average levels of discrimination in the home suburb are only marginally higher (difference: 0.005 or 0.07 standard deviations, not significant). Taken at face value, this suggests that people have information on discrimination in their place of residence but that this is more difficult to assess for suburbs where one is searching for jobs.

7 Concluding Remarks

This study uses a unique data set of job seekers using a free job advertisement website in South Africa's Western Cape province. This is to my knowledge the first study of this kind in Africa, a continent in which the rapidly increasing internet access offers new opportunities to create and facilitate markets. In particular, internet fora like gumtree offer an opportunity to match employers with job seekers in spatially segregated urban areas and reduce the reliance of firms to hire through social networks, currently the most common form of hiring in most developed countries (Beaman and Magruder (2012)).

One may wonder why such a large share of job seekers reveal that they are immigrants on their profile page if they get penalized by employers. One explanation is that employers can easily verify job seekers' nationality (as they typically require to see the national ID card). It is therefore beneficial for immigrants to reveal their nationality upfront to avoid going through the hiring process with discriminating employers.

One limitation of the study is that we know relatively little about the identity of potential employers using the website. From the survey we know that 96% were employed by a private household. In the Western Cape, these employers are predominantely White (63%) compared to Blacks (27%) and Coloured (10%). Whether different race groups have differential preferences over hiring immigrants vs. natives should be the subject of future research. A second limitation of the data set is that I can at best estimate discrimination at the first stage of the hiring process.Yet, in a country like South Africa with excess supply of labor and employers often reporting that they receive hundreds of applications for a single job, this first screening likely plays an important role in explaining employment outcomes.

References

Abel, M., R. Burger, and P. Piraino (2017). The value of reference letters: Experimental Evidence from South Africa. *Working Paper*. 10

Altonji, J. and C. Pierret (2001). Employer learning and statistical discrimination. *The Quarterly Journal of Economics* (February 2001). 3, 17, 38

Arrow, K. (1973). The Theory of Discrimination. In *Discrimination in Labor Markets*, pp. 3–33. 2, 12

Arrow, K. (1998). What has economics to say about racial discrimination? *Journal of Economic Perspectives 12*(2), 91–100. 4

Baert, S. and A.-S. De Pauw (2014). Is ethnic discrimination due to distate or statistics? *Economics Letters 125*(2), 1061–72. 17

Bartos, V., M. Bauer, J. Chytilova, and F. Matejka (2016). Attention Discrimination: Theory and Field Experiments. *The American Economic Review 106*(6), 1437–1475. 4

Beaman, L. and J. Magruder (2012). Who gets the job referral? Evidence from a social networks experiment. *American Economic Review 102*(7). 23

Becker, G. (1957). *The Economics of Discrimination.* University of Chicago Press. 2, 3

Bertrand, M. and E. Duflo (2016). Field Experiments on Discrimination. In *Handbook of Economic Field Experiments.* 1

Bertrand, M. and S. Mullainathan (2003). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. (1996). 3, 4

Bhorat, H., K. Pauw, and L. Mncube (2009). *Understanding the Efficiency and Effectiveness of the Dispute Resolution System in South Africa : An Analysis of CCMA Data.* Number May. Development Policy Research Unit Working Paper 137. 16

Black, D. (1995). Discrimination in an equilibrium search model. *Journal of labor Economics 13*(2), 309–334. 2

Black, D. A., N. Kolesnikova, S. G. Sanders, and L. J. Taylor (2013). The role of location in evaluating racial wage disparity. *IZA Journal of Labor Economics 2*(1). 20

Charles, K. and J. Guryan (2008). Prejudice and wages: an empirical assessment of Becker's The economics of discrimination. *Journal of Political Economy 116*(5), 773–809. 12, 13

Cornell, B. and I. Welch (1996). Culture Information and Screening Discrimination. *Journal of Political Economy 104*(3), 542–571. 13

Crush, J. (2008). *The perfect storm: The realities of Xenophobia in comtemporary South Africa.* Institute for Democracy in South Africa. 5, 6, 18

Crush, J. (2011). Complex Movement, confused responses: Labour migration in South Africa. 6

Dinat, N. and S. Peberdy (2007). Restless worlds of work, health and migration: domestic workers in Johannesburg. *Development Southern Africa 24*(1), 186–203. 8

Dinkelman, T. and V. Ranchod (2012). Evidence on the impact of minimum wage laws in an informal sector: Domestic workers in South Africa. *Journal of Development Economics 99*(1), 27–45. 8

DOL (2007). Labour Migration and South Africa: Towards a fairer deal for migrants in the South African Economy. 6

Farber, H. and R. Gibbons (1996). Learning and wage dynamics. *The Quarterly Journal of Economics* (November). 17, 38

Heckman, J. (1998). Detecting discrimination. *Journal of Economic Perspectives 12*(2), 101–116. 1, 2, 12, 19

Hellerstein, J. and D. Neumark (2008). Workplace segregation in the United States: Race, ethnicity, and skill. *The Review of Economics and Statistics 9*(3), 459–477. 20

Horton, J. J. (2010). *Online labor markets.* Springer Berlin Heidelberg, 2010. 4

Kerr, A. (2015). Tax(i)ing the poor? Commuting costs in South Africa. 3

Kuhn, P. and K. Shen (2013). Gender discrimination in job ads: Evidence from china. *The Quarterly Journal of Economics*, 287–336. 4

Lang, K. and J.-y. K. Lehmann (2012). Racial Discrimination in the Labor Market : Theory and Empirics. *Journal of Economic Literature 50*(4), 959–1006. 2, 3

Lazear, E. and S. Rosen (1981). Rank-order tournaments as optimum labor contracts. *Journal of Political Economy*, 841–864. 13

Neumark, D. (2016). Experimental research on labor market discrimination. *NBER Working Paper 22022*. 1, 17

Oreopoulos, P. (2011). Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment with Six Thousand Resumes. *American Economic Journal: Economic Policy 3*(4), 148–171. 3

Pager, D. (2007, jan). The Use of Field Experiments for Studies of Employment Discrimination: Contributions, Critiques, and Directions for the Future. *The ANNALS of the American Academy of Political and Social Science 609*(1), 104–133. 1, 4

Phelps, E. (1972, feb). The Statistical Theory of Racism and Sexism. *American Economic Review 62*(4), 659–661. 3

Pope, D. and J. Sydnor (2011). What is in a Picture? Evidence of Discrimination from Prosper. *Journal of Human Resources 46*(1), 53–92. 4

Rathelot, R. (2013). Ethnic differentials on the labor market in the presence of asymmetric spatial sorting: Set identification and estimation. *Mimeo CREST*. 22

Riach, P. and J. Rich (2002). Field Experiments of Discrimination in the Market Place. *The Economic Journal 29*(5), 1–12. 4

Rich, J. (2014). What do field experiments of discrimination in markets tell us? A meta analysis of studies conducted since 2000. *IZA Discussion Paper No.8584*. 1

Rosen, S. (1981). The economics of superstars. *American Economic Review*, 845–858. 13

SALDRU (2009). National Income Dynamics Study, Wave 1. 19

Seekings, J. and N. Nattrass (2008). *Class, race, and inequality in South Africa.* Yale University Press, 2008. 5

Tables

Table 1: Housekeeper characteristics by nationality

	Mean	N	Nationality				
			NoInf	SA	Malawi	Zimbabw	p-value
Nr profile clicks	6.61	3220	7.02	7.19	6.21	6.65	.001
Female	.515	3220	.322	.779	.509	.809	0
Male	.194	3220	.107	.221	.313	.092	0
Report age	.283	3220	.237	.516	.21	.474	0
Age (yr)	30.2	912	30.8	32.6	28.6	30.7	0
Report experience	.448	3220	.322	.484	.58	.385	0
Experience (yr)	3.95	527	4.87	5.42	3.47	3.61	.007
Reference	.328	3220	.296	.495	.336	.341	.013
Refer. phone nr	.024	3220	.026	.021	.022	.024	.948
Employer posts	.089	3220	.165	.063	.051	.04	.407
Workpermit	.048	3220	.03	.	.036	.101	0
Sleep in	.164	3220	.143	.105	.164	.207	.006
Able to drive	.041	3220	.028	.032	.068	.009	0
Certificate	.022	3220	.023	.011	.024	.019	.423
Virtuout traits	.412	3220	.352	.368	.422	.498	.002
Nr words	30.4	3220	30.1	33.9	30.4	30.2	0
Picture	.045	3220	.07	.063	.027	.036	.182
Punctuation	.305	3172	.297	.361	.302	.231	025
Wrong grammar	.132	3173	.163	.18	.127	.052	0
Nr Mistakes	.162	3174	.166	.131	.147	.144	0.971
Capitalized	.104	3174	.078	.098	.121	.085	.222
N			1138	102	1422	715	

Note: The table reports mean profile characteristics of job seeker profiles compared across nationalities. P-values are reported of a test of equal means across nationality groups.

Table 2: Housekeeper Analysis (dep var: Log Profile Clicks)

	(1)	(2)	(3)	(4)	(5)
South African	0.0755	0.0306	0.0195	0.0353	0.0327
	(0.0476)	(0.0395)	(0.0398)	(0.0410)	(0.0410)
Zimbabwian	-0.0337	-0.0413*	-0.0463**	-0.0409*	
	(0.0237)	(0.0214)	(0.0212)	(0.0225)	
Malawian	-0.109***	-0.105***	-0.104***	-0.0846***	
	(0.0201)	(0.0185)	(0.0191)	(0.0199)	
Foreign					-0.0689***
					(0.0180)
Male				-0.0572***	-0.0639***
				(0.0216)	(0.0213)
Female				0.0176	0.0220
				(0.0174)	(0.0174)
Age: <25				0.0332	0.0386
				(0.0332)	(0.0331)
Age: 26-30				-0.0217	-0.0161
				(0.0226)	(0.0226)
Age: 31-35				-0.0222	-0.0125
				(0.0345)	(0.0342)
Age: 35-60				-0.0557	-0.0455
				(0.0384)	(0.0378)
Experience (yr)				0.0129*	0.0132*
				(0.00715)	(0.00715)
Reference				-0.0327*	-0.0325*
				(0.0171)	(0.0171)
Employer posts				0.00163	0.00187
				(0.0467)	(0.0467)
Refer. phone				0.0993***	0.100***
				(0.0320)	(0.0321)
Live w employer				0.0833***	0.0845***
				(0.0223)	(0.0223)
Picture		0.446***	0.455***	0.438***	0.442***
		(0.0449)	(0.0453)	(0.0448)	(0.0445)
Punctuation		-0.0282	-0.0241	-0.0238	-0.0203
		(0.0217)	(0.0219)	(0.0219)	(0.0218)
Wrong grammer		-0.0663**	-0.0676**	-0.0583*	-0.0586*
		(0.0337)	(0.0338)	(0.0341)	(0.0341)
Nr Mistakes		-0.0137	-0.0140	-0.0182	-0.0150
		(0.0247)	(0.0246)	(0.0252)	(0.0252)
Profile Syntax	N	Y	Y	Y	Y
Suburb	N	N	Y	Y	Y
Control Var	N	N	N	Y	Y
R^2	0.011	0.198	0.204	0.216	0.217
N	3218	3218	3217	3217	3217
p-value: Zim=SA	0.028	0.083	0.115	0.050	
p-value: Mal=SA	0.000	0.001	0.002	0.003	
p-value: For=SA					0.012

Notes: standard errors in parentheses

Base group for nationality are job seekers who do not reveal their nationality. *Employer posts* refers to whether a former employer posts on their behalf. *Reference phone* captures whether a phone nr of a reference is provided. *Reference* captures any other forms of references mentioned.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Cross Sector Analysis

	Full Sample	Housekeeper	Nanny	General
	(1)	(2)	(3)	(4)
South African	0.0335	0.0324	0.0688	0.00791
	(0.0337)	(0.0411)	(0.0814)	(0.0830)
Foreign	-0.0753***	-0.0670***	-0.135***	-0.0252
	(0.0141)	(0.0180)	(0.0322)	(0.0397)
Male	-0.0120	-0.0648***	0.0672	0.0415
	(0.0163)	(0.0213)	(0.0446)	(0.0344)
Female	-0.00235	0.0224	-0.0682**	-0.0126
	(0.0143)	(0.0174)	(0.0318)	(0.0650)
Age: <25	0.0408	0.0374	-0.001000	0.0470
	(0.0249)	(0.0332)	(0.0454)	(0.0745)
Age: 26-30	-0.0145	-0.0172	0.0190	-0.0384
	(0.0195)	(0.0226)	(0.0423)	(0.0595)
Age: 31-35	-0.0155	-0.00951	0.0138	-0.118
	(0.0282)	(0.0341)	(0.0556)	(0.103)
Age: 36-60	-0.0295	-0.0443	-0.0410	0.0960
	(0.0307)	(0.0376)	(0.0522)	(0.147)
Experience (yr)	0.00138	0.0131*	-0.00586	-0.0148
	(0.00545)	(0.00716)	(0.00957)	(0.0144)
Live w Employer	0.0661***	0.0848***	0.0388	0.0303
	(0.0185)	(0.0223)	(0.0351)	(0.0870)
Workpermit	-0.0714**	-0.0537	-0.0720	-0.0323
	(0.0295)	(0.0390)	(0.0615)	(0.0652)
Urgency	0.102**	0.169***	0.0919	0.0163
	(0.0501)	(0.0607)	(0.0950)	(0.112)
Picture	0.427***	0.442***	0.478***	0.341***
	(0.0258)	(0.0444)	(0.0405)	(0.0505)
R^2	0.266	0.217	0.320	0.191
N	5338	3217	1133	988
p-value: For=SA	0.002	0.014	0.019	0.740

Notes: Robust standard errors parentheses

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

Table 4: Pool Composition

	Full Sample		Housekeeper		Nanny		General Work	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Foreign	-0.0882***	-0.0554	-0.0764***	0.0361	-0.135***	-0.191**	-0.0403	0.0450
	(0.0139)	(0.0324)	(0.0171)	(0.0555)	(0.0321)	(0.0669)	(0.0382)	(0.0652)
Foreign Share (θ_{For})	0.0899**	0.122**	0.0144	0.124	0.107	0.0456	0.0639	0.166
	(0.0334)	(0.0458)	(0.0472)	(0.0718)	(0.0900)	(0.117)	(0.0733)	(0.104)
Foreign x Foreign Share		-0.0636		-0.185*		0.167		-0.202
		(0.0551)		(0.0856)		(0.160)		(0.127)
Profile Syntax	Y	Y	Y	Y	Y	Y	Y	Y
Suburb	Y	Y	Y	Y	Y	Y	Y	Y
Control Var	Y	Y	Y	Y	Y	Y	Y	Y
R^2	0.265	0.265	0.211	0.212	0.316	0.316	0.182	0.184
N	5338	5338	3217	3217	1133	1133	988	988

Note: Robust standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Statistical Discrimination analysis

	Full Sample		Housekeeper		Nanny		General Work	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
South African	0.558	-0.935	0.246	-2.031**	1.929	-8.333*	-0.815	-0.0466
	(0.369)	(0.946)	(0.269)	(0.748)	(1.085)	(3.941)	(0.518)	(2.292)
Zimbabwian	-0.538***	-0.0695	-0.0247	0.0429	-1.671***	-0.903	0.625	0.338
	(0.160)	(0.436)	(0.180)	(0.410)	(0.300)	(0.910)	(0.708)	(2.459)
Malawian	-0.841***	-0.302	-0.265	0.0770	-1.843***	-0.656	0.437	0.1000
	(0.141)	(0.346)	(0.154)	(0.404)	(0.361)	(1.100)	(0.377)	(0.875)
Information (I)		0.197**		0.0695		0.237		-0.0494
		(0.0710)		(0.0954)		(0.131)		(0.157)
SA x Inform (I)		0.269		0.458**		2.494*		-0.176
		(0.211)		(0.175)		(1.065)		(0.556)
Zim. x Inform (I)		-0.194		-0.0468		-0.279		0.103
		(0.107)		(0.114)		(0.216)		(0.677)
Mal. x Inform (I)		-0.210**		-0.112		-0.359		0.105
		(0.0957)		(0.118)		(0.242)		(0.250)
Profile Syntax	Y	Y	Y	Y	Y	Y	Y	Y
Suburb	Y	Y	Y	Y	Y	Y	Y	Y
Control Var	Y	Y	Y	Y	Y	Y	Y	Y
R^2	0.008	0.010	0.001	0.003	0.035	0.050	0.003	0.003
N	5340	5340	3219	3219	1133	1133	988	988
p-value: *Zim x I=SA x I*		0.044		0.002		0.012		0.806
p-value: *Mal x I=SA x I*		0.028		0.001		0.009		0.637

Notes: Robust standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Spatial Job Search Analysis

Panel A: **Job Search *Distance***

	y = distance to target suburb (in m)				y=log(distance)	
	(1)	(2)	(3)	(4)	(5)	(6)
Malawian	2850.2***		1159.8**		0.0769	
	(997.2)		(485.2)		(0.0554)	
Zimbabwian	955.2		679.3*		0.103*	
	(700.7)		(383.1)		(0.0538)	
Foreigner		2243.3***		958.8***		0.0902**
		(724.7)		(343.6)		(0.0437)
Covariates	Y	Y	Y	Y	Y	Y
Location F.E.	N	N	Y	Y	Y	Y
R^2	0.043	0.037	0.440	0.440	0.658	0.658
N	2984	2984	2984	2984	2984	2984
Mean (natives)	10391	10391	10391	10391		

Panel B: **Job Search and *Home Market Discrimination***

	y: 1=Search outside home suburb		marginal effects
	(6)	(7)	(8)
Suburb Discrimination	0.875***	0.824***	0.831**
	(0.299)	(0.314)	(0.373)
Covariates	N	Y	Y
N	2940	2940	2940
Mean (natives)	0.281	0.281	0.281

Notes: Standard erros (clustered at the grid cell level) in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Panel B estimates standard errors through bootstrapping (200 repetitions) to account for the fact that the 'suburb discrimination' variable is estimated.

Figures

Figure 1: Distribution of Profile Clicks

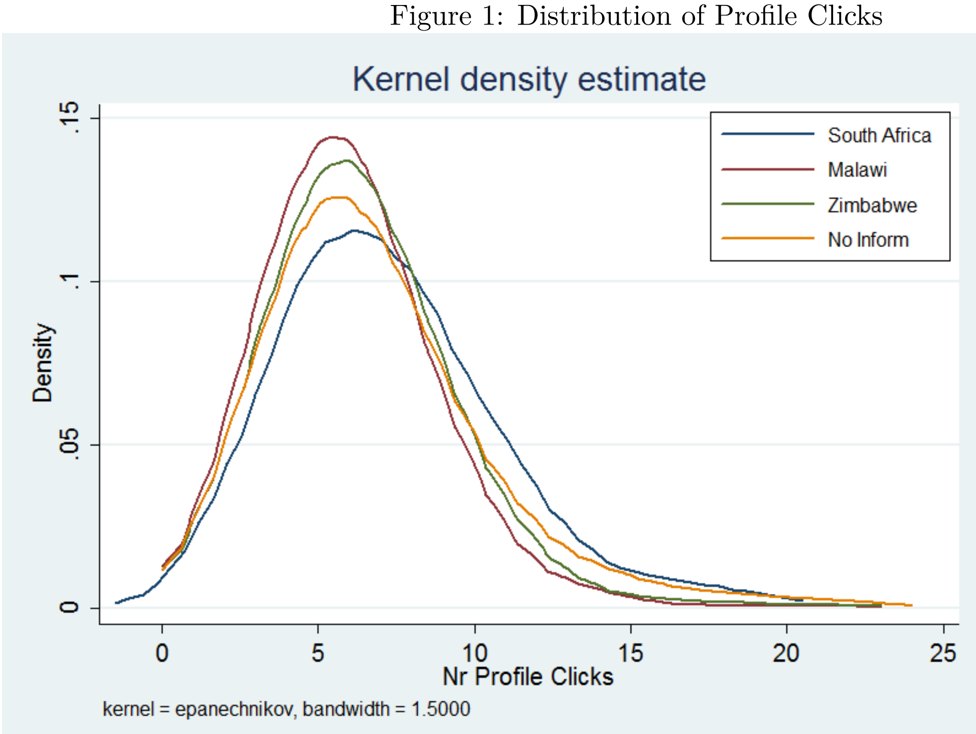
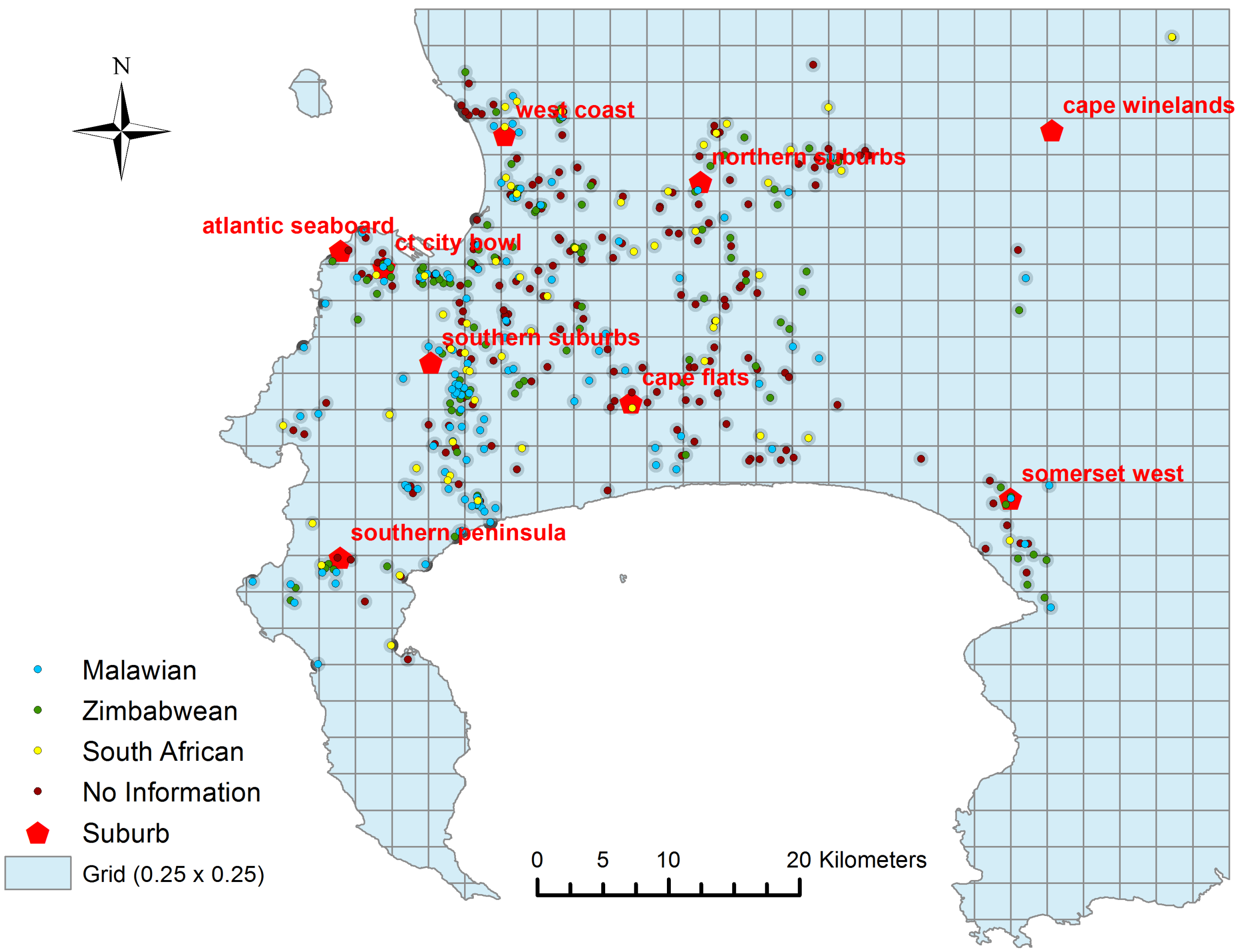


Figure 2: Spatial Analysis of Job Search (Cape Town)



Appendix

A. Figures

Figure A.1: Job Website: Search Result

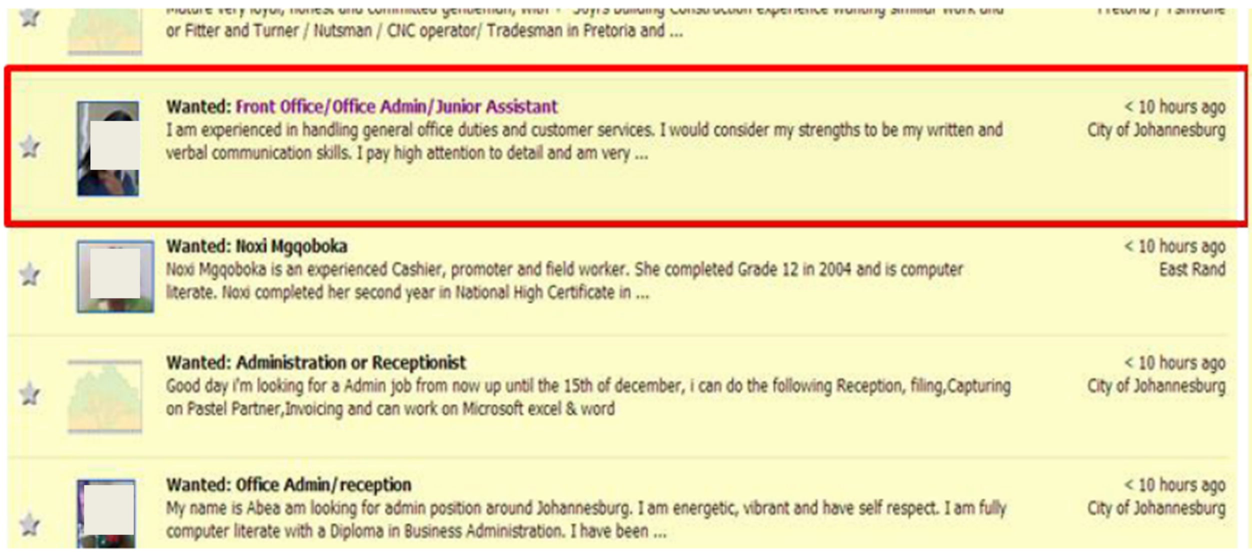
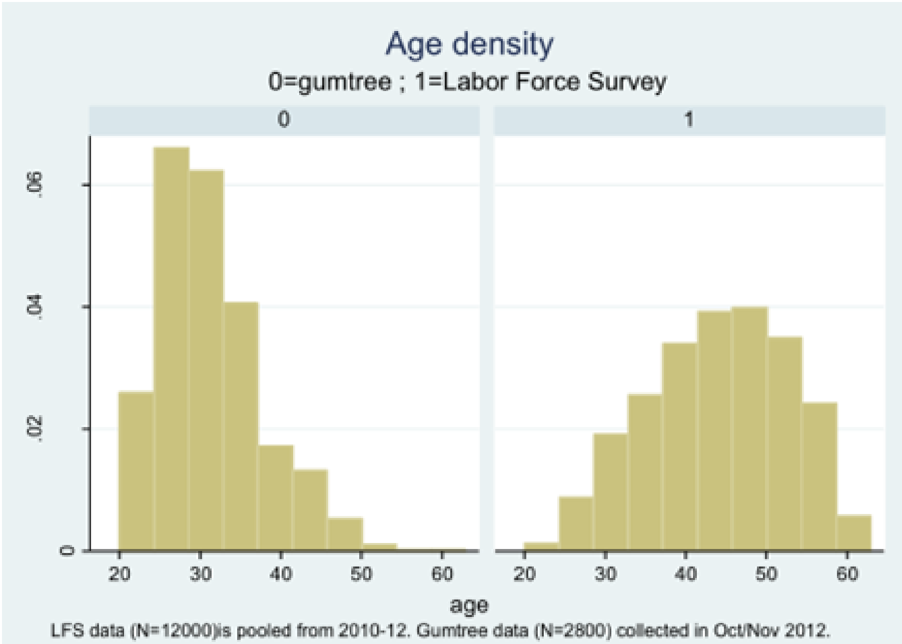


Figure A.2: Job Website: Profile Page

Figure A.3: Age Distribution: Job website vs. Labor Force Survey



B. Tables

Table A.2: Robustness Test: Short vs Long profiles

	Full Sample		Housekeeper		Nanny		General Work	
	Long	Short	Long	Short	Long	Short	Long	Short
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Zimbabwe	-0.0924***	-0.0303	-0.0535	-0.0260	-0.139***	-0.0966	-0.0625	-0.0373
	(-3.62)	(-1.11)	(-1.57)	(-0.83)	(-2.82)	(-1.54)	(-0.71)	(-0.34)
Malawi	-0.144***	-0.0496**	-0.119***	-0.0522*	-0.195***	-0.0900	-0.0406	-0.0180
	(-6.39)	(-2.06)	(-4.17)	(-1.82)	(-3.70)	(-1.33)	(-0.60)	(-0.27)
R^2	0.303	0.161	0.254	0.156	0.328	0.245	0.259	0.176
N	2871	2384	1571	1609	774	340	526	435

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.1: Job seeker characteristics by sector

	Mean	N	Sector: Housek	Sector: Nanny	Sector: General	p-value
Nr profile clicks	7.41	5341	6.61	9.66	7.43	0
South African	.035	5341	.03	.045	.04	.038
Malawian	.356	5341	.42	.15	.381	0
Zimbabwian	.189	5341	.21	.208	.101	0
No nation. info	.427	5341	.344	.599	.5	0
1=female	.429	5341	.515	.502	.065	0
1=male	.234	5341	.194	.128	.488	0
report age	.291	5341	.283	.353	.243	0
age (yr)	29.42	1552	30.20	28.33	28.31	0
report experience	.433	5341	.448	.399	.423	.014
experience (yr)	4.4	843	3.9	5.4	4.9	0
Reference	.296	5341	.335	.26	.207	0
Refer. phone nr	.014	5341	.023	0	0	0
Employer posts	.115	5341	.089	.165	.142	0
1=workpermit	.048	5341	.048	.049	.051	.914
Sleep in	.154	5341	.164	.214	.054	0
able to drive	.071	5341	.041	.062	.178	0
Certificate	.034	5341	.025	.073	.017	0
virtuout traits	.356	5341	.412	.284	.257	0
Nr words	31.1	5341	30.4	33.5	30.9	0
Picture	.087	5341	.045	.161	.138	0

Note: The table reports mean profile characteristics of job seeker profiles compared across sectors. P-values are reported of a test of equal means across sectors.

Table A.3: Domestic Worker Phone Survey Results, Cape Town

	N	Sample Means: Pooled	SA	Docum For	Undocum For	p-values: SA=For	SA=Leg	SA=Ileg	Doc=
Age (yrs.)	206	32.102	38.2	31.1	29.6	0	0	0	.0
Education (yrs.)	201	11.06	10.46	11.45	11.01	.013	.001	.091	.0
Max hours willing to work	202	40.9	41.5	41.2	40.4	.588	.897	.531	.5
Wage, daily (ZAR)	191	181.3	179.4	181.2	182.3	.704	.816	.694	.8
Wage, hour (ZAR, imputed)	189	23.11	23.97	22.57	23.15	.479	.336	.567	.6
Reservation wage, daily	200	185.9	185.6	185.5	186.5	.975	.981	.903	.8
Contract	197	.228	.333	.213	.188	.084	.173	.091	.6
Treated well: 0=Never..3=Always	197	2.59	2.71	2.58	2.57	.257	.25	.338	.8
How not treated well?									
.... lower pay than agreed	208	.038	0	.052	.047	.004	.044	.044	.8
.....employer rude	208	.077	.043	.091	.082	.329	.292	.364	.8
.....had to work more hours	208	.048	.022	.078	.035	.286	.138	.648	.2
Can ask for time off	192	.958	.949	.96	.962	.741	.79	.76	.9
Paid overtime	198	.47	.476	.432	.5	.924	.653	.804	.4
Heard about CCMA	197	.756	.93	.789	.628	0	.023	0	.0
Take employer to CCMA	176	.705	.947	.701	.577	0	0	0	.1
Know about minwage law	197	.33	.395	.355	.269	.285	.669	.167	.2
Should earn minwage	168	.792	.846	.8	.75	.277	.55	.232	.
Ever negotiated wage	196	.464	.524	.453	.443	.376	.469	.402	.8
N			4[illegible]	77	85				

Notes:

C. Simple Model of Hiring

A simple model may help to clarify these three channels: assume that firms have the following utility function $U = U(Y, \lambda, \delta, c)$. Aside from the productivity of workers (Y), firms care about the fines (δ) they receive for hiring in undocumented immigrant with probability p which is a function of the legal status l. Last, employers may incur utility loss (λ) from interacting with foreigners. As evidence shows that South Africans' attitude towards foreigners varies across nationalities (Crush 2008), the distate is modeled to be a function of the applicants' country of origin $\lambda(d)$. In addition, the firm accrues cost c for reviewing each applicant. While I later consider a more realistic tournament model of hiring, for simplicity assume that there are n vacancies and n applicants. For each job applicant, the risk-neutral firm decides to hire iff:

$$E[Y|d, x] - \lambda(d) > c + \bar{w} + p(l)\delta \tag{6}$$

The condition simply states that the firm hires a native if the expected output is higher than the wage plus the screening costs c. For a foreign applicant, the expected profit must also exceed the expected fine and the distate from hiring a foreigner. For now, I assume that wages are fixed ($\bar{w}$) so firms strictly prefer to hire more productive workers. The model predicts that conditional on applicants' observable characteristics X, employers prefer to hire South Africans for three possible reasons:

1. Employers risk paying a fine if they are caught hiring an immigrant who is undocumented (i.e. $l = 0$).

2. Consistent with models of taste discrimination, employers may receive disutility from hiring and interacting with foreigners.

3. Consistent with models of statistical discrimination, employers may belief that South Africans are more productive with respect to unobservables e, i.e. $E[Y|x, d = SA] > E[Y|s, d = Foreign]$.

D. Employer Learning Model

To shed light on the nature of discrimination, I test a simple employer learning model based on Faber and Gibbons (Farber and Gibbons (1996), FG) and Altonji and Pieret (Altonji and Pierret (2001), AP). FG model productivity Y as a function of $\tilde{Y}(x, q, n, z)$. x captures information available to both the researcher and employer, q information available only to employers, z information only available to the researcher and n unobserved factors. Since I observe exactly the same information set as the employer at the time of the screening decision, the production function thus simplifies to $Y = \tilde{Y}(x_i, n_i)$.

Let's posit an additive separable production function in which a worker produces output $y = f(d) + g(x) + e$ with $e \sim N(0, \frac{1}{h_e})$ and let the ability y of each job seeker be a random draw from the population distribution of their nationality d. The employer observes the applicants' nationality d and forms beliefs $E[y|d \sim N(m_0(d), \frac{1}{h_0})]$. The employer then receives additional information x with $y|x \sim N(m_1, \frac{1}{h_1})$ which she uses to form the posterior belief $y|x, d \sim N(\frac{h_0 m_0 + h_1 m_1}{h_0 + h_1}, \frac{1}{h_0 + h_1})$.[31] It is straightforward to show that the posterior productivity belief is a weighted average of the information on nationality and other signals with the weights determined by the relative informativeness presented by the inverse of the variance of the error term.

This simple model of employer learning offers a test to distinguishing taste from statistical discrimination similar to AP (2001). Assume that the vector X consists of J potential predictors of productivity Y such as age, experience or education, $X = (x_1, x_2, ..., x_J)$, that the job seeker may reveal on her profile.[32] Let's define $I = \frac{\sigma_x^2}{\sigma_e^2}$ as the relative informativeness of the observed set X relative to the unobserved error term e and assume that I monotonically increases in the number of signals j actually provided, i.e. each additional information increases the predictiveness of Y.

AP test whether the effect of the easy to observe variables (e.g. education, race) decrease as the employer learns additional information on the hard-to-observe variables over time. The equivalent

[31] I assume that h_i, which captures the inverse of the population variance σ_i^2, is constant, independent of d.

[32] Given that there is no standardized form to report information, job seekers may be strategic about what they reveal. For 201 job seekers in the domestic work sector, I observe *both* what is revealed online and detailed data from an anonymous survey. I find that there is no significant correlation between the reported age, years of experience, legal status or nationality with the decision to whether a person reveals this information in the classified ad. This could be explained by two factors: either job seekers do not know about the benefit of revealing positive characteristics and/or negative characteristics (e.g. immigrant status) is easily verifiable by employers so job seekers have an incentive to reveal it before occuring interview expenses.

test in this setting is to test if nationality becomes less important in the screening decision as job seekers provide more information. In a model of *statistical* discrimination,

$$\frac{\partial^2 P(Hire = 1|X, I, d)}{\partial d \partial I} < 0 \tag{7}$$

, i.e. the effect of nationality d on the probability of an applicant beeing hired decreases as additional information I is available (conditional on the actual content of the new information $X = x$). The intuition behind this prediction is straightforward: if firms use nationality as a proxy for productivity, then the importance of nationality should decrease as other predictors of productivity become available and are factored into employer beliefs. By contrast, if employers have the same productivity expectation of South Africans and foreigners but prefer hiring locals due to *taste*-based discrimination, we would expect the revelation of additional information to have no effect on the foreign-national gap in hiring decisions, i.e. $\frac{\partial^2 P(Hire=1|X,I,d)}{\partial d \partial I} = 0$.[33]

E. Domestic Worker Phone Survey

Protocol: We collected phone numbers of all job seekers who posted in the domestic work category in Gauteng and the Western Cape between December 1st, 2015 and January 10th, 2016. Experienced phone surveyors called these people and explained that they are calling as part of a research project to "*understand the situation of domestic workers in South Africa.*" and that you "*will be asked about your job search and work history.*" Surveyors stressed that we are not offering employment and that the survey is completely volunary, fully anonymous, and would take 15 minutes to complete. "*As a thank you for participating in this research, you will receive 30 Rand Airtime, regardless of your responses and how many questions you choose answer.*" The compensation is about 20% of the daily income of domestic workers. It was paid via an airtime transfer to a phone number of their choice.

Selection: We attempted to call a total of 444 people of which we reached 343 (77.2%). Of these people we reached 303 (88.3%) agreed to participate in the survey. This is a remarkably high share compared to other phone surveys. To assess how selective this sample is, in particular with regard

[33]This test is distinct from related audit studies. Bertrand and Mullainathan (2004) and Oreopolous (2011) test how callback rates change for higher quality CVs. The quality of resumes is improved both by adding new information (e.g. additional certificates) *and* by changing the quality of signals provided (e.g. domestic vs. international work experience). The test in equation (7) by contrast looks at the effect of providing more signals (j) holding the quality of provided information (X) constant.

to nationatlity composition, I test whether the nationality information posted on the website is correlated with the probability of being reached or the decision to participate (conditional on being reached). Compared to people who do not post their nationality, foreigners are 0.3% more likely to be reached (p-value:0.96) and 2.7% more likely to participate (p-value: 0.54). South Africans are 3.1% more likely to be reached (p-value:0.29) and 3.3% more likely to participate (p-value:0.77). These results suggest that the sample is representative of the population of job seekers, at least with regard to nationality composition.