The Coloured Walls: Ethnic differences in outcome in the Academic Job Market for PhD Economists

Viet Vu *

April 29, 2016

Abstract

Using a novel data set for the Job Market for PhD Economists, this paper examines the factors that allow an applicant to be successful, measured by the number of interviews they receive. This paper contributes to the literature by examining the link between perceived ethnicity as well as English ability and job market success. I find that there are significant differences in outcome between predicted ethnic groups. Consistently individuals identified to be in the East Asian ethnic groups receive less interviews than comparison groups (Anglo-Saxon). There is evidence that some of these differences can be explained due to differences in perceived English ability between ethnic groups. However, the effect of observed English ability by candidates is only significant for individuals in the East Asian ethnic group.

*The author would like to thank the following individuals for being instrumental in writing this honours thesis: Professor Craig Riddell for his continuous mentorship and guidance as the thesis advisor; Professor Michael Peters for providing the data set and working closely with the author initially; Professor Nicole Fortin for her guidance as the Honours program coordinator; and Allen Lee for his assistance in cleaning the initial dataset; Professor Nisha Malhotra, Professor Yaniv Yedid-Levi, and Professor Kevin Song for their advice on research approaches. This paper would not have been possible without these individuals.

1 Introduction

At the beginning of every September, all graduating PhD Economics students think about their future paths. Whether it be academia, government, or private sector, these students make decisions that has significant impact on their careers. Traditionally, many students apply to positions within academia. These academic positions are usually a first year tenure track Assistant Professorship at recognized universities for those wishing to do research, or either a tenure-track Instructorship or non-tenured Lecturers. This academic labour market, after many iterations, has been called "the Job Market'.'

This is a stressful period and many anecdotes are provided through various informal channels: a candidate's thesis advisor, an online forum such as econjobrumours.com, and other students. These anecdotes are provided on the basis of experience. All tenured (or non-tenured) economists have had to participate in "the Job Market" at an early point in their careers. Thanks to the rigidity of academia, the job market process has remained relatively the same for many decades. The only recent major change in the Job Market has been the introduction of online applications (most famously Econjobmarket.org (EJM) platform, when it first introduced the online application capabilities in 2005; followed by the American Economic Associations' Job for Economists (JOE) platform) that dramatically reduced the cost of applying for applicants.

The introduction of this application platform allowed researchers to collect data for the Job Market. For the first time, researchers could assess the popular claims made anecdotally. However, whether due to the lack of interest in this topic or the reluctance to study their own field, previous research has not focused heavily on the outcome difference between demographic groups (ethnicity and gender) in the academic job market for economists. This paper aims to contribute to that field of research by utilizing a novel data set, collected by Econjobmarket.org for the 2014-2015 Job Market and assess factors that affect an applicant's success, measured by the number of interviews they receive. In particular, I examine the role of ethnicity and gender in a candidate's success.

Overall, I find that gender, when coded in a binary manner, does not play a significant role in how successful an applicant is, across a broad range of specifications. In terms of ethnicity, I find that individuals with ethnic names, especially those associated with Asian, African, or Middle Eastern heritage, consistently receive fewer interviews compared to individuals with a Western European or Anglo name. Within my base specification, when I model interview as a random arrival process, individuals with East Asian and Southeast Asian names have 0.3 lower log expected interviews. Even after decomposing the gap into an "explained" and "unexplained" components, much of the ethnicity effects are unexplained by the observed data. However, some of the differences could be explained by looking at language differences between ethnic groups.

I find, through estimating the readability of academic papers submitted by candidates, that the observed writing skills (and thus English ability) played a significant predictive role for East Asian and Southeast Asian ethnic individuals, with individuals who had less readable texts in this demographic receiving less interviews. No such result was found for other ethnic groups. These results point to differential treatment (at least amongst the US and Canadian schools) for individuals based on their perceived ethnic names.

This paper proceeds by providing an overview of existing literature on the Job Market for Economists, followed by a summary of the Job Market process. The fourth section summarizes the data used while the fifth section provides my approach as well as regression results and my interpretation. The sixth section concludes.

2 Literature Review

The literature on this topic can be divided into three main groups. The first is a direct analysis of the job market for economists. The second is focused on analyzing the gender and racial disparity in the economics professions. The third is focused on developing effective methods (especially in terms of ranking departments) to produce a ranking of economists and economics departments.

This section explores each of the three groups in detail. Broadly speaking, research on the job market focuses on the applicant side, identifying attributes that make an applicant successful. Existing papers rely heavily on surveys of applicants and thus ignore the demand side (departments decision) of the labour market and make the analysis prone to biases. In addition, applications cannot be matched to specific job openings, making detailed analysis on this topic challenging. The research on discrimination in economic departments has focused on both the pre-job-market choices by female PhD candidates and performance of female Associate Professors after they have a placement though limited research exists for studying discrimination in the job market process or ethnic discrimination in the profession. Finally, there are a variety of ranking methods that exist for economics departments, each focusing on a specific measure of success for economists.

2.1 Literature on the Job Market

Analyzing the Job Market for Academic Economists has been a long standing interest within the field. Substantial work in the area dates back to at least 1988 with (Carson & Navarro, 1998). The analysis of the job market tend to focus on one of two aspects: how candidates apply and how departments choose candidates. Carson and Navarro (1998) mainly looks at the issue from the department perspective, surveying the decision making process of the department as well as identifying qualities economics departments deem important in candidates. Approaching this problem from the department side is in the minority however, whether due to lack of data (a department may not be willing to give up application data or simply does not track them) or lack of interest.

The vast majority of studies on the job market focused on the applicant side as a result. The data is usually collected through surveys. Some of these studies focus on one particular applicant quality, such as the reputation of the dissertation advisor in Hilmer and Hilmer (2007) and Hilmer and Hilmer (2011); applicant's graduate school performance in Athey, Katz, Krueger, Levitt, and Poterba (2007); or the applicant's gender in Steinpreis, Anders, and Ritzke (1999) though Steinpreis' study was done for the Psychology Academic Job Market. Some studies also focus on looking at the role of specific mechanisms in the market, such as the role of signalling in Coles et al. (2010), or anonymous job applications in Krause and Rinne (2012). Yet other studies focus on examining overall job market performance, as in Barbezat (1992) and McFall, Murray-Close, Willis, and Chen (2014). Finally, there are those that look at initial job matching and subsequent performance in terms of publications 10 years after the initial placement (Oyer, 2006).

In most cases, these studies emphasize the importance of the applicant's PhD degree institution ranking, the reputation of one's thesis advisor, and the strength of their recommendation. Some studies also showed potential discrimination in terms of applicant origin (with one study finding that an applicant who completed an undergraduate degree in China is less likely to be successful in the US market.) Most of these studies also control for factors such as field chosen by candidates to control for individual differences.

It is also important to point out that most of the regressions run in these studies are logit or probit models due to most empirical models having the dependent variable coded as a binary (successful or not successful). The outcome variables are usually extensive thanks to surveying applicants and are broadly divided into initial interview, fly-outs, and job offers.

However, as mentioned, all of the currently available studies on the job market rely on applicant surveys. Though some papers such as McFall et al. (2014) aims to alleviate potential bias by creating weighting factors for their collected data, surveying applicants is inherently prone to biases. Surveys were conducted because there were no available and comprehensive database on the job market.

In addition, one unique disadvantage of using survey data from job applicants is a complete lack of information from the department perspectives and the interdependence nature of the job market. For example, an applicant applying to a mid-rank department has to compete with other applicants who choose to apply to mid-rank departments. This means that the competition is different for each advertised position. This is the biggest issue with existing papers analyzing the job market.

Finally, many of these studies focused much more on characteristics that applicants had some control over (such as their graduate degree, their publishing history, etc.) and not on qualities that may indicate discrimination such as race or gender. This may both be due to lack of data and the assumption that such discrimination does not exist in the job market. It is also notable that in cases where gender data is available, no apparent discrimination based on gender or only a small level of discrimination is found.

2.2 Literature on Discrimination in the Economics Profession

Much of the literature on discrimination in the Job Market has been focused on the publishing pattern between male and female economists such as in Maske, Durden, and Gaynor (2003) or promotion patterns in economics academia (McDowell, Singell Jr., & Ziliak, 1999). Yet others focus directly on gender differences in career paths of economists such as that in Kahn (1993) or McMillen and Singell Jr. (1994). Most do find substantial differences in the economics profession.

Notably, these studies find that much of the difference is not explained. For example, Maske et al. (2003) finds that the 59% of the gender productivity gap cannot be explained due to observable factors. One issue that is common with these studies is that they consider the discrimination in isolation; that is, discrimination that happens in one aspect of job performances is not related to other aspects. This may explain the high degree of unobserved gap.

Most of these studies also focus on citation analysis, thanks to the relative ease of obtaining and organizing such data. This is a valuable and powerful tool for analyzing discrimination after a candidate received an academic position. In addition, many ranking methods for economists take citation analysis into account, making this choice of methodology appropriate.

However, the same method may not be applied for analyzing applicants in the Job Market. This is because only a small number of PhD graduates have already been published and referred to in academic journals. McMillen and Singell Jr. (1994) is the only study that has discrimination analysis focused directly on the job market. However, one major weakness in McMillen and Singell Jr. (1994) is the lack of depth in empirical specifications. This paper also only looks at the gender difference in career choices and not relative success in the job market after this choice is made.

At this point, I can also demonstrate one possible missing link between studies focused on gender discrimination in economics and studies of the job market for economists. As McMillen and Singell Jr. (1994) demonstrated, women's career choices are different from those of males. In addition, as McFall et al. (2014) shows, the majority of job market candidates come from high ranking departments. It could therefore be that only women who are in high ranking departments (thus those who are extremely competitive, comparable to their male counterparts) choose to enter the market, diminishing observable discrimination at the job market level. We do not have data to substantiate this claim in this paper however.

2.3 Methods in ranking departments

Finally, we focus on existing methods used to rank economists and economic departments. The ranking method is important as different ranking methods used may produce different analytical outcomes. Though many studies alleviate the small variation in rankings by aggregating these rankings (such as grouping departments into top 10, top 20 and so forth), other nuances such as field ranking of departments are not often looked at.

The majority of ranking methods use citation analysis to rank economists. Most notably, Research Papers in Economics (RePEc) ranking uses extensive citation analysis to produce a variety of different ranking for economist (Zimmermann, 2012). However, the methodology employed to rank economic departments has a much higher variety. Most studies agree that the best departmental ranking comes from using an aggregation of some form (most commonly economists associated with that department) but one ranking method is often optimized for a specific purpose.

For example, the RePEc ranking as described in Zimmermann (2012) mostly uses aggregate of citation ranks of economists associated with the departments. Then they control for the relative size of the departments so bigger departments do not have an inherent positive bias associated with them. However, many commentators do note that this may not be the best way to rank departments. Athey et al. (2007) argues that field ranking, for many purposes, is much more important than overall citation ranking. It makes sure that when a PhD hopeful is considering different schools for study, they will not make uninformed decision focused solely on overall department ranking.

On the other hand, Amir and Knauff (2008) argues that regardless of the field ranking, the most important function of a good PhD program is placing their graduates at top departments after graduation. Therefore, they develop a ranking method that focuses on department's PhD placement.

In general, different ranking methods do change the departmental ranking slightly. However, in general, if a department is considered to be a "top" department, such as Harvard, its ranking is very likely to remain in that top range. The most drastic change happens in the field ranking produced by Athey et al. (2007) as there is a clear field focus apparent in many departments.

Ultimately, no one ranking method can convincingly make a case for the best ranking. Each ranking is optimized for some purposes, have some limitations and is useful. Therefore, a variety of ranking methods will be used to differently weigh the strength of each applicant in my analysis.

3 The Job Market

I briefly describe the Job Market process to those who are not familiar with the idea in this section. A more extensive discussion can be found in Cawley (2009). Economics departments around the world are always on the lookout for new talent, whether to add new faculty members or replace faculty who retire (or are dismissed). Though there is a market for experienced senior faculty members, most of the open market focuses on prospective PhD Graduates in Economics.

Every year, as PhD students complete their dissertation, they are "put on the market." This means that the department sends out a signal to all Economics departments that this student, in their opinion, is ready to graduate and contribute research. Most often, the student, called the Job Market Candidate or JMC, will also have a Job Market Paper (JMP) which is focused on their primary field of research, allowing prospective employers to gauge their ability.

After this initial signalling period, two things can happen. Well-resourced departments often have a placement officer who will personally call economic departments that are a good match for specific students on the behalf of the student. This acts as a signal of quality (especially because the placement officer's reputation is at risk if they recommend the wrong JMC to the wrong position). For individuals who are not placed in this manner (either a student who is not considered a "Super Star" or those from less-resourced departments), they engage in the open market, applying to jobs that are advertised through one of two major channels: Econjobmarket.org, a non-profit operated by a group of Economics professors or Job Openings for Economists (JOE) run by the American Economics Association.

In each of their applications, the candidate encloses their CV (academic record), a Cover Letter, their JMP, several recommendation letters from Economics faculties, and other statements (such as statement of teaching philosophy, evaluation of teaching performance, etc.). The departments then evaluate each candidate and decide whether to interview them or not. Most interviews are conducted in February at the annual American Economics Association's conference held in a major city in the United States. The 2016 AEA Conference was held in San Francisco. Here, candidates selected for interviews will move through several hotel rooms booked by the AEA and engage in an interview, presenting their Job Market Paper and defending it.

After those interviews are held, Economics department often continue the process by offering a fly-out to the most promising candidates. A fly-out means that the department will arrange for the potential candidate to travel to their department, present their research to faculty members, and interact with the faculty members in a formal (scheduled meetings) and informal (dinners) manner. At this stage, the candidate's research is evaluated by faculty members as a whole and a fly-out also gives the department a chance to see how well a candidate would fit into the culture at that specific department.

If a candidate makes a great impression during their fly-outs, and if their research is of top-notch quality, the department may send the candidate an offer for placement. The candidate and the department then enter a period of negotiation over the candidate's compensation packages, benefits, as well as research and teaching obligations. Once these details are sorted, the candidate accepts (or rejects) the department's offer and concludes the Job Market.

However, this only concludes the primary job market. In some instances, some departments may not be able to find a match during the first round of applications and some candidates may not receive any satisfactory offers from economic departments. These unmatched agents engage in the secondary market which is much more informal and decentralized in the nature than the process described for the primary market.

As can be seen, the Job Market is a long process that requires a high level of investment by candidates. Further, the process is highly centralized and is repeated every year, creating opportunities to understand this market.

4 Data

I use cross-sectional data collected from Econjobmarket.org for this paper. Econjobmarket.org (EJM) is, according to their website, "a nonprofit organization (501c3 tax-exempt charity) that facilitates the flow of information in the economics job market by providing a secure central repository for the files of job-market candidates (including papers, reference letters, and other materials) accessed on line" (Econjobmarket.org, 2015). EJM is amongst the most popular platforms used by job market participants.

The data spans the 2014 job market, looking at openings advertised between January 1st, 2014 and May 19th, 2015. This data represents 58,915 applications by 4,787 applicants towards 530 openings from 335 institutions.

However, I will focus on a specific subset of this overall data for my analysis. In 2014, EJM introduced an internal platform to allow recruiters to schedule interviews with applicants. In the 2014 Job Market, 65 institutions, who posted 117 openings, opted to use this feature. As a result, EJM was able to collect full interview offers for these openings. I thus limit my analysis to this subset (referred to as the interview sample from now on), covering 24,871 applications submitted by 2,715 applicants.

In this section, I describe and discuss the strengths as well as limitations of my dataset. I start by outlining the transformations performed on select variables. I then discuss the appropriateness of using the interview sample to represent the Job Market. I follow by describing the applicants' characteristics in detail. Finally, I look at the interview dynamics in the data.

4.1 Initial Transformations

I performed several transformations to the dataset. Firstly, I corrected typing errors and unified spelling variations (such as abbreviations). Secondly, as demographic information such as gender and ethnicity was optionally reported by applicants, I devised strategies to obtain complete demographic characteristics. I was not able to use the optionally reported demographic data due to potential endogeneity in the applicants' choice to reveal personal data.

To overcome this reporting issue, I use a name-gender and name-ethnicity classifier to infer an applicant's demographic characteristics from their full name. For gender, through the gender package in R (Mullen & Blevins, 2015), I utilize the US Census data to create a probability estimate for an individual's gender. For ethnicity, I use the algorithm devised by Ambekar, Ward, Jahangir, Male, and Skiena (2009) performing similar probability estimation. Using algorithms to predict an individual's identifying features is appropriate for two reasons. Firstly, both of the algorithms have a high level of accuracy (at least as good as other classifiers) and perform well with test data. Secondly, I postulate that application reviewers form racial and gender expectations using applicants' names; therefore, using an algorithm simulates this process. This process is documented in some well known papers such as Bertrand and Mullainathan (2004). Therefore, using applicant names to form demographic expectation models reality well.

In addition, I use the name of the PhD-granting university reported by applicants and the "Research Papers in Economics" (RePEc) ranking to produce a numerical measure of an applicant's PhD degree strength. I group these rankings into top 10 (universities with rank 1-10), top 20 (universities with rank 11-20), top 50 (universities with rank 21-50), top 100 (rank 51-100), top 250 (rank 101-250), and outside 250 (rank below 250). The RePEc ranking is produced for each economic department by aggregating the scores for authors affiliated with that department. I group the rankings to account for minor variations in different ranking methods and to identify the general effect of attending an elite school, regardless of field specialization.

In addition, EJM has a comprehensive database on recommenders who submitted recommendation letters for each application. I used the recommenders' name and their RePEc ranking to produce a measure of recommendation strength. Due to lack of availability of data, exact ranking was used only for economists listed in the top 5% in RePEc. For each additional percentile increment (top 5%-6% and 6%-7%), the median rank for that percentile was used for all economists falling into that percentile. Anyone ranking below 7% was assigned a rank of 3700 for identification purposes. I recognize that this is an imperfect measure: a strong positive recommendation letter from an unranked economist may benefit the applicant more than lukewarm letters from highly ranked economists. However, I assume here that all recommendations are equally strong as they are often chosen carefully by applicants, making the relative respectability of economists writing them important in application evaluation. Further, high ranking economists may only choose to advise top quality students, allowing recommender strength to serve as a proxy for candidate strength.

From these raw rankings, I calculate two metrics for each application. The first metric I calculate is the rank of the highest ranking recommender for a particular application. Notice here that this metric is produced for each application and not each applicant. I account for the fact that applicants may ask different recommenders for different applications. The second metric I derive is the average ranking of recommenders for a particular application. As each application submits multiple recommendations, both the "best" recommendation rank and the average recommendation rank are important in whether or not a candidate receives an interview.

4.2 Sample Appropriateness

One question that comes to mind when assessing this sample is whether the sample is representative of the market. I aim to demonstrate this by analyzing the similarities between characteristics of the interview sample and the full data. I first look at institution rankings of departments that scheduled interviews using EJM. I use RePEc rankings as above to rank these departments. The result is displayed in Figure 1. The two sided KS test between the two ranking densities produces a D statistics of 0.025 and a p-value of 0.9965, showing that the two distributions are not statistically different from each other.

The second characteristic I look at is the origin country of the position openings. Table 1 shows the 7 countries where the highest number of positions are located in the collected data and the interview subset. 6 out of 7 countries that appear in the table for the full data also appear in the interview subset. In particular, 4 countries: United States, Canada, Spain, and Italy appear in the same position for both data. Though there are some discrepancies (notably China missing from the interview subset), the similar geographic variation of the subset improves my confidence in this sub-sample serving as a good proxy for the entire sample.

4.3 Applicants

This section will provide an overview of applicants within the interview sample. In my data, 2715 unique applicants were identified, with the earliest application from these applicants being submitted on June 2nd, 2014. The vast majority of the applicants (86%) listed PhD as their highest degree achieved or soon to be achieved. A small number (3.8%) listed some form of Masters, while the rest did not have this information. In terms of their degree location, the majority of the applicants (60%) completed their training in the United States. Figure 2 shows the density count.

For comparative purposes, I also present the RePEc ranking of applicants' degree universities. Overall, rankings for 1888 applicants (70% of the sample) were identified. The unidentified applicants are assumed to have gone to institutions ranking below 250 as rankings for these universities are not readily available. For identified applicants, there is a concentration in higher ranking departments (specifically at the top 50 departments.) Table 2 shows the full breakdown. There are two explanations for this trend. The first explanation concerns the difference in available resources for high ranking departments. Higher ranking departments tend to be better resourced than lower ranking departments, being able to afford to train higher number of graduates, giving them a higher number of applicants.

The second explanation surrounds the competitive nature of academic job in economists.

As explained earlier, going to stronger departments act as a credible signal for applicants seeking an academic job. This means that an Economics PhD graduate who goes to top schools is more likely to enter the academic market due to higher chance of success. Applicants who go to lower ranking schools, lacking this signal, decide not to enter the market and seek other opportunities. This means that only top students in lower ranking departments (those who can compete with applicants from higher ranking departments) apply, making them underrepresented.

Next, as explained, I used an algorithm to predict an applicants' ethnicity, the method of which is documented in Ambekar et al. (2009). The algorithm performs extremely well in the training data for identifying Greater European names (with a precision of 0.96 and recall of 0.93) and Asian names (with a precision of 0.82 and recall of 0.83). However, the algorithm does not perform as well for identifying Greater African names (with a precision of 0.56 and recall of 0.76.) Therefore, when interpreting regression results, it is important to be wary of any significance (or insignificance) with the Greater African names. At the same time, regression results can be trusted for Asian names with the Greater European names serving as the base case. The summary is presented in table 3

Finally, in terms of applicants' gender makeup, using the name-gender algorithm, 759 individuals were identified to be females, 1849 were identified to be males and 107 had ambiguous names and did not self-report their gender. The proportion of female to male applicants is proportional to the market and is consistent with the gender makeup of individuals who graduate with a PhD in Economics. For example, in the 2013 National Science Foundation's survey of Doctoral Recipients, 24% of doctoral recipients in Economics were female (National Science Foundation, 2013), as opposed to 28% in the interview sample.

4.4 Interview Dynamics

In total, 2046 interviews were offered for the 117 positions advertised. This represents 971 individuals interviewed. The distribution of the number of interviews received by each candidate is presented in table 4. Overall, the majority of applicants who applied to these positions did not receive an interview while there is a sizable number of applicants who received multiple interviews. Notably, 251(9%) applicants who received 3 or more interviews were responsible for 1175 (57%) of the interviews given. It is important to note that though there are many applicants without any interviews, these applicants may have received an interview from positions outside the considered samples.

In addition, it is apparent that within the sample, specializations in some fields are more advantageous than others. Table 5 summarizes the interview rate (offered any number of interviews versus none at all) for different field specializations reported by the applicant. Some fields such as Development & Growth, Industrial Organization, Labour, Macroeconomics and Econometrics have high percentage of applicants receiving at least one interview. Others such as Law & Economics and Political Economy have relatively lower yield, possibly due to a weak market for these individuals within academic economic departments.

I further explore the difference in characteristics between individuals who received interviews with a minority status (not being Western European or Anglo-Saxon) against non-minorities. I find that minorities who receive interviews are not statistically different from non-minorities that receive interviews, though there is a lower concentration of attending a highly ranked program for minorities than non-minorities. The difference in the population that did not receive any interviews are starker however. Non-minorities who did not receive interviews have attended, on average, higher ranked programs than minorities who did not receive an interview. These densities can be seen in Figure 3

5 Estimation

5.1 Theoretical Model

I assume that each department will produce a score for each of the application it receives. Once the score passes a certain threshold, the candidate receives an interview. That is, the model will be in the following binary form for individual i applying to department j:

$$Y_{ij} = 1\{\beta_{0j} + \beta_{1j} X_{ij} \geq u_{ij}\} \tag{1}$$

Therefore, for each applicant, the model that expresses the total number of interviews they receive after applying to J schools is:

$$\sum_{j \in J} Y_{ij} = \sum_{j \in J} 1\{\beta_{0j} + \beta_{1j} X_{ij} \geq u_{ij}\} \tag{2}$$

This result is especially important when I approach this from how departments make interview decisions. Now, this ignores the constrained optimization problem faced by departments. As interviewing candidates may be costly, given their application pool, a department chooses the portfolio of applicants $\{R_{i_1}...R_{i_{max}}\}$ they can interview by maximizing:

$$p(R_{i_1})u(R_{i_1}) + \sum_{j=i_2}^{i_{max}} (\prod_{k=i_1}^{j} (1 - p(R_k)))p(R_j)u(R_j) \tag{3}$$

Where $p(R_{i_j})$ measures the probability a candidate at the j^{th} place in the interview portfolio will be a successful match for the department and $u(R_{i_j})$ the utility associated with such applicants. Intuitively, departments will only consider hiring the j^{th} candidate when any candidates ranking above the j^{th} candidate did not result in a successful match. However, estimating the hiring dynamic in this manner mean we have to have further assumptions on the probability function as well as utility function, making the bivariate model desirable in producing simpler interpretations.

However, estimating this bivariate model is difficult given the number of position openings. Alternatively, I can interpret the dynamics in the model as candidates, depending on their individual characteristics, receive interviews as a Poisson process with a unique arrival rate of λ_i. Though the interpretation using a Poisson model will be less accurate, the deviation using a Poisson model will likely to be smaller than estimating the sum of bivariate models. This is the approach I will take. This will take the following functional form:

$$\log(E[Y|X]) = \alpha + \beta x \tag{4}$$

This equation estimates the log expected count of interviews as a function of independent variables. Intuitively, one can think of the coefficients attached to each variable as contributors to the arrival rate of interviews of an individual with similar characteristics.

5.2 Empirical Model

To examine the predictors of interviews, I perform a variation of a regression model. I first use the number of interviews each candidate received as the dependent variable. For the ethnicity measure, Anglo Saxon is used as the base comparison case. In field specializations, Microeconomics is used as the base comparison case. For ranking groups, I use "Outside 250" as the base comparison case. I also include country fixed effect to control for countries in which applicants are trained in.

In addition, I perform a logistics regression on the predictor of successful applications (if an application receives an interview or not). The comparison cases are identical to the regression of the first model. I also add the average ranking of recommenders as an additional predictor in this model.

For the degree ranking groups, I expect the coefficient attached to these to be positive and greater for higher rankings. It is common knowledge that going to a higher ranking school improves one's chance of receiving an interview.

One potential problem with my approach surrounds omitted variables bias. As observing all characteristics considered by a hiring committee (such as strength of an applicant's job market paper) is difficult, some bias is likely present due to variables that were omitted from the regression. Further, the error present in the estimation of ethnicity provide additional layer of uncertainty in the estimation results. Reverse causality may not be a problem here as the interview decisions are made after application factors are considered.

I now discuss my expected results. For ethnic factors, I expect to see a modest negative coefficient attached to ethnic minorities when compared with the base group of Greater European and Anglo Saxon (for the second level ethnicity). This is consistent with the theory of labour force discrimination. The effect is expected to decrease once further explanatory variables are accounted for.

Overall, without any other explanatory variables, identified ethnicity was highly significant. Table 6 summarizes the result. Compared to the base group of Anglo Saxon, the Asian Ethic group, on average, has 0.46 lower log expected count of interviews. Similar results hold for the Greater African Ethnic group which has 0.55 lower log expected count of interviews. Controlling for country effect and field effects from the base specification did change the coefficients significantly for the Greater African Ethnic, reducing the effect to 0.45 lower log expected count.

When further specification is added to control for degree ranking groups, an expected pattern emerges. Compared to group of lower than 250 in degree ranking, going to the top 10 schools shifts log expected count by 1.05. Similar effects hold for applicants going to universities in the rank 11-20. While the effect dramatically reduces for schools ranked between 51-100 (0.36 shift) and outside of the top 100, the difference is statistically insignificant. Gender effect was not significant at the 5% level regardless of field and country controls.

Notably, when both the ranking and the country effect as well as the field fixed effects are taken into account, the coefficient on ethnicity becomes less negative. In the 4th model considered, Asian ethnic group saw a 0.33 negative shift in log expected count (compared to the Greater European group) and African Ethnic is associated with a 0.32 lower log expected count. However, these effects are still at around a third of the positive effect of going to the top 10 departments (compared to unranked departments).

I follow up this analysis by using the second level ethnicity data for applicants. Table 7 presents the regression result for this class of models. When only ethnicity is considered, applicants with a Middle Eastern name sees a 0.46 negative shift in log expected count than applicants with Anglo Saxon names. Groups such as East Asian names and Indian names sees a 0.24 negative shift in log expected count. Surprisingly, individuals with Western European ethic names see a positive shift in log expected count of 0.37 compared to the base group. This supports my hypothesis that ethnicity is an important factor in the application dynamics. The coefficients on ranking is virtually identical to the first class of models.

The coefficient on the West European names can potentially be explained by the inclusion of many European departments in the interview sample (namely Spain). It is common for departments operating in a specific geographical region to prefer candidates who are fluent in the language spoken in that region. Therefore, I decompose my sample further by only considering applications toward US and Canadian departments to examine whether this hypothesis is true. These two countries are chosen due to the large sample available and the near-identical job market environment in these countries.

For this purpose, instead of regressing over the number of interviews an applicant receive, I regress over whether an application made by an applicant is successful in receiving an interview or not. I chose this model instead of the previous model as restricting the position I regress over not only diminishes the pool of applicants considered but also many characteristics within the data (overall competitiveness of the market). By using application data instead, I make sure that I compare like to like.

Instead of thinking of interviews as arriving at a specific arrival rate and candidate characteristics affecting that arrival rate, I assume that certain characteristics are likely to increase (decrease) the chance of an application being successful (unsuccessful) in receiving interviews.

I further assume (for simplicity) at this point that all candidates are evaluated in the same manner by different departments and the only difference between departments is due to randomness with mean 0. This allows us to measure the importance of different factors in interview success for the overall market.

Table 8 presents the baseline regression model. For this model, the average ranking of recommenders attached to each application is also added as a regressor. The dependent variable is a binary, indicating whether an application receives an interview or not. Therefore, a logistics regression is used. Only regressions at the second level ethnicity are shown. As not enough data is available (in number of applicants studying in a specific region), country fixed effects are also omitted. The effect is available as a baseline comparison to the restricted sample.

Table 9 shows the result of the regression for the application made towards US and Canadian positions. First, East Asian (-0.5 standard deviation shift) and Middle Eastern (-0.6 standard deviation shift) populations still stand out as the groups that have lower probability of a successful application than Anglo-Saxon named individuals. The effect is still as strong when factors such as ranking is considered for the East Asian and the Middle Eastern population.

It is also interesting to note that the statistically significant coefficient attached to the African Ethnic population disappears after controlling for factors such as graduate school ranking as well as average recommendation rank. This hints to the fact that the perceived disadvantage for the African ethnic population may be explained by observable differences between groups in the model.

Finally, at least for the US-Canada sample, when not controlling for field effects, there seems to be a slightly positive effect for females (a shift of 0.15 standard deviation). This effect loses statistical significance when field effects are taken into account, implying that the positive correlation between females receiving interviews can be explained by their field specialization.

5.3 Omitted Variable Bias

In this section, I expand further on the idea touched on in the previous paragraphs. If the negative coefficients that appear on the ethnic groups are due to language abilities, the ethnicity dummies could be correlated with the uncertainty in the model, creating an omitted variable bias. I seek to resolve this issue by introducing a variable that measures the language ability of applicants.

5.3.1 EF English Proficiency Index (EPI)

Starting in 2011, Education First (EF), a company specializing in English education as a second language (ESL) started publishing a report, assigning an index score of English proficiency for 70 countries around the world. EF complies this report through a free online English quiz that measures the quiz taker's ability to read and listen in English. They then aggregate the result at the country level and publish an index.

A point of caution needs to be noted here. The published methodology of the EPI suffers from various problems. First, as the English quiz requires internet connection, its results in less developed nations may be positively biased (as more wealthy citizens who are also more educated will be the individuals who have access to Internet). Further, the ability to answer a reading and listening test may not correlate perfectly to actual English language ability (especially those at the level of PhD Economics). However, it is still one of the most publicly accessible proxies on English ability of many countries around the world. Further, any PhD Economics students in the US or Canada has had to pass such language test and place in the top distribution, providing us valuable insight as to the link between perceived English ability and actual English ability.

For my study, I use the index in the following manner. For each of the following country groups: West European, East European, East Asian, Middle East & North African, and Indian Subcontinent, corresponding EPIs were identified. In particular, I excluded South America (though the dominant language spoken is Spanish) from the calculation of EPI as the West European group included hispanic names and included Scandinavia in the calculation of the EPI score for West European.

After countries are sorted into regions, I take a simple arithmetic mean of the scores, producing a score for each group in the level 2 ethnicity. I then assign that score to all applicants that were identified as belonging to that group. I then include this new variable in my regression. This new variable is obviously perfectly collinear with the ethnicity data. Therefore, when I run a regression using EPI as a regressor, I can interpret the coefficient attached to EPI as the estimated marginal effect on one's probability of receiving interviews by improving one's perceived language ability.

However, some caution is warranted here. Assume for example that the true model for determining the success of an application is in the following form:

$$Y_i = \beta_0 + \beta_1 \alpha_{1i} + \beta_2 \alpha_{2i} + u_i \tag{5}$$

Where α_{1i} and α_{2i} are random variables such that:

$$\begin{cases} \alpha_{1i} = \gamma_0 + \gamma_1 D_{1i} + \gamma_2 D_{2i} + v_i \\ \alpha_{2i} = \pi_0 + \pi_1 D_{1i} + \pi_2 D_{2i} + p_i \end{cases} \tag{6}$$

where D_1 and D_2 are ethnic dummies. In effect, I assume there are two variables that affects the probability of success where both of these coefficients are correlated with the ethnic dummies. I substitute the fitted values for α_1 and α_2 to get:

$$Y_1 = \beta_0 + \beta_1(\hat{\gamma_0} + \hat{\gamma_1} D_{1i} + \hat{\gamma_2} D_{2i}) + \beta_2(\hat{\pi_1} D_{1i} + \hat{\pi_2} D_{2i}) + u_i \tag{7}$$

$$\Rightarrow Y_1 = \beta_0 + \beta_1 \hat{\gamma_0} + \beta_2 \hat{\pi_1} + (\beta_1 \hat{\gamma_1} + \beta_2 \hat{\pi_1}) D_{1i} + (\beta_2 \hat{\gamma_2} + \beta_2 \hat{\pi_2}) D_{2i} + u_i \tag{8}$$

Therefore, if I omit the α_2 variable, the estimated coefficients on α_1 becomes biased. Further, because I assume with my model that only α_1 is relevant, the estimate of the coefficients on D_{1i} and D_{2i} are also likely biased. Therefore this variable should only be used to see if English ability (albeit an imperfect measure of) is a significant factor in explaining some of the difference observed in different ethnic groups without placing strong weight on the size of the coefficients.

Table 10 shows the result of the regression using English Ability as an explanatory variable. The result indicates that English Ability can be used to predict the probability that an application is successful, supporting my theory that perceived (or actual) English ability does correlate with the probability of successful applications. As discussed, this estimate, due to omission of other variables, may be biased. Other factors that could explain difference in outcome include the quality of one's job market paper and the true strength of recommendation letters, amongst others.

It could be that applicants who are classified as "Asian" or "Middle Eastern" obtained their degree from less prestigious universities (typically the top schools are concentrated in US and Europe). A word of caution is needed here. The majority of my sample are individuals who completed their PhD degrees in the United States (or Canada). Very frequently, the admission process for these programs require prospective students to submit proof of English ability (most commonly the submission of TOEFL, or other standardized test scores). These scores need to pass a certain threshold (often 80^{th} percentile or higher). Therefore, individuals who complete this program should be towards the upper-end distribution of all test takers in the world, and consequently their countries. It is therefore peculiar that average national English proficiency predicts application outcomes with such significance.

5.3.2 Readability Score of Job Market Papers

In the previous section, it was noted that the perceived English ability may play a role in determining whether a candidate receives interviews or not. I aim to contrast this perceived English score with actual English ability. One important factor in determining one's English ability that a department observes is their writing, particularly through the writing of their Job Market Paper.

I use this logic to conduct the following test. I use 5267 PDFs submitted by candidates in support of their application (many applicants submitted more than 1 PDFs), then extracted the introduction of those papers. Specifically I extract texts that appear between two keywords: "Introduction" and "Section 2". Although the title of the first section is relatively standardized, the title of the second section is not. However, almost all of the papers I examine has a section that states "The paper proceeds as follows, in section 2..." at the end of their introduction. I use the introduction as it is the most common element of any job market papers and is often indicative of a candidate's ability to communicate the main ideas presented in their paper, producing desirable sample texts to evaluate the documents with. The introduction also doesn't include many equations or tables which interfere with textual analysis.

I then use the Automated Readability Index (ARI) to measure the readability of the paper introductions. The ARI (Senter & Smith, November 1967) is a readability score designed to measure the understandability of a given text. Intuitively, it calculates a weighted average of word complexity and sentence complexity for a given text. Specifically, it uses the following formula:

$$4.71\frac{\text{Characters}}{\text{Words}} + 0.5\frac{\text{Words}}{\text{Sentences}} - 21.43 \tag{9}$$

For example, the ARI of the introduction to this paper is 13.77, corresponding to a level of 14, the reading level of a university student. Within my sample, the average readability score was 13.4 (corresponding to 14 or the reading level of a university student) with a standard deviation of 2.37. This is a reasonable and expected score for an introduction of a PhD dissertation.

The generated ARI was not significant in predicting either the perceived English score nor the ethnic levels in the sample, allowing the generated score to give further insight.

When this reading score is included in the initial set of Poisson regressions (over the number of interviews received by applicants), the English ability variable is not significant. However, as explained in the previous section, English ability is only a consideration for positions in English speaking countries. When I include the ARI score in the logit regressions of the restricted sample (to applications toward US and Canada), the coefficient attached to the ARI score is significant. Further, when we include interactions between the ARI score and the ethnic dummies, the only variable of the set that is statistically significant is the interaction between individuals identified to belong to the Greater East Asian ethnicity and the ARI score. The dummy variables by themselves lose significance and the ARI score by itself also loses significance. The results of various models ran is contained in table 13. The result indicates that Greater East Asians with 1 higher readability score (having less readable text) lowers the log expected count of interviews by 0.12 on average.

This has several implications. First, I note that lacking any interaction terms, having a less readable job market paper is correlated with a higher chance of the application being successful. However, this positive effect is dwarfed by the negative ethnic effects. The significant and negative coefficient attached to the interaction term between ARI score and East Asian suggest one possible interpretation that writing quality (and understandability) is only considered for candidates with East Asian ethnic names.

5.4 Oaxaca Decomposition

Oaxaca (1973) provides a method to decompose the difference between two observable groups into the explained part by the explanatory variables in the model and the unexplained part. To see how this method works, assume that there are two groups (Male and Female, Caucasian and non-Caucasian, etc.) and a following model:

$$Y = \beta_0 + \beta_1 X + u_i \tag{10}$$

Where Y is the dependent variable and X is the explanatory variable that is different between the two groups. I start by estimating the mean Y for two groups (A and B) independently.

I assume here that the uncertainty term is centred around 0 for both groups:

$$\bar{Y_A} = \beta_0^A + \beta_1^A \bar{X}^A \tag{11}$$

$$\bar{Y_B} = \beta_0^B + \beta_1^B \bar{X}^B \tag{12}$$

Therefore, the differences in Y between the two groups will be:

$$\Delta Y = \bar{Y_A} - \bar{Y_B} = \beta_0^A + \beta_1^A \bar{X}^A - \beta_0^B - \beta_1^B \bar{X}^B \tag{13}$$

There are two popular ways to continue. To understand this, let us go back to my specific case of interview numbers given to two groups: Caucasians and non-Caucasians. The first way to think about this decomposition method is to base my analysis on Caucasian groups. It could be that Non-Caucasians will receive the same average number of interviews they receive without discrimination but the discrimination takes the form of Caucasian receiving higher number of interviews (positive discrimination for Caucasians). The second way to think about this decomposition is to base my analysis on Non-Caucasian groups. It could be that Caucasians will receive the same number of interviews (on average) without discrimination and discrimination takes the form of non-Caucasians receiving lower number of interviews (negative discrimination for non-Caucasians). In the first case, I add and subtract $\beta_1^A \bar{X}^B$ from (12)

$$\beta_0^A + \beta_1^A \bar{X}^A - \beta_0^B - \beta_1^B \bar{X}^B + \beta_1^A \bar{X}^B - \beta_1^A \bar{X}^B$$

$$= (\beta_0^A - \beta_0^B) + (\beta_1^A - \beta_1^B)\bar{X}^B + (\bar{X}^A - \bar{X}^B)\beta_1^A \tag{14}$$

Alternatively, if the second group (Non Caucasians) are chosen as the base, I add and subtract $\beta_1^B \bar{X}^A$ from (12)

$$\beta_0^A + \beta_1^A \bar{X}^A - \beta_0^B - \beta_1^B \bar{X}^B + \beta_1^B \bar{X}^A - \beta_1^B \bar{X}^A$$

$$= (\beta_0^A - \beta_0^B) + (\beta_1^A - \beta_1^B)\bar{X}^A + (\bar{X}^A - \bar{X}^B)\beta_1^B \tag{15}$$

In both cases, $(\beta_1^A - \beta_1^B)$ term represents the difference between coefficient, or the unexplained (unobserved) portion of the difference, and $(\bar{X}^A - \bar{X}^B)$ represents the difference between the mean of explanatory variables, or the explained (observed) portion of the difference. Other weighting could also be placed on the two groups that implies a mix of positive and negative discrimination.

For my data, I decompose both the gender result and the ethnicity result. For gender, the two groups will be male and female. For ethnicity, the two groups will be Caucasians (Greater Europeans in my classification) and non-Caucasians (Asians and Greater Africans in my classification). I present the decomposition result in table 11 and ethnicity decomposition result in table 6. The data used here is the applicant data with the dependent variable being the number of interview candidate receives.

The decomposition result indicates that for ethnicity, the majority of the difference between the groups are unexplained by the factors in the data. This supports my initial hypothesis that there may be intrinsic differences between the perceived ethnicity of applicants. As for the gender decomposition, the difference in the explained and the unexplained portion is minor and given the insignificance of the coefficient, unimportant.

5.5 Discussion of the estimates: Are the unexplained portion due to discrimination?

We first note that throughout the various specifications and models provided in this paper, identified gender plays a small and insignificant role in determining candidate success at the initial job market interview level. This indicate that at least at the level when candidates are given interviews, there are no bias for or against female economists. However, this does not rule of potential bias prior to submitting an application (gender difference in the career choice) or after receiving a placement (in citation research).

The Oaxaca decomposition showed that a large portion of the observed difference between ethnic groups cannot be explained by observable factors. This deserves further discussion to identify the potential source of this difference. One variable that remains relatively unexplored in this paper is the quality of recommendation letters. Recommendation letters by faculties are one of the top characteristics departments look at when making decisions and it is likely that the difference in outcome observed can be explained if better measure of recommendation letter quality can be found.

However, where does the difference in letter quality stem form? Some reasons why could be comments regarding a candidate's English ability (as explored previously) or their communication skills (as a direct way in which high English ability is desirable), amongst other things. In addition, it is unlikely (though still possible) that recommenders will comment on these aspects directly. It is more likely that recommenders use different words as a signal of quality (for example, exceptional and good, though both positive, have different levels of positivity attached to it), as an agglomeration of different factors that a candidate possess.

This does not rule out the possibility of discrimination however. These expectations about each candidate may be formed based on ethnic characteristics (or perceived characteristics based on ethnic groups) and may not reflect the true ability of candidates.

Further, my textual analysis in proxying a candidate's writing ability indicate that East Asian ethnic individuals are the only ethnicity to be penalized by having a less readable job market paper. Possible interpretations include the fact that individuals of East Asian ethnicity are compared against one another or apply to the same positions. Regardless, the readability of the text is a significant factor only for individuals identified to be of Greater East Asian ethnicity.

Finally, the result of the textual analysis is contrasted with the generated ethnic groups used in this paper. We did not base our analysis on actual nationality or ethnicities. We based our analysis on identified and estimated ethnic groups using an applicant's first and last name. The fact that we identified outcome differences using generated ethnicity after controlling for a wide range of variables should be strong evidence in favour of the existence of ethnic-based bias in the job market for academic economists.

6 Concluding Remarks

Economists have studied discrimination for many years. In fact, much of the research on discrimination is routinely used by politicians and policy makers to justify affirmative action policies and long term investment policies. It was therefore surprisingly to see a marked difference in application outcomes for applicants of varying ethnicity and no difference in outcome based on gender. More remarkable is the fact that the ethnicities attached to individuals were generated purely on the basis of their names and may not correspond to their passport country nor their socio-economic status.

Now, the difference in outcome for different ethnic group may not constitute discrimination. It could be that there are other unobserved differences between these groups that can explain the difference in outcome. However, the high level of correlation found between perceived (not actual) English ability of applications and its probability of success is definitely worrying. Within my sample, the majority of the applicants come from US universities, meaning that they have to satisfy stringent English language requirements, which in effect would put them towards the top end of the distribution of the measure used to create perceived English ability, making the perceived difference in English ability an irrelevant factor. The result of my textual analysis also indicate that readability of the job market paper is an important predictive tool for application success for East Asian ethnic individuals.

Another effect could be differences in immigration laws. It could be, for example, more difficult to obtain a work visa as a Chinese national than as a Canadian national to work in the United States. Studies surrounding immigration laws and the resulting hiring dynamics should also be examined. Regardless, one fact is concretely shown here. Whether due to discrimination or not, candidates with non Anglo-Europoean name consistently perform worse than candidates with Anglo-European names.

This paper contributed to the body of research by using a new extensive dataset and demonstrated differences in outcome between machine-predicted ethnic groupings. Future research could work on identifying more extensive measures to identify the source of this difference, primarily through analyzing applicant characteristics (publication records, nationality, etc.) not discussed in this paper.

References

Ambekar, A., Ward, C., Jahangir, M., Male, S., & Skiena, S. (2009). Name-ethnicity Classification From Open Sources. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 49-58.

Amir, R., & Knauff, M. (2008). Ranking Economics Departments Worldwide on the Basis of PhD Placement. *The Review of Economics and Statistics*, *90*(1), 185-190.

Athey, S., Katz, L. F., Krueger, A. B., Levitt, S., & Poterba, J. (2007). What Does Performance in Graduate School Predict? Graduate Economics Education and Student Outcomes. *AEA Papers and Proceedings*, *97*(2), 512-518.

Barbezat, D. A. (1992). The Market for New Ph.D. Economists. *Journal of Economic Education*, *23*(3).

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review*, *94*(4), 991-1013.

Carson, R., & Navarro, P. (1998). Sellers and Buyers Guide to the Job Market for Beginning Academic Economists. *Journal of Economic Perspectives*, *2*(2), 137-148.

Cawley, J. (2009). A Guide (And Advice) For Economists On the U.S. Junior Academic Job Market. *Job Opening for Economists*.

Coles, P., Cawley, J., Levine, P. B., Niederly, M., Roth, A. E., & Siegfried, J. J. (2010). The Job Market for New Economists: A Market Design Perspective. *Journal of Economic Perspectives*, *24*(4), 187-206.

Hilmer, M. J., & Hilmer, C. E. (2007). Dissertation Advisors and Initial Job Placements for Economics PhD Recipients. *Applied Economics Letters*, *14*(5), 311-314.

Hilmer, M. J., & Hilmer, C. E. (2011). Is It Where You Go Or Who You Know? On the Relationship Between Students, PhD Program Quality, Dissertation Advisor Prominent, and Early Career Publishing Success. *Economics of Education Review*, *30*, 991-996.

Kahn, S. (1993). Gender Differences in Academic Career Paths of Economsits. *American Economic Review*, *83*(2), 52-56.

Krause, A., & Rinne, K. F., Ulf Zimmerman. (2012). Anonymous Job Applications of Fresh Ph.D. Economists. *Economics letters*, *117*, 441-444.

Maske, K. L., Durden, G. C., & Gaynor, P. E. (2003). Determinants of Scholarly Productivity Among Male and Female Economsits. *Economic Inquiry*, *41*(4).

McDowell, J. M., Singell Jr., L. D., & Ziliak, J. P. (1999). Cracks in the Glass Ceiling: Gender and Promotion in the Economics Profession. *American Economic Review*, *89*(2), 392-396.

McFall, B. H., Murray-Close, M., Willis, R. J., & Chen, U. (2014, November). Is It All Worth It? The Experiences of New PhDs On the Job Market, 2007-2010. *NBER Working Paper Series*.

McMillen, D. P., & Singell Jr., L. D. (1994). Gender Differences in First Jobs for Economists. *Southern Economic Journal*, *60*(3), 701-714.

Mullen, L., & Blevins, C. (2015). Jane, John ... Leslie? A Historical Method for Algorithmic Gender Prediction. *Digital Humanities Quarterly*.

National Science Foundation. (2013). *Survey of Doctorate Recipients.*

`http://ncsesdata.nsf.gov/doctoratework/2013/`.

Oaxaca, R. (1973). Male-Female Wage Differentials in US Labor Markets. *International Economic Review*, *14*(3), 693-709.

Oyer, P. (2006). Initial Labor Market Conditions and Long-Term Outcomes for Economists. *Journal of Economic Perspectives*, *20*(3), 143-160.

Senter, R., & Smith, E. (November 1967). Automated Readability Index. *University of Cincinnati*.

Steinpreis, R. E., Anders, K. A., & Ritzke, D. (1999). The Impact of Gender on the Review of the Curricula Vitae of Job Applicants and Tenure Candidates: A National Empirical Study. *Sex Roles*, *41*.

Zimmermann, C. (2012). Academic Ranking with RePEc. *Federal Reserve Bank of St. Louis Working Paper Series*.

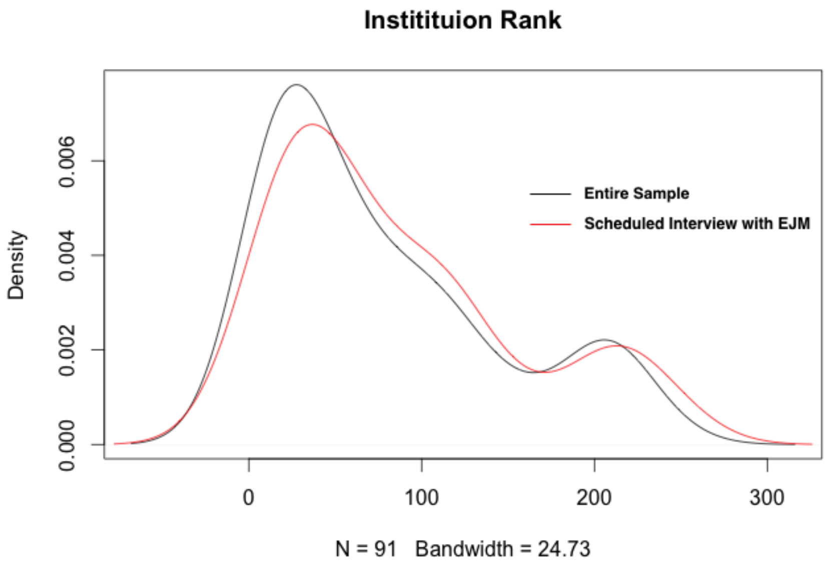

Figure 1: Density estimate of institution rankings

Country	**No. of Position**	**Country**	**No. of Position**
United States	203	United States	35
Canada	57	Canada	24
United Kingdom	54	Australia	10
Spain	34	Spain	8
China	16	United Kingdom	7
Australia	13	Turkey	5
Italy	12	Italy	5

Table 1: **Countries with most position openings. (Left: Full data; Right: Interview subset)**

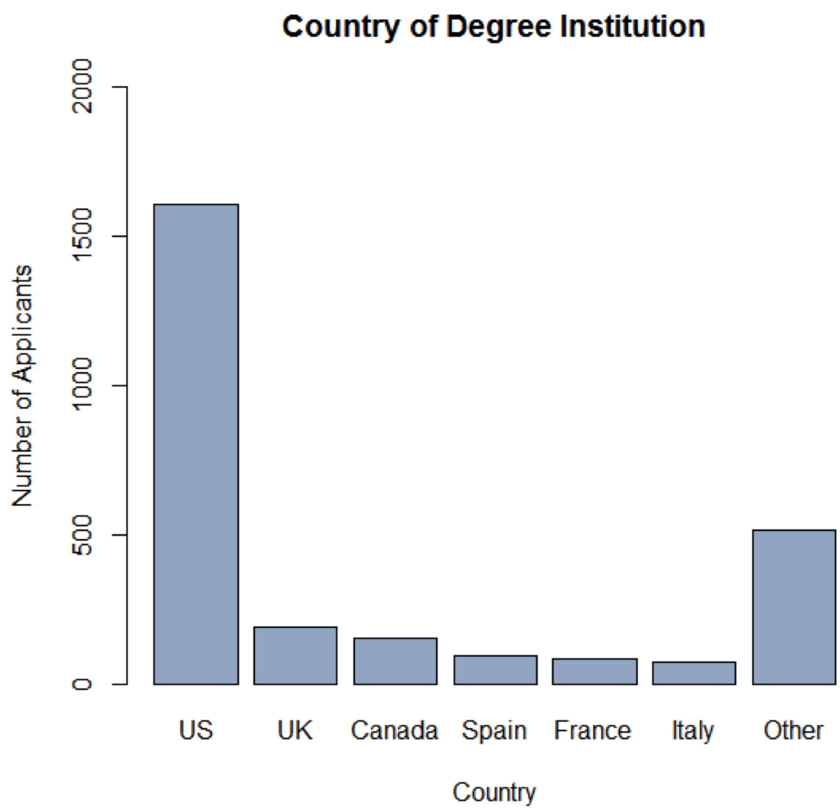

Figure 2: Histogram of country of departments

Group	**No. of Applicants**	**Implied Applicants Per School**
Top 10	274	27.4
Top 20	242	24.2
Top 50	414	13.8
Top 100	303	6.1
Top 250	341	2.3
Outside Top 250	1142	NA

Table 2: **Degree Ranking of Applicants**

Ethnicity level 1	Ethnicity level 2	No. of Applicants	Percentage
Asian	Total	915	33.7 %
	Greater East Asian	706	26 %
	Indian Sub Continent	209	7.7 %
Greater European	Total	1462	54 %
	Anglo-Saxon	443	16.3 %
	East European	160	5.9 %
	West European	638	23.5 %
	Judaish	221	8.1 %
Greater African	Total	291	10.7 %
	Continental Africa	108	4 %
	Middle East & North Africa	183	6.7 %
Unidentified	Total	47	1.7 %

Table 3: **Name-Ethnic Breakdown of Applicants**

No. of Interviews	**No. of Applicants**
0	1803
1	453
2	209
3	113
4	69
5	40
> 5	29

Table 4: **Number of Interviews per candidate**

Field	**No Interview**	**Any Interview**	**Total Number of Applicants**
Behavioural Economics	78%	22%	50
Development & Growth	59.6%	40.4%	223
Econometrics	61.8%	38.2%	178
Economic History	68.4%	31.6%	19
Environmental; Ag. Econ	80.2%	19.8%	162
Experimental Econ	80%	20%	35
Finance	74.2 %	25.8%	391
Health, Education; Welfare	62.3%	38.7%	150
Industrial Organization	59.4%	41.6%	154
Intl. Finance/Macro	67.9%	32.1%	81
Intl. Trade	59.3 %	40.7%	118
Labmy & Demographic Econ	58.7%	41.3%	213
Law and Economics	92.9%	7.1%	14
Macro/Monetary	57.9%	42.1%	349
Microeconomics	64.9%	35.1%	242
Political Economy	82.1%	17.9%	39
Public Economics	68.5%	31.5%	92
Theory	51.1%	48.9%	45
Urban;Rural;Regional	68.7%	31.3%	32

Table 5: **Rate of applicants interviewed by primary field**

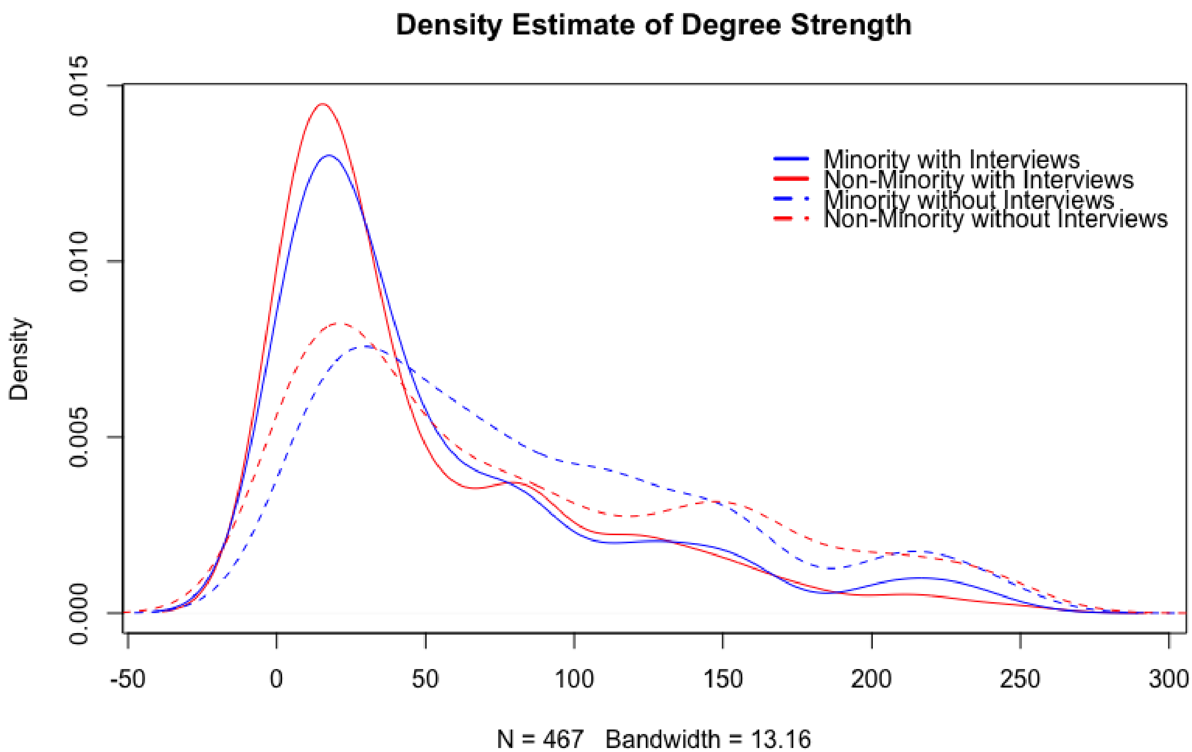

Figure 3: Density of degree rankings by interview results and ethnic groupings

Table 6: Regression Model for number of interviews received [1]

Variable	**(1)**	**(2)**	**(3)**	**(4)**
Intercept	-0.17*** (0.03)	-0.18* (0.08)	-0.64*** (0.05)	-0.8*** (0.10)
Asian Ethnic	-0.46*** (0.05)	-0.44*** (0.06)	-0.40*** (0.06)	-0.33*** (0.06)
Greater African Ethnic	-0.55*** (0.09)	-0.45*** (0.09)	-0.40*** (0.09)	-0.32*** (0.09)
Rank 1-10			1.05*** (0.07)	1.14*** (0.08)
Rank 11-20			1.03*** (0.07)	1.10*** (0.08)
Rank 21-50			0.78*** (0.07)	0.84*** (0.07)
Rank 51-100			0.36*** (0.09)	0.38*** (0.09)
Rank 101-250			-0.04 (0.1)	0.01 (0.1)
Gender			-0.07 (0.05)	0.08 (0.06)
Country effect	No	Yes	No	Yes
Field effect	No	Yes	No	Yes
Residual Deviance	0.016	4091.2	4145.7	3573.2
Degrees of Freedom	2666	2559	2554	2448

[1] Results from Poisson Regression and MLE. Dependent variable in all column is the number of interviews each applicants received. Interview samples are used. The coefficients are marginal effects of each variable. *p<0.05; **p<0.01; ***p<0.001

Table 7: **Regression Model for number of interviews received** [2]

Variable	**(1)**	**(2)**	**(3)**	**(4)**
Intercept	-0.39*** (0.06)	-0.35*** (0.1)	-0.92*** (0.07)	-0.94*** (0.1)
East Asian	-0.24** (0.08)	-0.26*** (0.08)	-0.16* (0.08)	-0.21*** (0.08)
Indian Subcontinent	-0.24* (0.11)	-0.20 (0.11)	-0.09 (0.11)	-0.07 (0.1)
African Ethnic	-0.14 (0.14)	-0.13 (0.14)	-0.06 (0.14)	-0.001 (0.14)
Middle Eastern Ethnic	-0.46*** (0.13)	-0.35*** (0.13)	-0.28* (0.13)	-0.26* (0.13)
Judaism Ethnic		0.26** (0.09)	0.20* (0.09)	-0.15 (0.09)
East European Ethnic	0.03 (0.11)	0.03 (0.1)	0.17 (0.11)	-0.09 (0.09)
West European Ethnic	0.37*** (0.07)	0.33*** (0.07)	0.44*** (0.07)	0.28*** (0.08)
Rank 1-10			1.08*** (0.07)	1.15*** (0.08)
Rank 11-20			1.05*** (0.07)	1.10*** (0.08)
Rank 21-50			0.83*** (0.07)	0.85*** (0.07)
Rank 51-100			0.40*** (0.09)	0.40*** (0.09)
Rank 101-250			-0.001 (0.01)	0.01 (0.1)
Gender			-0.08 (0.05)	-0.09 (0.06)
Country effect	No	Yes	No	Yes
Field effect	No	Yes	No	Yes
Residual Deviance	4640.1	4065.7	4099.2	3554.8
Degrees of Freedom	2661	2554	2549	2443

[2]Results from Poisson regressions and MLE. Dependent variable in all column is the number of interviews each applicants received. Interview samples are used. The coefficients are marginal effects of each variable. *p<0.05; **p<0.01; ***p<0.001

Table 8: **Regression Model for whether an application is successful** [3]

Variable	**(1)**	**(2)**	**(3)**	**(4)**
Intercept	-2.25*** (0.04)	-2.23*** (0.08)	0.09 (0.43)	0.09 (0.45)
East Asian	-0.60*** (0.07)	-0.65*** (0.07)	-0.46*** (0.07)	-0.52*** (0.08)
Indian Subcontinent	-0.44*** (0.11)	-0.48*** (0.11)	-0.25* (0.11)	-0.30** (0.11)
African Ethnic	-0.38** (0.14)	-0.38** (0.14)	-0.14 (0.14)	-0.13 (0.14)
Middle Eastern Ethnic	-0.67*** (0.12)	-0.67*** (0.12)	-0.47*** (0.12)	-0.49*** (0.13)
Judaism Ethnic	-0.01 (0.09)	0.02 (0.09)	0.04 (0.09)	-0.05 (0.09)
East European Ethnic	-0.23* (0.11)	-0.27* (0.11)	-0.09 (0.11)	-0.22 (0.11)
West European Ethnic	-0.10 (0.07)	-0.09 (0.07)	0.11 (0.07)	0.11 (0.08)
Rank 1-10			0.54*** (0.08)	0.57*** (0.08)
Rank 11-20			0.48*** (0.08)	0.49*** (0.08)
Rank 21-50			0.44*** (0.07)	0.42*** (0.07)
Rank 51-100			0.11 (0.09)	0.08 (0.09)
Rank 101-250			-0.01 (0.1)	-0.02 (0.1)
Log Mean Recommendation Rank			0.34*** (0.05)	0.35*** (0.05)
Gender			-0.05 (0.05)	-0.05 (0.06)
Field effect	No	Yes	No	Yes
Degrees of Freedom	24510	24478	23546	23515

[3]Results from logistics regressions and MLE. Dependent variable in all column is a binary of whether an application receives an interview. Full interview samples are used. The coefficients are marginal effects of each variable. *p<0.05; **p<0.01; ***p<0.001

Table 9: **Regression Model for whether an application is successful** [4]

Variable	**(1)**	**(2)**	**(3)**	**(4)**
Intercept	-2.31***	-2.11***	0.12	0.06
	(0.07)	(0.12)	(0.56)	(0.59)
East Asian	-0.45***	-0.54***	-0.41***	-0.50***
	(0.10)	(0.10)	(0.10)	(0.11)
Indian Subcontinent	-0.36	-0.47**	-0.29	-0.38*
	(0.14)	(0.15)	(0.15)	(0.15)
African Ethnic	-0.46*	-0.50**	-0.31	-0.30
	(0.19)	(0.19)	(0.20)	(0.20)
Middle Eastern Ethnic	-0.69***	-0.69***	-0.53**	-0.54**
	(0.17)	(0.17)	(0.17)	(0.18)
Judaism Ethnic	-0.10	-0.10	-0.14	-0.14
	(0.12)	(0.12)	(0.12)	(0.12)
East European Ethnic	-0.22	-0.27	-0.19	-0.23
	(0.15)	(0.16)	(0.16)	(0.16)
West European Ethnic	-0.31**	-0.36**	-0.30**	-0.35**
	(0.10)	(0.10)	(0.10)	(0.10)
Rank 1-10			0.91***	0.98***
			(0.12)	(0.12)
Rank 11-20			0.71***	0.70***
			(0.12)	(0.12)
Rank 21-50			0.79***	0.76***
			(0.11)	(0.10)
Rank 51-100			0.27*	0.25
			(0.14)	(0.14)
Rank 101-250			0.46***	0.40**
			(0.14)	(0.14)
Log Mean Recommendation Rank			0.39***	0.36***
			(0.07)	(0.07)
Gender			0.15*	0.12
			(0.08)	(0.08)
Field effect	No	Yes	No	Yes
Degrees of Freedom	14788	14758	14153	14124

[4]Results from logistics regressions and MLE. Dependent variable in all column is a binary of whether an application receives an interview. Applications to US-Canada are included in the sample. The coefficients are marginal effects of each variable. *p<0.05; **p<0.01; ***p<0.001

Table 10: **Regression Model for whether an application is successful** [5]

Variable	**(1)**	**(2)**
Intercept	-0.51 (0.61)	-0.65 (0.66)
English Ability	0.008*** (0.002)	0.01*** (0.0002)
Rank 1-10	0.92*** (0.13)	0.94*** (0.13)
Rank 11-20	0.74*** (0.13)	0.69*** (0.13)
Rank 21-50	0.82*** (0.11)	0.73*** (0.12)
Rank 51-100	0.26 (0.15)	0.18 (0.15)
Rank 101-250	0.44** (0.15)	0.34* (0.15)
Log Mean Recommendation Rank	0.41*** (0.07)	0.39*** (0.08)
Gender	0.18* (0.08)	0.15 (0.08)
Field effect	No	Yes
Degrees of Freedom		

[5] Results from logistics regressions and MLE. Dependent variable in all column is a binary of whether an application receives an interview. Full interview samples are used. The coefficients are marginal effects of each variable. *p<0.05; **p<0.01; ***p<0.001

Table 11: **Oaxaca Decomposition on Gender Differences** [6]

Decomposition	Total Gap	Unexplained Gap	Explained Gap
Returns to male are baseline	0.117	0.087	0.030
Return to female are baseline	0.117	0.093	0.024

Table 12: Oaxaca Decomposition on Ethnic Differences [7]

Decomposition	Total Gap	Unexplained Gap	Explained Gap
Returns to European are baseline	0.324	0.279	0.045
Return to non-European are baseline	0.324	0.260	0.064

[6]Regression controls are degree ranking of each applicants
[7]Regression controls are degree ranking of each applicants

[8] Results from logistics regressions and MLE. Dependent variable in all column is a binary of whether an application receives an interview. Full interview samples are used. The coefficients are marginal effects of each variable. *p<0.05; **p<0.01; ***p<0.001

Table 13: **Regression Model for whether an application is successful** [8]

Variable	**(1)**	**(2)**	**(3)**	**(4)**
Intercept	-3.2*** (0.24)	-1.84* (0.84)	-1.54 (0.86)	-2.55* (1.08)
East Asian			-0.53*** (0.14)	1.03 (0.76)
Indian Subcontinent			-0.40* (0.20)	-0.99 (1.07)
African Ethnic			-0.02 (0.25)	2.30 (1.42)
Middle Eastern Ethnic			-0.46 (0.25)	0.60 (1.23)
Judaism Ethnic			-0.30 (0.17)	0.62 (1.16)
East European Ethnic			-0.34 (0.21)	-0.28 (1.54)
West European Ethnic			-0.36* (0.14)	-0.49 (0.85)
Reading Score	0.05** (0.02)	0.04* (0.02)	0.04* (0.02)	0.09 (0.05)
Reading Score × West European Ethnic				0.01 (0.06)
Reading Score × East Asian				-0.12* (0.05)
Reading Score × Indian Subcontinent				0.04 (0.08)
Reading Score × African Ethnic				-0.17 (0.11)
Reading Score × Middle Eastern Ethnic				-0.08 (0.09)
Reading Score × Judaism Ethnic				-0.07 (0.08)
Reading Score × East European Ethnic				-0.01 (0.1)
Rank 1-10		1.09*** (0.16)	1.05*** (0.16)	1.10*** (0.16)
Rank 11-20		0.77*** (0.15)	0.73*** (0.16)	0.77*** (0.16)
Rank 21-50		0.79*** (0.14)	0.79*** (0.14)	0.83*** (0.14)
Rank 51-100		0.17 (0.18)	0.17 (0.18)	0.20 (0.18)
Rank 101-250		0.49** (0.19)	0.46* (0.19)	0.45* (0.19)
Log Average Recommendation Rank		0.22* (0.1)	0.22* (0.1)	0.18 (0.1)
Gender		0.06 (0.1)	0.11 (0.1)	0.12 (0.1)
Field effect	No	Yes	Yes	Yes
Degrees of Freedom	8503	8119	8111	8103