

Essays on discrimination in the marketplace

Linnaeus University Dissertations

No 213/2015

ESSAYS ON DISCRIMINATION
IN THE MARKETPLACE

LUCA FUMARCO

LINNAEUS UNIVERSITY PRESS

Essays on discrimination in the marketplace

Doctoral dissertation, Department of economics and statistics, Linnaeus University, Växjö, Sweden, 2015

ISBN: 978-91-87925-51-1

Published by: Linnaeus University Press, 351 95 Växjö

Printed by: Elanders Sverige AB, 2015

Abstract

Fumarco, Luca (2015). *Essays on discrimination in the marketplace*, Linnaeus University Dissertation No 213/2015, ISBN: 978-91-87925-51-1. Written in English

This thesis is composed of four self-contained papers and focuses on discrimination in the market place.

Essay 1: *“Disability Discrimination in the Rental Housing Market – A Field Experiment on Blind Tenants.”* Although discrimination against disabled people has been investigated in the labor market, the housing market has received less attention in this regard. This paper focuses on the latter market and investigates whether blind tenants assisted by guide dogs are discriminated against in the rental housing market. The data are collected through a field experiment in which written applications were sent in response to online advertisements posted by different types of advertisers. I find statistically significant evidence that one type of online advertiser, that is, the apartment owner (i.e., a person who advertises and rents out his/her own apartment(s) on his/her own), discriminates against blind tenants, because of the presence of the guide dog, not because of the disability. According to the legislation, this behavior qualifies as illegal discrimination.

Essay 2: *“Does the design of correspondence studies influence the measurement of discrimination?”* (co-authored with Carlsson and Rooth). Correspondence studies can identify the extent of discrimination in hiring as typically defined by the law, which includes discrimination against ethnic minorities and females. However, as Heckman and Siegelman (1993) show, if employers act upon a group difference in the variance of unobserved variables, this measure of discrimination may not be very informative. This issue has essentially been ignored in the empirical literature until the recent methodological development by Neumark (2012). We apply Neumark’s method to a number of already published correspondence studies. We find the Heckman and Siegelman critique relevant for empirical work and give suggestions on how future correspondence studies may address this critique.

Essay 3: *“Does Labor Market Tightness Affect Ethnic Discrimination in Hiring?”* (co-authored with Carlsson and Rooth). In this study, we investigate whether ethnic discrimination depends on labor market tightness. While ranking models predict a negative relationship, the prediction of screening models is ambiguous about the direction of the relationship. Thus, the direction of the relationship is purely an empirical issue. We utilize three (but combine into two) correspondence studies of the Swedish labor market and two distinctly different measures of labor market tightness. These different measures produce very similar results, showing that a one percent increase in labor market tightness increases ethnic discrimination in hiring by 0.5-0.7 percent, which is consistent with a screening model. This result stands in sharp contrast to the only previous study on this matter, Baert et al. (forthcoming), which finds evidence that supports a ranking model.

Essay 4: *“Relative Age Effect on Labor Market Outcomes for High Skilled Workers – Evidence from Soccer.”* In sports and education contexts, children are divided into age groups that are arbitrary constructions based on admission dates. This age-group system is thought to determine differences in maturity between pupils within the same group, that is, relative

age (RA). In turn, these within-age-group maturity differences produce performance gaps, that is, relative age effects (RAEs), which might persist and affect labor market outcomes. I analyze the RAE on labor market outcomes using a unique dataset of a particular group of high-skilled workers: soccer players in the Italian major soccer league. In line with previous studies, evidence on the existence of an RAE in terms of representativeness is found, meaning that players born relatively early in an age group are over-represented, while players born relatively late are under-represented, even accounting for specific population trends. Moreover, players born relatively late in an age group receive lower gross wages than players born relatively early. This wage gap seems to increase with age and in the quantile of the wage distribution.

Keywords: Disability Discrimination; Field Experiment; Housing Market; Correspondence Studies; Discrimination; Hiring Discrimination; Ethnic Discrimination; Labor Market Tightness; Ranking Models; Screening Models; Relative age; Labor Markets in Sports

"To Amber, and my whole family"

Acknowledgements

“Tous pour un, un pour tous”
Alexandre Dumas

My thesis work would have been impossible without the contribution of several people, inside and outside the academic world. I would like now to thank them all. I want to express my gratitude to my supervisor, Dan-Olof Rooth, and my assistant supervisor, Magnus Carlsson. They have given me guidance and feedback throughout the whole thesis. Both my supervisors have spent lots of time discussing with me ethics aspects of experiments and answering many trivial questions on academic life. I would like to thank them also for spurring me to travel and for teaching me to cultivate an academic network. They have had also the patience to deal with communication difficulties, and have provided me with technical and precise teaching on writing effectively academic papers within the economics field. They did much more, and have been the first people with whom I published a paper.

I have greatly benefited from the support of other researchers as well, from the Linnaeus University. Emma Neumann has been an important doctoral companion and friend from the very first moment I set my foot in Sweden. I thank Dominique Anxo for his human support and friendship. Simone Scarpa, a social work scholar, was always there to give me precious suggestions; he has reviewed part of my research and has given me directions on how to set up visiting periods. He is also an important friend, although we cannot agree either on soccer or culinary matters. Jonas Månsson, Håkan Locking, Lars Andersson and Peter Karlsson have given me the possibility to teach in a number of courses, which enriched me both as a person and as a researcher; with this respect, I would like to remember Thomas Lind, who introduced me to the teaching activity. I would like also to thank the rest of my department, for being always there to discuss of research.

I would like to extend my acknowledgements to Paul Nysted, who has been the opponent at the final seminar, and provided important feedback to improve my thesis.

I have a debt of gratitude also with people from the University of California, in Irvine. David Neumark, who supported my visiting period at UCI and provided me with important feedback on my thesis. I should thank other persons within UCI, namely Marianne Bitler, who has provided me with important feedback on my research, Patrick Button, who has been an important friend during my visiting period and commented my research, and Mary Ellen Wynn, who has given me probably the most fruitful English lessons ever, albeit she was not formally my teacher. With regards to this visiting period, I thank the Linnaeus University and FORTE (at the time, FAS) for their financial support.

I have greatly enjoyed the friendship of also all my other doctoral fellows, both economists and statisticians, from my same department. They have been emotional support and important research companions. I would like also to thank the administrative staff of the department, who helped me with a number of practical

issues. I would like to thank for their friendship and support also my fellow doctoral students from the Gothenburg University.

My former professors, instructors and classmates from the Università del Piemonte Orientale have always been there for me. Among other things, they have helped me in the first stage of one experiment conducted for my thesis and have received me for the Erasmus Teaching Assignment. And before that, they have instilled in me the love for economics, then they helped me to move my first steps abroad—especially with the master at the Université Rennes 1, and, finally, they have been supportive in moments of need. I would like to thank also my former professors at the Université Rennes 1.

The following paragraph is direct to my family, so it is in Italian. Un ringraziamento speciale va a tutta la mia famiglia, senza la quale non potrei essere qui, e con la quale ho speso davvero poco tempo da quando ho iniziato l'università. I miei genitori, Renata e Dino, mi hanno sempre incoraggiato, supportato ed insegnato tutto. Devo ringraziare anche mio fratello e le mie sorelle. Alessandro mi porta in avventurose uscite in mountain-bike, “senza cunisiun style;” Cinzia, che cerca sempre di non farmi sentire solo quando visito Mirabello; e Barbera, la quale rallegra sempre la compagnia. Ringrazio i miei nonni; in particolare, vorrei ricordare mio nonno Bruno, scomparso quando ero in USA. I miei zii, Giosuè ed Annalisa, cercano sempre di tirarmi su il morale con goliardiche cene in famiglia. Tutti hanno compreso le mie scelte che mi han condotto lontano dall'Italia, mi hanno motivato ed hanno partecipato ad alleviare momenti di stress del dottorato¹.

I must thank also my American side of the family. They represent fine examples of “Minnesota nice.” In particular, my wife Amber deserves special thanks. She has been supporting me through thick and thin. Not only she put up with me for five years by now, but she has also helped me with my written and oral English. She is also creator and author of the cover picture of this thesis. This doctoral degree is hers as much as mine.

Perhaps I have forgotten someone; please, pardon me, the list of people I should thank is very long.

¹ Special thanks goes to my family, without which I could not be here, and with whom I have spent really little time since I started being a university student. My parents, Renata and Dino, have always encouraged me, supported me and taught me everything. I have to thanks my siblings too. Alessandro brings me to reckless mountain-bike adventures; Cinzia, who always tries to make me feel not lonely when I visit Mirabello; and Barbera, who always spirits up the bunch. I thank my grandparents; in particular my grandpa Bruno, who disappeared while I was in the US. My uncles Giosuè and Annalisa, who always try to cheer me up with goliardic family dinners. All of them have understood my choices which have brought me away from Italy, they have motivated me and they have participated in alleviating doctorate periods stress.

Contents

Acknowledgements

Introduction

Essay 1: Disability Discrimination in the Rental Housing Market
– A Field Experiment on Blind Tenants

Essay 2: Does the design of correspondence studies influence the
measurement of discrimination?

Essay 3: Does Labor Market Tightness Affect Ethnic Discrimination in
Hiring?

Essay 4: Relative Age Effect on Labor Market Outcomes for High-Skilled
Workers – Evidence from Soccer

Content

Introduction	3
Disability Discrimination in the Rental Housing Market – A Field Experiment on Blind Tenants.....	4
Does the design of correspondence studies influence the measurement of discrimination?.....	6
Does Labor Market Tightness Affect Ethnic Discrimination in Hiring?	7
Relative Age Effect on Labor Market Outcomes for High Skilled Workers – Evidence from Soccer	9
Policy and Research Implications	10
References.....	11

Introduction

Contemporary discrimination is sporadically overt. Therefore, advanced techniques are required to expose discrimination and to bring it to the attention of the general public and policy makers, who ultimately adopt measures to address this problem. The economic field is equipped with an arsenal of advanced techniques that may help in this respect. Economists have used these tools to provide evidence of discrimination in different markets, such as the labor and product markets, in terms of different wages for the same job, different prices for the same good, different chances to find a job or to find housing, for instance.

One of the main techniques of investigation is the field experiment. It is of increasing popularity because of its attractive characteristics: It allows researchers to obtain identification through randomization, that is, the agents being studied are randomly allocated a specific treatment, the effect of which is studied by researchers; agent responses to the treatment are studied in natural environments, that is, in the “field;” and in some cases, the involved agents might not be aware that they are part of an experiment (Levitt & List, 2009). Researchers have control over the characteristics of the study, and at the same time, they can preserve a realistic setting (Levitt & List, 2009).

However, the economic discrimination literature based on field experiments presents a few shortcomings. It focuses on gender and ethnic discrimination, while discrimination against aged and disabled people (Riach & Rich, 2002), as well as discrimination based on sexual orientation, are less frequently studied. Moreover, field experiments seldom permit researchers to gain insights on the nature of discrimination (List, 2004). That is, studies are rarely able to explain whether certain groups of people are discriminated against because of personal preferences, stereotypes, or other factors. Information on the nature of discrimination is important to help policy makers in addressing the discrimination problem. Finally, although advanced, field experiments can still be improved. In fact, they may produce data that lacks information that is relevant in the real world but is unobservable to experimenters (e.g., Neumark, 2012; Pager, 2007; Riach & Rich, 2002;

Heckman, 1998; Heckman & Siegelman, 1993), thereby causing biased estimates.

This thesis addresses the above-mentioned shortcomings. It is composed of four self-contained essays, three of which use field experiments. The first essay contributes by analyzing discrimination against a group of tenants usually neglected by the economics literature: the visually impaired tenants. Moreover, this essay suggests a new strategy to gain insights on the nature of discrimination against this group of people. The analyses in this essay are conducted on data from a field experiment I have conducted in Italy. The second essay contributes to the field by illustrating the usage and results interpretation of an empirical methodology that has been recently created by Neumark (2012) to address criticisms against frequently used field experiments in the labor market related to the presence of unobservable characteristics. The third essay tries to gain more insight into the nature of discriminatory behavior in hiring. More specifically, this study focuses on the relationship between labor market tightness (i.e., the ratio between vacancy and unemployment rates) and ethnic discrimination in hiring. The analyses in this third essay, as well as in the second essay, are conducted on data from field experiments used in previously published studies. The fourth essay contributes by investigating the consequences, in the labor market, of early discriminatory treatment against a group of people that has only recently received more attention: “relatively young people” (i.e., people who are born near the end of the selection period that determines school classes and sports youth categories). This essay is different from all the others in terms of the data source. In fact, the analyses are based on a particularly rich panel dataset, in which information has been collected from websites and newspapers dedicated to sports. Under certain conditions, it is possible to claim that these data come from a natural experiment. The natural experiment differs from the field experiment because, although the allocation of the treatment occurred as well as randomly, this randomization is not induced by the researchers but is naturally found in the data (Levitt & List, 2009).

The remainder of this introduction includes a non-technical summary for each of the four essays and a conclusive section that contains both policy implications and recommendations for future related research.

Disability Discrimination in the Rental Housing Market – A Field Experiment on Blind Tenants

The first essay of this thesis investigates discrimination against blind tenants assisted by guide dogs in the rental housing market. In general, the studies that investigate the discrimination of disabled people in the housing market use the experimental methodology called “in-person audit test” (e.g., Turner et al., 2005). Within the context of the investigation of discrimination against blind

tenants with a guide dog, this technique would amount to matching two actors, who pretend to be housing applicants, over all characteristics except for two (i.e., one actor is blind and owns a guide dog, and the other actor is neither blind nor owns a dog). Both the non-disabled and the disabled actors visit the same housing agencies and inquire about available housing units. The disabled actor also requests the waiver of restrictions and/or of fees related to the presence of the assistance dog. Evidence of discrimination is found when the percentage of housing units made available for the disabled actor is statistically significantly lower than that made available for the non-disabled counterpart.

What are the possible problems with that methodology? First, usually, when this method is applied to the investigation of discrimination against disabled tenants, only a limited sample size is used, which precludes statistical tests from being performed. Second, the presence of “unobserved characteristics,”¹ that is, actor characteristics that are not appropriately accounted for and lead to biased results. Such characteristics might be subtle differences in the actors’ behavior with the housing brokers (Pager, 2007, and Riach & Rich, 2002). In the particular context of this study, which focuses on blind tenants with guide dogs, also the behavior of the dogs might be a source of unobservable characteristics. Third, the actors might be (sub)consciously motivated to gain evidence of discrimination and consequently adjust their behavior during the experiment; this problem is called “experimenter effect” (Pager, 2007).

The investigation carried out in this essay avoids these problems by utilizing data obtained through an alternative experimental technique: the “correspondence test,” that is, a field experiment in which written applications are matched, so that no actor is involved. The main difficulty with the correspondence test is that the target characteristic (in this case the disability status of a tenant) has to be signaled in a natural way. To circumvent this problem, in this experiment, blind applicants indirectly reveal their health conditions by including in their written application the information on the presence of the assistance dog. The experiment is implemented in Italy, and concerns two main groups of applicants: a couple of married tenants, and a couple of married tenants where the wife is blind and owns a guide dog. The data suggest that married tenants suffer from discriminatory treatment if the wife is blind and owns a guide dog; however, this result per se would not be very insightful.

To gain further insights on the causes of the discriminatory behavior and its possible remedies, I add two variations to the standard correspondence test. First, I include a second control group comprised of married tenants with a pet dog. When this group is compared to the group in which the wife is blind, it is

¹ The reason for using this term is that such characteristics are unobserved by the researcher, although they are observed by the housing brokers.

possible to investigate whether blind tenants are discriminated against simply because of their disability, assuming guide and pet dogs are similar from the point of view of the landlords. Second, I analyze two separate sub-samples of online advertisers of rental apartments: housing brokers (i.e., professional intermediary agents who advertise and rent out apartments that belong to someone else, who in return pays the housing brokers with a commission) and apartment owners (i.e., people who advertise and rent out their own apartments, taking care of the whole process on their own). The results provide evidence that household tenants that include a blind wife who is assisted by a guide dog are discriminated against by typical landlords because of the presence of the guide dog, not because of their disability.

Does the design of correspondence studies influence the measurement of discrimination?

The second essay of this thesis, co-authored with Magnus Carlsson and Dan-Olof Rooth, improves upon the understanding of a new econometric technique recently proposed by Neumark (2012), which makes it possible to gain unbiased evidence of discrimination from “correspondence tests” in the labor market. Correspondence tests, briefly described in the first essay of this thesis, are an increasingly popular method for measuring discriminatory treatment in the labor market as well (see Riach & Rich, 2002, for a survey). When the analysis focuses on measuring the extent of ethnic/gender discrimination, the target characteristics, that is, the ethnicity or gender of the applicant for a job, is signaled by the name of the applicant. The ethnicity/gender of the applicant is the only characteristic that differs between otherwise-equal matched pairs of applicants. As explained in Essay 1, the degree of discrimination in hiring is quantified by calculating the difference in the invitation rate to a job interview between the groups; if this difference is statistically significant, there is evidence of discrimination. This methodology provides a major advantage over the more traditional analyses of administrative data: It circumvents the problems caused by the presence of unobservable characteristics.

However, as attractive the correspondence test might be relative to the administrative data, it should be noted that it cannot differentiate between preference-based discrimination (Becker, 1957) and statistical discrimination (Aigner & Cain, 1977; Arrow, 1973; Phelps, 1972). Preference-based discrimination is based on employer prejudice, while statistical discrimination arises when employers behave differently based on perceived group differences, either in the mean or in the variance of the unobservable characteristics of the applicants. This implies that the results from a correspondence are only informative about the extent of discrimination and not the exact mechanism by which it occurs. An additional potential problem with this methodology, which this study focuses on, is that the results depend

on how the experiment is conducted. When employers perceive a group difference in the variance of unobserved characteristics,² the degree of discrimination revealed by a correspondence study depends on the standardization level of the fictitious job applicants, which is set by the experimenters. This level of discrimination is not representative for the whole population, and it corresponds only to that of female workers or workers from ethnic minorities with similar level of qualifications to those of the fictitious applicants.

This issue is at core in Neumark (2012). He proposes a method that consists of estimating the perceived relative variance of the unobservable characteristics across groups; this estimate is then decomposed into two parts. The first part captures discrimination in hiring due to employer preferences and/or employer stereotypes based on the average unobservable characteristics, while the second part captures discrimination in hiring due to employer stereotypes based on the variance of the unobservable characteristics. This second part represents the extent to which the estimated discrimination is affected simply by the level of standardization of the observed characteristics included in the job applications. The focus of our study is particularly on this part.

In the current study, we apply Neumark's method to data used in published papers (Carlsson & Rooth, 2007; Carlsson, 2010; and Rooth, 2011). These data are from three experiments conducted in the Swedish labor market between. Our results suggest that the measure of the estimated discrimination might in fact depend on perceived group differences in the variance of unobservable characteristics.

Does Labor Market Tightness Affect Ethnic Discrimination in Hiring?

The third essay of this thesis, co-authored with Magnus Carlsson and Dan-Olof Rooth, investigates whether labor market tightness influences the level of ethnic discrimination in hiring. Labor market tightness is defined by the vacancy-unemployment ratio. A high ratio means a tight labor market: in this case, there are many vacancies and few unemployed workers looking for jobs; thus vacancies are difficult to fill. In a slack labor market, the opposite is true: There are a few vacancies, and many unemployed workers are looking for a job; thus, vacancies are easy to fill. We refer to two economic theories to understand why and how the level of ethnic discrimination in hiring could depend on labor market tightness. Ranking models (e.g., Blanchard & Diamond, 1994) predict a negative relationship between the degree of ethnic

² The variance is a measure that describes, in a given dataset, how far on average each value of a variable is from the mean value of that variable.

discrimination and labor market tightness, and screening models (e.g., Vishwanath, 1989) predict a positive relationship.

To the best of our knowledge, this is only the second study to analyze this topic with data obtained through field experiments. Baert et al. (2014, henceforth, BCGV) have already investigated this relationship. They use data from a correspondence study in the Belgian labor market and find evidence supporting the ranking model because they find that ethnic discrimination is lower when labor market tightness is higher. However, the measure of labor market tightness used in BCGV lacks a necessary characteristic for being considered a valid measure of such a characteristic. There should be a positive general effect of the proxy on the likelihood of finding a job,³ but the measure of labor market tightness in the BCGV study seems to lack such a main effect.

In our study, we reexamine this relationship. We use data from published papers (Carlsson & Rooth, 2007; Carlsson, 2010; and Rooth, 2011); these data are from three separate correspondence studies that were conducted in the Swedish labor market and focused on ethnic discrimination. In our analyses, we use two measures of labor market tightness that fulfill the necessary property of having a general positive effect on the probability of finding a job. The first measure of labor market tightness is the callback rate of female applicants, which is taken from another correspondence study. This measure should closely mimic the degree of occupation-specific labor market tightness because other studies have demonstrated that, in Sweden, women are not discriminated against in hiring (see, e.g., Carlsson, 2011 and Eriksson & Lagerström, 2012). The second measure is the actual number of job applicants per job vacancy; this measure was collected a posteriori, through a telephone survey, for one of the three correspondence tests. Both of these measures present a general effect on the callback rate, being strongly and positively associated with the callback rate for native Swedish men.

As a further contribution of our study, we address potential omitted variable bias. This problem arises, causing bias in the estimates of discrimination, when important causal factors are left out of the empirical model used to investigate the data. We address this problem, controlling for heterogeneous effects at the occupation and firm level. In fact, a number of factors that affect the level of discrimination might vary across occupations and firms. Following this model adjustment, we can provide a clearer causal interpretation of the results.

We find that an increase in labor market tightness, using either of our two proxies, increases ethnic discrimination in hiring, which is supportive of a screening model.

³ That is, the likelihood of finding a job should be higher when vacancies are difficult to fill.

Relative Age Effect on Labor Market Outcomes for High Skilled Workers – Evidence from Soccer

The fourth essay of this thesis studies the long-run effect on labor market outcomes of chronological differences between peers born in the same year. There is extensive empirical evidence that children born late in the education and sport admission year are systematically disadvantaged throughout childhood until the late teens. This phenomenon is called the relative age effect and has gained popularity even outside the academic environment (e.g., Gladwell, 2008; Dubner & Levitt, 2010). Age-groups—both in education and in sports—are formed using arbitrary admission dates that determine some children to be older than others within the same age-group. This chronological difference is the “relative age (RA),” which is responsible for differences in maturity (e.g., Bedard & Dhuey, 2006; Musch & Hay, 1999). In turn, the RA causes a performance gap, called the “relative age effect (RAE),” which potentially affects children’s achievements. Because of its nature, this effect is expected to dissipate with age and eventually to disappear. However, it might persist or even widen because of the human capital accumulation process. This essay aims to investigate whether the RAE has long-lasting effects, which are even visible in the labor market. I study the long-run RAE on a particular group of high skilled workers: professional soccer players from the Italian major league, that is, Serie A. I use a unique dataset containing detailed information on players from the last seven seasons.

Two hypotheses are tested. The first hypothesis being tested is the presence of RAE in terms of representativeness. The RAE mechanism suggests that relatively old players are often perceived as more talented at early ages—when in fact they are initially just more mature. Because of this perceived higher level of talent, they are streamed (Allen & Barnsley, 1993) and reach Serie A more frequently than the relatively young peers. Any evidence of under-representation should hold even when accounting for the underlying birthdate distribution of the general population. The second hypothesis being tested is the RAE in terms of wage gaps. The RA framework does not provide clear expectations in this regard because the RAE in terms of wages depends upon several factors, which are discussed in detail in the essay.

The results reveal the presence of RAE in terms of representativeness and wage gaps. Italian players born at the beginning of the admission year are over-represented; moreover, this over-representation decreases and turns into under-representation as the end of the admission year approaches. This is an expected result based on the RAE theory. Furthermore, there is statistically significant evidence that players born toward the end of the admission year earn lower wages. Additional analyses suggest that the wage gap might be the largest at the entry of the labor market, and then, it drops and tends to increase for the remainder of the career. This particular development of the wage gap could be due to player career choices.

Policy Implications and Future Research

This thesis indicates that different groups, such as disabled tenants, workers with a foreign background and relatively young soccer players, all suffer from economic discrimination. Workers and tenants are discriminated in terms of opportunities to obtain a job and in terms of opportunities to find an apartment, respectively; whereas relatively young soccer players are under-represented and receive lower wages as long-term consequences of discriminatory treatment in youth. Moreover, this thesis indicates that the extent of the measured discrimination in the labor market might differ depending on the circumstances, such as the level of refinement of the experimental methodology and the labor market tightness. What are the policy implications and what could be the direction of future research based on the results from this thesis?

The experimental strategy used in the first essay provides evidence for disability discrimination and suggests the possible nature of this discrimination. With this additional knowledge, it is possible to recommend specific awareness and information campaigns to be implemented to decrease discrimination. Awareness campaigns could focus on the guide dog role in the assistance of a disabled owner; moreover, information campaigns should be conducted to educate landlords about the legislation that protects disabled tenants. Finally, the results in this essay call for future research to further explore disability discrimination in the rental housing market. The field experiment employed in this essay could be extended to the investigation, for instance, of discrimination against tenants with mobility impairments; research on this group of tenants could provide important knowledge on discrimination against elderly tenants because the likelihood to develop mobility impairments increases with age.

The empirical method illustrated in the second essay allows the measurement of the bias in the initial estimate of discrimination from field experiments. As a possible solution to this bias, we propose that future experimenters invest more effort in the preparation of the experimental design. Particular attention should be paid to the preparation of job applications when the experiment addresses discrimination in hiring; it is possible that online databases could be useful in this respect. Future researchers may also try to apply this same empirical methodology to data from experiments on discrimination in the housing market.

The results from the third essay suggest that ethnic discrimination increases in improved labor market conditions. This result is opposite to that obtained in the research on both ethnic employment gaps and the business cycle. As a consequence, policy makers should keep in place measures to combat ethnic discrimination in hiring even when the business cycle improves.

The results from the fourth essay reveal that professional soccer players born toward the end of the selection year are under-represented in the Italian

major league, and on average, they receive a lower wage compared to their older counterparts. Additional analyses may encourage future studies to employ the so-called “quantile regression” in the study of wage gaps; this is motivated by empirical and theoretical reasons described in more detail in the essay. However, the results of this study should be considered carefully: The lack of variation in the admission dates does not allow me to definitely rule out different explanations for the under-representation and lower wages of relatively young players. Therefore, I recommend future studies on this same topic to exploit data characterized by variation in the admission date to gain clearer results. Moreover, other aspects of the relative age effect on labor market and education outcomes should be analyzed. If further studies supported the existence of long-run RAEs, a restriction of the age-groups could be beneficial (e.g., 6-9 months in lieu of 12-24 months); the periodic rotation of admission dates could also have positive effects because the RAE would not consistently provide advantages to people born in a given month (Wattie et. al, 2015).

References

- Aigner, D. J., & Cain, G. G. (1977). Statistical theories of discrimination in labor markets. *Industrial and Labor relations review*, 30(2), 175-187.
- Allen, J., & Barnsley, R. (1993). Streams and Tiers: The Interaction of Ability, Maturity, and Training in Systems with Age-Dependent Recursive Selection. *Journal of Human Resources*, 28(3), 649-659.
- Arrow, K. (1973). The theory of discrimination. In: Ashenfelter O, Rees A (eds) *Discrimination in labor markets*. Princeton University Press, Princeton.
- Baert, S., Cockx, B., Gheyle, N., & Vandamme, C. (2014). Is There Less Discrimination in Occupations where Recruitment is Difficult, forthcoming in *Industrial and Labor Relations Review*.
- Becker, G.S. (1957). *The economics of discrimination*. University of Chicago Press, Chicago.
- Bedard, K., & Dhuey, E. (2006). The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects. *The Quarterly Journal of Economics*, 121(4), 1437-1472.
- Blanchard, O. J., & Diamond, P. A. (1994). Ranking, Unemployment Duration, and Wages. *Review of Economic Studies*, 61(3), 417-434.
- Carlsson, M. (2011). Does hiring discrimination cause gender segregation in the Swedish labor market? *Feminist Economics*, 17(3), 71-102.
- Carlsson, M. (2010). Experimental Evidence of discrimination in the Hiring of First- and Second-generation Immigrants. *Labour*, 24 (3), 267-278.
- Carlsson, M., & D.-O., R. (2007). Evidence of Ethnic Discrimination in the Swedish Labor Market Using Experimental Data. *Labor Economics*, 14 (4), 716-729.
- Dubner, S. J., & Levitt, S. D. (2010). *Superfreakonomics: Global Cooling, Patriotic Prostitutes and Why Suicide Bombers Should Buy Life Insurance*. Penguin UK.

- Eriksson, S., & Lagerström, J. (2012). Detecting discrimination in the hiring process: evidence from an Internet-based search channel. *Empirical Economics*, 43(2), 537-563.
- Gladwell, M. (2008). *Outliers: The story of success*. Hachette UK.
- Heckman, J.J. (1998). Detecting discrimination. *Journal of Economic Perspective*, 12(2), 101-116.
- Heckman, J.J., & Siegelman, P. (1993). The Urban Institute audit studies: Their methods and findings. In: Fix M, Struyk R (eds). *Clear and convincing evidence: Measurement of discrimination in America*. Urban Institute Press, Washington DC, 187-258.
- Levitt, S. D., & List, J. A. (2009). Field Experiments in Economics: The Past, the Present, and the Future. *European Economic Review*, 53(1), 1-18.
- List, J. A. (2004). The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field. *The Quarterly Journal of Economics*, 119(1), 49-89.
- Musch, J., & Hay, R. (1999). The Relative Age Effect in Soccer: Cross-Cultural Evidence for a Systematic Distribution Against Children Born Late in the Competition Year. *Sociology of Sport Journal*, 16(1), 54-6.
- Neumark, D. (2012). Detecting discrimination in audit and correspondence studies. *Journal of Human Resources*, 47(4), 1128-1157.
- Pager, D. (2007). 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.
- Phelps, E. S. (1972). The statistical theory of racism and sexism. *The American Economic review*, 62(4), 659-661.
- Riach, P. A., & Rich, J. (2002). Field Experiments of Discrimination in the Market Place. *The economic journal*, 112(483), F480-F518.
- Rooth, D.-O. (2011). Work Out or Out of Work - The Labor Market Return to Physical Fitness and Leisure Sports Activities. *Labour Economics*, 18 (3), 399-409.
- Turner, M. A., Herbig, C., Kaye, D., Fenderson, J., & Levy, D. (2005). *Discrimination against Persons with Disabilities: Barriers at Every Step*. Washington D.C.: HUD USER, Department of Housing and Urban Development, 39105.
- Vishwanath, T. (1989). Job Search, Stigma Effect, and Escape Rate from Unemployment. *Journal of Labor Economics*, 7(4), 487-502.
- Wattie, N., Schorer, J., & Baker, J. (2015). The Relative Age Effect in Sport: A Developmental Systems Model. *Sports Medicine*, 45(1), 83-94.

I

Disability Discrimination in the Rental Housing Market – A Field Experiment on Blind Tenants

by

Luca Fumarco^a

Abstract. Although discrimination against disabled people has been investigated in the labor market, the housing market has received less attention in this regard. This paper focuses on the latter market and investigates whether blind tenants assisted by guide dogs are discriminated against in the rental housing market. The data are collected through a field experiment in which written applications were sent in response to online advertisements posted by different types of advertisers. I find statistically significant evidence that one type of online advertiser, that is, the apartment owner (i.e., a person who advertises and rents out his/her own apartment(s) on his/her own), discriminates against blind tenants, because of the presence of the guide dog, not because of the disability. According to the legislation, this behavior qualifies as illegal discrimination.

JEL-Classification: J14; R21

Keywords: Disability Discrimination; Field Experiment; Housing Market

^a Linnaeus University, e-mail: luca.fumarco@lnu.se

Acknowledgements. I would like to thank: my supervisors Dan-Olof Rooth and Magnus Carlsson for guidance and many useful suggestions; Simone Scarpa, Giangiacomo Bravo and Stijn Baert, for providing helpful feedback; the participants of the seminars series of the Centre for Labour Market and Discrimination Studies, at the Linnaeus University; PhD fellows from the Department of Economics and Statistics of the Linnaeus University; the participants of the Annual Ph.D. Conference in Economic Demography, Lund University; Piermassimo Pavese for the provision of some administrative data; Stefania Ottone, Ferruccio Ponzano and Matteo Migheli for logistic support.

1 Introduction

Over 2.7 million people are blind in Europe (Pascolini & Mariotti, 2012).¹ This population, along with those who are affected by other types of impairments, is discriminated against despite the protection provided by the European Union legislation (e.g., Jones, 2008, for disability discrimination in the labor market). Art. 21 of the Charter of Fundamental Rights of the European Union, 2000, states that any discrimination based on disability shall be prohibited.² Although discrimination against disabled people has been investigated in the labor market (Jones, 2008, for a literature review), the housing market has received less attention in this regard. However, also this type of discrimination deserves attention. A dwelling is a primary necessity that determines social inclusion, job opportunities, and the enjoyment of public services; therefore, discrimination in the housing market against disabled people harms their quality of life.

Studies that have investigated whether blind tenants, and more in general impaired tenants, are discriminated against in the housing market have used the experimental methodology referred to as the “in-person audit test.” Within the context of disability discrimination, this experimental technique is used in the US by the Urban Institute (Turner et al., 2005) and by specialized law firms for audit reports (e.g., Murphy, 2007). Two actors, playing the role of tenants enquiring about housing, are matched over all characteristics except for one (e.g., one actor is blind

¹ I use the term “blind” because I am referring to persons who specifically have the maximum visual impairment. Alternative terms, such as “visually impaired,” lack specificity regarding the intensity of the impairment; differently, there are clinical and legal parameters that define whether a person is “blind.” For example, according to the WHO definition, a person is to be considered blind when her best eye acuity is lower than 20/500 or is characterized by a visual field lower than 10 degrees (Maberley et al., 2006). The disability benefit systems used by individual countries are based on slightly different legal definitions.

² There also are precise EU directives to which individual member countries should align their legislations. The 2000/78/CE Directive also prohibits employment discrimination on grounds of disability. Directive Proposal COM, 2008, 426, aims at also protecting disabled people in the employment market; because this directive is still a proposal, single countries’ legislations remain heterogeneous on this matter.

and the other is not) or two (e.g., one actor is blind and owns a guide dog; the other actor neither is blind nor owns a dog). The disabled actor visits a number of housing agencies to inquire about available housing units; if he owns an assistance dog, he also requests a waiver of restrictions on and/or fees related to the dog. The non-disabled actor who owns no dog visits the same housing agencies and inquires about housing units. There is evidence of discrimination against disabled tenants when the percentage of housing units made available for the disabled actor is statistically significantly lower than that made available for the non-disabled actor (e.g., Turner et al., 2005, for a similar analysis on discrimination against deaf tenants).³ Additionally, there is evidence of discrimination whenever the blind applicant receives a refusal to accommodate his guide dog (e.g., Turner et al., 2005; Murphy, 2007).⁴

Results from American in-person audit tests suggest that blind tenants might be treated differently, with a lower percentage of housing units made available to them (e.g., Murphy, 2007). Additionally, blind tenants receive refusals to waive restrictions and/or fees on their guide dogs (e.g., Murphy, 2007).

There are at least three concerns with these studies. First, in-person audit experiments on disabled tenants do not typically involve large samples and do not perform statistical hypothesis testing (Turner et al., 2005), so that it is not possible to infer predictions about larger populations than the sample. Second, in general, in-person audit tests may produce biased results if some actors' characteristics are not appropriately accounted for (Heckman & Siegelman, 1993; Heckman,

³ Other studies investigate discrimination against blind tenants but do not make any statistical inferences. Turner et al. (2005) propose an experimental design to investigate discrimination against blind tenants with and without guide dogs and then conduct a pilot experiment, which is not suitable for statistical inference. Murphy (2007) is an example of the type of audit report typically implemented by American law firms; although these experiments use matched pairs of applicants, they do not generally implement any statistical inference.

⁴ In the US, refusal to accommodate guide dogs violates the Reasonable Accommodation Under the Fair Housing Act, 2004.

1998; Riach and Rich, 2002; Pager, 2007), so that these characteristics are observed by the landlords but not by the researchers. In the jargon used for experiments in the social sciences, it is standard to call them unobserved characteristics, and they might result from poorly matching the actors. Pager (2007) and Riach and Rich (2002) suggest that such characteristics might be reflected in subtle differences in how applicant tenants conduct the interactions with the housing brokers. Furthermore, in studies on blind tenants with a guide dog, the dog's features might represent an additional source of unobservable disturbances.⁵ Taken together, these unobservable characteristics might unwillingly convey systematically different information on the matched applicants, driving the differential treatment and thus causing biased estimates of discrimination (Pager, 2007). Third, a serious general threat to the validity of in-person audit tests is the experimenter effect (Pager, 2007). Actors might be (sub)consciously motivated to obtain evidence of discrimination and consequently adjust their behavior during their interactions with brokers (Pager, 2007).

This study circumvents these three problems by utilizing data from a written field experiment, known as correspondence test, which permits obtaining unbiased estimates of discrimination (Riach & Rich, 2002).⁶ Written applications are delivered via the Internet, which facilitates contacting a large number of landlords and thus allowing for statistical inferences to be made. Time and budget limitations are much less important than they are in-person audit experiments because it takes only a few additional minutes to send additional inquiries and register the re-

⁵ For instance, the conditions of the dog's hair, its smell and its behavior during interactions with housing brokers might be perceived as relevant and affect the experimental results if they were interpreted by the brokers as signals of the applicant's socioeconomic status and ability to manage his possessions.

⁶ Because of the unbiased results, this methodology is also used to investigate other types of discrimination in the rental housing market, namely, discrimination based on ethnicity (e.g., for Italy, Baldini & Federici, 2011; for Sweden, Carlsson & Eriksson, 2013), gender (e.g., for Sweden, Ahmed & Hammarstedt, 2008), sexual orientation (e.g., for Sweden, Ahmed & Hammarstedt, 2009), age and employment status (for Sweden, Carlsson & Eriksson, 2013).

sponses and the cost of contacting additional landlords is virtually null. Because no actor is involved, the use of written field experiments also allows for almost perfect control over the applications (Pager, 2007), thus reducing the bias caused by unobservable characteristics and by the experimenter effect.

However, correspondence tests present one problem. The target characteristic, in this case the blindness, has to be signaled in a way that does not seem unnatural (Pager, 2007).⁷ Blind applicants would rarely reveal their health conditions directly when applying for apartments. However, because guide dogs are an international symbol for the blind and dogs in general are often seen as family members by their owners, this experiment solves this problem by mentioning in the application the presence of the assistance dog.⁸ This information serves as a cue for the tenant's disability.

The actual implementation of the correspondence test is simple. First, I prepared the bogus applications for the control group, composed of married tenants,⁹ and for the target group, composed of married tenants where the wife is blind.¹⁰ Then, I sent these applications in response to online housing advertisements in Italy. Finally, I compared the invitation rates to visit the apartments across the control and target groups.

⁷ For instance, in experiments aimed at analyzing ethnic or gender discrimination in hiring, the applicant's name is used as a signal (e.g., for Sweden, Carlsson & Rooth, 2007; for the US, Bertrand & Mullainathan, 2004).

⁸ For instance, in the US and in Italy, many people consider their dogs friends/family members (Eurispes, 2013, and the online website of the American Veterinary Medical Association). Thus, informing a landlord about the presence of the guide dog seems a natural behavior.

⁹ The usage of households rather than individuals is a strategy that is also adopted in studies on sexual orientation discrimination in the rental housing market (e.g., for Sweden, Ahmed & Hammarstedt, 2009), where the tenants' sexual orientation is indirectly communicated through a brief description of the household.

¹⁰ I do not opt for married tenants where the husband is blind to avoid the risk that online advertisers would have discriminated these households based on expected lower household income, which would confound the estimates for disability discrimination. This risk is attenuated with a blind wife because women's employment rates and wages are already lower than men's in many OECD countries. Source: OECD Employment and Labour Market Statistics.

Although Italian legislation protects disabled people from discrimination (Italian Law no. 67, 2006),¹¹ the data suggest that on average married tenants with a blind wife assisted by a guide dog are less likely to receive an invitation to visit the apartment.

However, this information alone would say little about the reason online advertisers of apartments for rent discriminate and how to possibly reduce this discrimination. Therefore, I add two variations to the standard correspondence tests. First, this test includes a second control group composed of married tenants with a pet dog. The comparison between the invitation rates to visit apartments for this second control group and for married tenants where the wife is blind permits investigating whether blind tenants assisted by a guide dog are discriminated against specifically because of their disability. Second, I analyze two separate sub-samples of online advertisers. There are “housing brokers,” that is, professional intermediary agents who advertise and rent out apartments that belong to someone else, who in return pays the housing brokers’ commissions,¹² and there are “apartment owners,” who advertise and rent out their own apartments, managing the process entirely on their own;¹³ no housing broker is involved in managing these apartments. It is likely that these two groups of online advertisers differ in terms of knowledge of and respect for the legislation. In contrast to apartment owners, in Italy, housing brokers must either pass a specific exam or spend a period of at least 12 months as practitioners in a housing agency in order to obtain a brokerage license (Federazione Italiana Mediatori Agenti d’Affari, 2006). Thus, housing brokers are more likely to have complete knowledge of the legislation.

¹¹ According to this law, disabled people cannot be discriminated against based on their impairment (i.e., blindness) or, more indirectly, on factors related to it (i.e., the presence of the guide dog). This law embodies the principles suggested in Directive Proposal COM, 2008, 426.

¹² For instance, after the housing broker has found a new tenant for an apartment, she might be paid the equivalent of one month of rent, or a certain percentage of the total yearly rent, by the actual apartment owner.

¹³ For instance, apartment owners might rent out part of the housing unit they are living in, or they might own a few apartments and rent out some of them.

They also conduct many more housing transactions, which increases their risk of being taken to court in case of misconduct and thus might decrease illegal behaviors. All things considered, a legitimate expectation seems to be that housing brokers discriminate against disabled tenants less often than apartment owners. If this is the case, policies aiming at decreasing discrimination against blind tenants should be tailored to apartment owners.

The results provide evidence that household tenants where the wife is blind and is assisted by a guide dog are discriminated against by apartment owners because of the presence of the guide dog, not because of their disability, and this behavior qualifies as discrimination. In contrast, there is no evidence of discrimination from housing brokers.

The remainder of the paper is organized as follows. In Section 2, I present the experiment design and the descriptive results. Section 3 describes the model and the results. Section 4 concludes.

2 Experiment Design and Descriptive Results

2.1 Experiment Design

I use data from a correspondence test I conducted in the Italian rental housing market from the 12th of April to the 22nd of June, 2013. I sent 1,000 fictitious written applications in response to advertisements on the Italian classified website *Subito.it*.

The fictitious applications are “non-paired,” meaning that each online advertiser received only one application. The usage of non-paired written applications provides at least four advantages over the paired written applications technique. First, only one fictitious application per group needs to be sent to a landlord, further reducing the risk of being exposed. Second, one application per advertiser minimizes the advertisers’ inconvenience because each receives only one

inquiry. Third, it is possible to apply for a larger sample of house units over a given period. Fourth, this methodology is not affected by any bias related to the order of inquiries.

The randomization of the applications occurred as follows. I created three tenants' identities (i.e., Andrea Rossi, Francesco Russo, and Alessandro Ferrari)¹⁴ and also an email account for each of them.¹⁵ Then, I entered them into a spreadsheet on which the three identities were repeated approximately 330 times each; after that, I randomly ordered the identities by drawing without resampling using a normal distribution. Afterwards, I created three applicants' statuses (i.e., married tenants, married tenants with blind tenant plus guide dog, and married tenants with pet dog), I prepared a list on which the applicants' statuses were repeated approximately 330 times each, and again I randomly ordered them by drawing without resampling using a normal distribution. Thereafter, I paired the list of applicants' statuses with that of applicants' identities. Finally, for all of the fictitious applicants, I also randomly determined the general location of the apartment for which they would be applying (i.e., the region of the apartment and whether the apartment had to be in a metropolitan city or not).¹⁶ The number of applicants per region reflects the proportion of the regional population in the national population, and within each region, the number of applicants per metropolitan city reflects the proportion of its population in the regional population.¹⁷

¹⁴ These are the most frequent Italian names and surnames. Sources: demo.istat.it and italygen.com. The matching name/surname occurred randomly.

¹⁵ The email accounts took the form of `name.surname####@gmail.com`.

¹⁶ Similar to the two previous steps, I first created the list of randomly ordered locations, and then I paired this list with the list of applicants.

¹⁷ A "metropolitan city" is an administrative institution that is expected to be operative as of 2015. However, this administrative institution was already described in Law no. 142, 1990, and, before that, it was solicited in the Italian Constitution, art. 114. A metropolitan city includes a large core city and its smaller surrounding towns; the core city and towns are closely related in terms of economic activities, provision of public services, cultural aspects and territorial features. The metropolitan cities have large populations, ranging from hundreds of thousands to a few million inhabitants.

Because of the particular nature of this experiment, I use household tenants rather than single tenants. Although technologies exist that allow blind persons to use computers, some people might ignore their existence. There would have been the risk of detection if some landlords viewed with suspicion applications written by blind persons. Therefore, each application was written by the (normal-sighted) husband, who revealed the composition of the household, which implied revealing whether the wife owned a dog and whether this dog was either a pet or a guide dog.

The standard application message for a vacant housing unit can be translated as follows:

“Good xxxxx,

My family is interested in the apartment for rent described in the advertisement you posted at the website Subito.it. I would like to move in with my wife [and her (guide) dog].

If the apartment is still available, we would like to visit it.”

When the actual experiment began, I sent applications, following the list of applicants with assigned apartment locations. Applications were sent to the most recent advertisements in order to minimize the probability of contacting an advertiser whose apartment had already been rented out. However, advertisements for apartments smaller than 40 square meters or more expensive than 1,500€ per month were not taken into consideration. The size restriction was adopted because of the Italian law that legislates the maximum number of tenants per square meter;¹⁸

¹⁸ Ministerial Decree, 20th of June, 1975, art. 5, and its modification in 1986.

this limit does not apply to pets. The rent restriction was adopted to avoid overly conservative estimates of discrimination,¹⁹ assuming that discrimination is lower for very high rents.

2.2 Descriptive Results

The mean invitation rate for each of the three groups is presented in Table 1. A landlord's response represents an invitation to visit the apartment either when it includes a direct invitation or when it requests a phone call to discuss day and time of the visit.

Table 1. Mean invitation rates for the groups of applicants.

Tenants' group	N	Mean	Std. Dev.
	(1)	(2)	(3)
A (Not blind, no dog)	334	0.784	0.412
B (Blind, guide dog)	332	0.666	0.472
C (Not blind, pet dog)	334	0.635	0.482

Note: Columns (1)-(2)-(3) report, respectively: number of observations and means and standard deviations of groups' invitation rates.

Group A is the reference group and comprises applications from married tenants; Group B consists of applications from married tenants where the wife is blind and is assisted by a guide dog; and Group C is composed of applications from married tenants with a pet dog.

This table helps to illustrate how the three-group comparison strategy allows for more insights. Comparing the mean invitation rates for Groups A and B suggests the presence of disability discrimination. In fact, according to the Italian legislation, landlords cannot discriminate based on either i) disability status (i.e., being blind) or ii) specific conditions related to that disability (i.e., owning a guide dog); these are the two characteristics that differ between groups A and B.

¹⁹ The estimated average monthly rent in Italy is 1,000€. Source: number.com.

Comparing Groups A and C suggests the presence of discriminatory behaviors against pet dog owners, which is legal; the only difference between these two groups is the presence of a pet dog in Group C. Comparing the invitation rates for Groups B and C suggests that tenants with a guide dog are treated equivalently to tenants with a common pet dog.²⁰ Additional insights can be gained from these comparisons contingent on what I call the “*equality assumption*.” This assumption requires that, from the point of view of the online advertisers, there were no differences between pet and guide dogs in terms of burden on the apartment and that online advertisers had no preferences for one over the other.²¹ If the reader accepted this assumption, the wife’s blindness would be the only difference between Groups B and C, and thus, this comparison would suggest that blindness per se is not the cause of discrimination. Consequently, because there appeared to be no discrimination based on blindness, the gap of 12 percentage points in the mean invitation rates for Groups A and B might be caused by the presence of the guide dog alone. Independent from the reason landlords discriminate based on the presence of the guide dog—they are unaware of the legislation or they consciously ignore it—this form of discrimination is illegal. Vice-versa, if the reader did not accept the equality assumption, the finding that blind tenants are discriminated against would still hold, although it would not be possible to gain any additional insights on the causes of discrimination.²²

²⁰ Independent group t-tests for the three comparisons give statistical support to the interpretation of the data in Table 1. The difference in the invitation rates between Groups A and B is highly statistically significant ($df=664$, $t=-3.459$, $p=0.000$); that between Groups A and C is also highly statistically significant ($df=666$, $t=-4.314$, $p=0.000$); the difference in the invitation rates between Groups B and C is not statistically significant at any standard level ($df=664$, $t=-0.836$, $p=0.403$).

²¹ Accounts from members of associations for the visually impaired people lead to the thought that this is a credible assumption. Many Italians appear to consider guide dogs as having the same behavior as pet dogs within housing units (e.g., they bark and scratch the furniture), even though guide dogs are trained and selected from particular breeds.

²² If the equality assumption does not hold, it is not possible to understand whether discrimination were due to the guide dog or to the blindness status. In fact, the similar invitation rates (lower than that of the reference group) for

The analysis of separate sub-samples for each type of online advertiser offers additional insights. In the Italian rental housing market, there are two very different online advertisers, namely, housing brokers and apartment owners. As was previously explained, housing brokers have greater knowledge of the legislation; moreover, they conduct many more housing market transactions, which increases their risk of being taken to court in case of misconduct and thus decreases illegal behaviors on their parts. Therefore, it should be reasonable to expect them to be more respectful of the legislation that protects disabled tenants. Table 2 reports the main statistics for the three groups of tenants, divided by online advertiser type.

Table 2. Mean invitation rates for the groups of applicants by online advertiser type.

Tenants' groups	Online advertiser type					
	Housing brokers			Apartment owners		
	N	Mean	Std. Dev.	N	Mean	Std. Dev.
	(1)	(2)	(3)	(4)	(5)	(6)
A (Not blind, no dog)	138	0.819	0.386	191	0.764	0.425
B (Blind, guide dog)	133	0.812	0.392	191	0.555	0.498
C (Not blind, pet dog)	148	0.709	0.455	183	0.579	0.495

Note: In the panel “Housing brokers”, columns (1)-(2)-(3) report, respectively: number of observations and means and standard deviations of groups’ invitation rates from housing brokers. In the panel “Apartment owners”, columns (4)-(5)-(6) report the equivalent statistics for apartment owners. For each group of tenants, the total number of observations in the two panels is lower than that reported in Table 1 because of a small number of missing observations about the online advertisers’ type.

tenants with pet dogs and blind tenants with guide dogs might be a result of different combinations of landlords’ preferences for (guide) dogs and their owners. For instance, suppose that landlords are actually indifferent toward guide dogs (which are usually trained Labradors or Golden Retrievers or German Shepherds), so that they are not bothered by their presence within the housing unit, but they do not like pet dogs (a much broader range of species, which also include Chihuahuas or Bull Terriers, and they are not necessarily trained) because they think pet dogs would more likely ruin the apartment or annoy the neighbors. At the same time, landlords might discriminate against blind tenants only because of their disability. In this case, by chance, blind tenants’ invitation rate might be the same as that for tenants with pet dogs. Other combinations of landlords’ discriminatory behaviors may lead to similar results. In these cases, the policy implications of the results would be similar, although less specific, to those mentioned in the concluding section.

As expected, the statistics in Table 2 suggest that the two types of online advertisers behave very differently: housing brokers do not discriminate against blind tenants with guide dogs, whereas apartment owners do discriminate.²³

3 Model and Results

In this section, I further analyze the data using a linear probability model. The main purpose of this analysis is to verify whether the randomization process described in Section 2 worked as intended. The dependent variable, $Invitation_rate_i$, is a dummy that equals one if the applicant received an invitation to visit an apartment and zero otherwise. This outcome variable is regressed on two variables of interest: the dummy $Blind_i$, where i is a household with a blind wife who owns a guide dog, and Dog_i , where i is a household with a pet dog.²⁴ A vector of control variables, X_i , includes apartment and advertisement characteristics: logarithms of apartment square meters and of monthly rent; rent per square meter;²⁵ a dummy for the apartment being in a metropolitan city; a dummy for the apartment being furnished; and dummies for apartment pictures as well as for a phone number in the advertisement. The variables for the logarithms of apartment square meters and of monthly rent as well as for the rent per square meter were rescaled by subtracting their minimum values; the rent per square meter is also rescaled, specifically, divided by

²³ Independent group t-tests for the three comparisons in the two sub-samples of advertisers give statistical support to the interpretation of the data in Table 2. In the sub-sample of housing brokers, the difference in the invitation rates between Groups A and B is not statistically significant (df=269, t=-0.144, p=0.885). In contrast, the differences between Groups A and C and Groups B and C are statistically significant (respectively, df=284, t=-2.182, p=0.030 and df=279, t=-2.012, p=0.045). In the sub-sample of apartment owners, the differences in the invitation rates between Groups A and B as well as Groups A and C are highly statistically significant (respectively, df=380, t=-4.417, p=0.000, and df=372, t=-3.884, p=0.000), but the difference between Groups B and C is not (df=372, t=0.472, p=0.637).

²⁴ Therefore, household tenants without a pet dog and where the wife is not disabled (Group A) is the reference group.

²⁵ This variable is intended to act similarly to an interaction between the logarithm of apartment square meters and the logarithm of monthly rent.

10. The model also includes F_i , a vector of fixed effects,²⁶ as well as a stochastic individual term, ε_i . The linear probability model looks as follows:

$$Invitation_rate_i = \beta_0 + \beta_1 Blind_i + \beta_2 Dog_i + \beta X_i + \beta F_i + \varepsilon_i$$

This model is first estimated only with the variables of interest, then the vector of control variables is added, and finally the vector of fixed effects is also included. Because the descriptive results in Section 2 suggest that housing brokers and apartment owners are quite different in terms of discriminatory behaviors, the model is estimated on the two separate subsamples of online advertisers. The main estimates are reported in Table 3.

Table 3. Linear probability model main estimates by online advertiser type.

	Housing brokers			Apartment owners		
	(1.A)	(1.B)	(1.C)	(2.A)	(2.B)	(2.C)
Constant	0.818*** (0.032)	0.581*** (0.170)	0.621*** (0.188)	0.764*** (0.034)	0.940*** (0.098)	0.983*** (0.109)
$\hat{\beta}_1$ (Blind, guide dog)	-0.007 (0.045)	-0.008 (0.047)	-0.009 (0.050)	-0.209*** (0.065)	-0.225*** (0.061)	-0.238*** (0.064)
$\hat{\beta}_2$ (Not blind, pet dog)	-0.109*** (0.041)	-0.117*** (0.040)	-0.111*** (0.043)	-0.185*** (0.041)	-0.211*** (0.042)	-0.234*** (0.041)
Control variables	N	Y	Y	N	Y	Y
Fixed-effects	N	N	Y	N	N	Y
P-value ($\hat{\beta}_1 - \hat{\beta}_2$)	0.029	0.027	0.043	0.705	0.833	0.952
R-squared	0.015	0.031	0.072	0.038	0.067	0.127
N	419	409	409	565	526	526

²⁶ This vector includes: dummies for each applicant's identity (e.g., Andrea Rossi and Francesco Russo; Alessandro Ferrari is the base identity), dummies for each Italian region (with Lombardy being the base region) and a dummy for the application being sent after a new condominium national regulation came into force (i.e., Law no. 220, 11th of December 2012, which came into force on the 18th of June, 2013). According to this law, condominium regulations can no longer include pet restrictions, but this does not apply to other types of apartments. For most of the advertisements, I was not able to differentiate between condominium apartments and other types.

Note: Missing apartment characteristics, including the type of agent, cause the total sample to be smaller than 1,000 observations. Group A, which includes married tenants, is the reference group. Robust standard errors corrected for day of inquiry are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3 confirms the insights provided by the descriptive statistics. These estimates show no evidence of disability discrimination by housing brokers; in fact, $\hat{\beta}_1$ is close to zero and never statistically significant. However, households with a pet dog experience discriminatory behaviors; $\hat{\beta}_2$ is highly statistically significant and negative. The difference between $\hat{\beta}_1$ and $\hat{\beta}_2$ is also statistically significant. The model augmented with control variables and fixed-effects gives equivalent estimates.²⁷ The combination of these results suggests that housing brokers treat household tenants where the wife is blind and owns a guide dog similarly to the reference group (i.e., household tenants with no dog). Table 3 also provides evidence of disability discrimination by apartment owners. Given that $\hat{\beta}_1$ is negative and highly statistically significant, blind tenants are discriminated against; their invitation rate to visit an apartment is 20 percentage points lower than that of household tenants with no dog. Moreover, household tenants with a pet dog have an invitation rate that is 18.5 percentage points lower than that of the reference group; $\hat{\beta}_2$ is highly statistically significant. The difference between the estimates for $\hat{\beta}_1$ and $\hat{\beta}_2$ is not statistically significant. The model augmented with control variables and fixed-effects finds $\hat{\beta}_1$ and $\hat{\beta}_2$ to be both larger and even closer to each other. All of the estimates in the table are robust to different specifications.²⁸ Heterogeneity analyses suggest the absence of any statistically significant difference

²⁷ The estimates of the constants in columns (1.C) and (2.C) depend, though, on the reference region, which is Lombardy. The Appendix explores the possibility that discrimination varies by geographic dimensions.

²⁸ I have implemented two alternative linear models. In one model, I introduce a partial interaction between *Blind_i* and *Dog_i* without including the main effect for *Blind_i*. The results are equivalent. In this alternative linear model specification, the variable *Dog_i* equals 1 regardless of the nature of the wife's dog that is, without distinguishing between pet dogs and guide dogs. In the other model, I pool the observations on housing brokers and apartment own-

in terms of discriminatory behaviors between metropolitan and non-metropolitan cities or between different Italian macro-regions, i.e., north, central and south.²⁹ See the results in the Appendix.

If the reader accepted the equality assumption, the direction and size of the estimates suggest that discrimination against blind tenants occurs because of the presence of their guide dogs alone. According to EU and Italian legislation, this behavior qualifies as discrimination.

4 Discussion and Conclusions

In this paper, I have analyzed discrimination against blind tenants who are assisted by guide dogs in the Italian rental housing market. I have found evidence that blind tenants are discriminated against by apartment owners (i.e., landlords who advertise and rent out their own apartments) because of their guide dogs alone despite the legislation. According to both EU and Italian legislation, this behavior can be referred to as *indirect discrimination* because it is based on a tenant's characteristic that is indirectly related to the disability. These legislations define indirect discrimination as occurring when an apparently neutral requirement that is assumed to apply to everyone (e.g., pet restrictions) has an unfair effect on disabled people (e.g., blind tenants) and is also unreasonable considering the circumstances of the case.

These results are in line with those from American studies which investigate disability discrimination and reasonable accommodation for guide dogs. Within that context, the refusal to provide reasonable accommodation could be interpreted as indirect discrimination.

ers and introduce the interaction between the variable *Company_i*, which equals 1 if the advertiser is a housing broker, and *Blind_i* as well as *Dog_i*. The results are equivalent.

²⁹ Northern regions are: Emilia-Romagna, Friuli-Venezia-Giulia, Liguria, Lombardy, Piedmont, Trentino-South Tyrol, Aosta Valley, Veneto. Regions from central Italy are: Lazio, The Marche, Umbria, Tuscany. Southern regions are: Abruzzo, Apulia, Basilicata, Calabria, Campania, Molise, Sardinia, Sicily. Italy's regions are divided into these three macro-regions following the Italian national institute (i.e., ISTAT).

Awareness and information campaigns should be tailored to apartment owners to decrease discrimination. Awareness campaigns could focus on the role of guide dogs, conveying the message that they not only provide emotional support to their handlers, as all other dogs do, but that they also assist them in multiple activities throughout the day. For instance, guide dogs identify, and help to avoid, obstacles that the handler would not identify alone; guide dogs help the handler to board public transportation, and they also help their handlers proceed safely along roads and cross them. These campaigns could also illustrate the particular characteristics of guide dogs: they are well trained and are selected from particularly obedient breeds. These awareness campaigns could potentially favorably shape the perception of those apartment owners who know the law but choose to ignore it. These awareness campaigns centered on the importance of guide dogs would be more relevant if the equality assumption illustrated in Section 2.2 were fully accepted. Differently, information campaigns should be conducted to educate apartment owners about the legislation; some of them might still not be fully aware of what is considered illegal discrimination. Information campaigns would be useful even if indirect discrimination were not the (only) cause of discrimination.

These results call for future research to further explore disability discrimination in the rental housing market. For example, a correspondence test might be adapted to investigate discrimination against tenants with mobility impairments, such as tenants who need wheelchairs, walkers, or rollators. Given that mobility declines with age (Rantakokko et al., 2013), this research could add important knowledge also on discrimination against elderly tenants.

Bibliography

Ahmed, A. M., & Hammarstedt M. (2008). Discrimination in the rental housing market: A field experiment on the Internet. *Journal of Urban Economics*, 64(2), 362-90.372

Ahmed, A. M. & Hammarstedt, M. (2009). Detecting Discrimination against Homosexuals: Evidence from a Field Experiment on the Internet. *Economica*, 76(303), 588-597.

Baldini, M. & Federici, M. (2011). Ethnic Discrimination in the Rental Housing Market *Journal of Housing Economics*, 20(1), 1-14

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.

Carlsson, M. & Eriksson, S. (2014) Discrimination in the rental market for apartments. *Journal of Housing Economics*, 23, 41-54.

Carlsson, M. & Rooth, D.-O. (2007). Evidence of ethnic discrimination in the Swedish labor market using experimental data. *Labour Economics*. 14(4), 716-729.

Eurispes (2013). Rapporto Italia 2013.

Federazione Italiana Agenti d'Affari (2006) L'Agente Immobiliare nel Contesto Europeo.

Heckman, J.J. & Siegelman P. (1993). The Urban Institute audit studies: Their methods and findings. In: Fix M., Struyk R. (eds.) Clear and convincing evidence: Measurement of discrimination in America, 187-258. Washington DC: The Urban Institute.

Heckman, J. J. (1998). Detecting Discrimination. *The Journal of Economic Perspectives*. 12(2), 101-116.

Jones, M. K. (2008). Disability and the labour market: a review of the empirical evidence *Journal of Economic Studies*. 35(5), 405-424.

Maberley, D. A. L., Hollands, H., Chuo, J., Tam, G., Konkak, J., Roesch, M., Veselinovic, A., Witzigmann, M. & Bassett, K. (2006). The prevalence of low vision and blindness in Canada *Eye*. 20, 341-346.

Murphy, T. P. (2007). *Disability Discrimination Audit of the Housing Market of Newton, Massachusetts*. Boston: Disability Law Center, Inc.

Pager, D. (2007). 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.

Pascolini, D. & Mariotti, P. S. (2012). Global estimates of visual impairment: 2010. *British Journal of Ophthalmology*, 96(5), 614-618.

Rantakokko, M., Mänty, M., & Rantanen, T. (2013). Mobility Decline in Old Age. *Exercise Sport Science Review*, 41(1), 19-25.

Riach, P. A., & Rich, J. (2002). Field Experiments of Discrimination in the Market Place. *The economic journal*, 112(483), F480-F518.

Turner, M. A., Herbig, C., Kaye, D., Fenderson, J., & Levy, D. (2005). *Discrimination against Persons with Disabilities: Barriers at Every Step*. Washington D.C.: HUD USER, Department of Housing and Urban Development, 39105.

Appendix

This appendix includes heterogeneity analyses based on different geographic dimensions. Discriminatory behaviors might differ between metropolitan and non-metropolitan cities, and also between different Italian macro-regions, i.e., north, central and south, for a wide range of reasons.

Overall, the results are not statistically significant and should be considered carefully. Whether a town is included within a metropolis is somewhat arbitrary; a similar consideration should be made during the interpretation of the results from the analyses by macro-regions.

Metropolitan and non-metropolitan cities

The behavior of landlords in metropolitan cities might differ from that of landlords in non-metropolitan cities for a number of reasons. For instance, apartments in metropolitan cities might have fewer outdoor spaces where pets can be kept, so that pets might have to remain indoors for longer and be considered more of a burden from the landlords' point of view; as a consequence, online advertisers might treat tenants worse whether they have a pet dog or a guide dog. As another example, apartment owners in metropolitan cities might conduct more frequent housing transactions, and thus they might discriminate less against blind tenants with guide dogs.

In this additional analysis, interactions of the variables $Blind_i$ and Dog_i with the dummy variable for the apartment being in a metropolitan city are added. The resulting model is the following:

$$Invitation_rate_i = \beta_0 + \beta_1 Blind_i + \beta_2 Dog_i + \beta_3 Metro_i + \\ \beta_4 Blind_i * Metro_i + \beta_5 Dog_i * Metro_i + \beta X_i + \beta F_i + \varepsilon_i$$

The estimates are provided below, in Table A.1.

Table A.1 Linear probability model estimates, interaction with metropolitan city.

	Housing brokers			Apartment owners		
	(1.A)	(1.B)	(1.C)	(2.A)	(2.B)	(2.C)
Constant	0.825*** (0.033)	0.576*** (0.170)	0.614*** (0.189)	0.795*** (0.034)	0.965*** (0.100)	0.921*** (0.102)
$\hat{\beta}_1$ (Blind, guide dog)	-0.018 (0.050)	-0.027 (0.050)	-0.034 (0.055)	-0.239*** (0.067)	-0.256*** (0.067)	-0.266*** (0.070)
$\hat{\beta}_2$ (Not blind, pet dog)	-0.107** (0.049)	-0.117** (0.048)	-0.112** (0.050)	-0.191*** (0.039)	-0.223*** (0.042)	-0.243*** (0.040)
$\hat{\beta}_3$ (Metro)	-0.075 (0.135)	-0.141 (0.131)	-0.161 (0.135)	-0.195** (0.090)	-0.165* (0.094)	-0.141 (0.103)
$\hat{\beta}_4$ (Blind, guide dog)×(Metro)	0.101 (0.147)	0.135 (0.146)	0.169 (0.148)	0.187 (0.138)	0.187 (0.148)	0.169 (0.150)
$\hat{\beta}_5$ (Not blind, pet dog)×(Metro)	0.007 (0.186)	0.011 (0.180)	0.028 (0.184)	0.062 (0.138)	0.071 (0.148)	0.056 (0.151)
Control variables	N	Y	Y	N	Y	Y
Fixed-effects	N	N	Y	N	N	Y
R-squared	0.017	0.033	0.076	0.049	0.070	0.130
N	419	409	409	565	526	526

Note: Missing apartment characteristics, including the type of agent, cause the total sample to be smaller than 1,000 observations. Group A, which comprises married tenants, is the reference group. Robust standard errors corrected for day of inquiry are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

The estimates do not provide statistically significant evidence of an interaction effect between the apartment being in a metropolitan city and having either a pet or a guide dog. However, the results suggest that blind tenants with a guide dog are slightly favored by housing brokers in metropolitan cities. Moreover, blind tenants also appear to experience less discrimination by apartment owners in metropolitan cities because the interaction term is positive. There are no similar differences between metropolitan and non-metropolitan cities for tenants with pet dogs because

the interaction term is very close to zero for these tenants and for both groups of advertisers and for all of the specifications.

These results should be considered carefully because it is to some extent arbitrary whether a city is labeled metropolitan.

Northern, central and southern regions

The behavior of landlords in different macro-regions might also differ for multiple reasons. For instance, attitudes toward pets might differ on a geographic basis.

In this additional analysis, interactions of the variables $Blind_i$ and Dog_i with dummy variables for the apartment being in northern or central Italy are added; the reference macro-region is southern Italy. In this model specification, dummy variables for individual regions are not included in the fixed-effect vector. The model is thus the following:

$$Invitation_rate_i = \beta_0 + \beta_1 Blind_i + \beta_2 Dog_i + \beta_3 North_i + \beta_4 Center_i + \beta_5 Blind_i * North_i + \beta_6 Blind_i * Center_i + \beta_7 Dog_i * North_i + \beta_8 Dog_i * Center_i + \beta X_i + \beta F_i + \varepsilon_i$$

The estimates are provided below, in Table A.2.

Table A.2 Linear probability model estimates, interactions with macro-regions.

	Housing brokers			Apartment owners		
	(1.A)	(1.B)	(1.C)	(2.A)	(2.B)	(2.C)
Constant	0.820*** (0.0538)	0.634*** (0.180)	0.632*** (0.189)	0.712*** (0.056)	0.934*** (0.101)	0.886*** (0.098)
$\hat{\beta}_1$ (Blind, guide dog)	-0.042 (0.075)	-0.068 (0.074)	-0.072 (0.073)	-0.204** (0.083)	-0.242*** (0.083)	-0.247*** (0.084)
$\hat{\beta}_2$ (Not blind, pet dog)	-0.139 (0.0853)	-0.167** (0.080)	-0.173** (0.080)	-0.212*** (0.077)	-0.251*** (0.077)	-0.259*** (0.077)
$\hat{\beta}_3$ (North)	-0.046 (0.078)	-0.091 (0.075)	-0.087 (0.076)	0.095 (0.073)	0.069 (0.067)	0.065 (0.066)
$\hat{\beta}_4$ (Center)	0.103 (0.074)	0.049 (0.073)	0.046 (0.075)	0.045 (0.081)	0.017 (0.075)	0.015 (0.075)
$\hat{\beta}_5$ (Blind, guide dog)×(North)	0.061 (0.104)	0.089 (0.103)	0.090 (0.102)	-0.041 (0.109)	-0.015 (0.098)	-0.011 (0.096)
$\hat{\beta}_6$ (Blind, guide dog)×(Center)	0.019 (0.125)	0.070 (0.131)	0.078 (0.129)	0.069 (0.137)	0.129 (0.135)	0.131 (0.133)
$\hat{\beta}_7$ (Not blind, pet dog)×(North)	0.090 (0.119)	0.116 (0.113)	0.118 (0.113)	0.022 (0.104)	0.030 (0.104)	0.036 (0.103)
$\hat{\beta}_8$ (Not blind, pet dog)×(Center)	-0.065 (0.129)	-0.028 (0.131)	-0.026 (0.130)	0.092 (0.115)	0.125 (0.119)	0.134 (0.119)
Control variables	N	Y	Y	N	Y	Y
Fixed-effects	N	N	Y	N	N	Y
R-squared	0.026	0.042	0.044	0.048	0.075	0.079
N	419	409	409	565	526	526

Note: Missing apartment characteristics, including the type of agent, cause the total sample to be smaller than 1,000 observations. Group A, which comprises married tenants, is the reference group. Robust standard errors corrected for day of inquiry are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

The estimates do not provide statistically significant evidence that discrimination against tenants with pet vs. guide dogs differ by macro-region. However, they suggest that blind tenants could be favored by housing brokers in the northern and central regions. Furthermore, apartment owners

appear to discriminate against blind tenants independently of the macro-region where the apartment is located.

These results should be considered carefully because the division of Italian regions into macro-regions may be assessed differently and it is somewhat arbitrary (e.g., in some studies ISTAT divides regions into sub-macro-regions, such as north-eastern and north-western or southern and insular regions).

II

ORIGINAL ARTICLE

Open Access

Does the design of correspondence studies influence the measurement of discrimination?

Magnus Carlsson^{1*}, Luca Fumaco^{1†} and Dan-Olof Rooth^{1,2,3†}

* Correspondence:
Magnus.Carlsson@lnu.se
† Equal contributors
¹Centre for Labour Market and
Discrimination Studies, Linnaeus
University, Kalmar, Växjö, Sweden
Full list of author information is
available at the end of the article

Abstract

Correspondence studies can identify the extent of discrimination in hiring as typically defined by the law, which includes discrimination against ethnic minorities and females. However, as Heckman and Siegelman (1993) show, if employers act upon a group difference in the variance of unobserved variables, this measure of discrimination may not be very informative. This issue has essentially been ignored in the empirical literature until the recent methodological development by Neumark (2012). We apply Neumark's method to a number of already published correspondence studies. We find the Heckman and Siegelman critique relevant for empirical work and give suggestions on how future correspondence studies may address this critique.

JEL classification: J71

Keywords: Correspondence studies; Discrimination

1. Introduction

Correspondence studies are an increasingly popular method for measuring discriminatory treatment against, e.g., ethnic minority and female workers in the labor market (see Riach and Rich, 2002, for a survey). In the standard correspondence study, matched pairs of qualitatively identical job applications are sent to employers that have advertised a job opening. The only difference between the fictitious applications is the name of the applicant, which signals ethnicity or gender. The degree of discrimination in hiring is quantified by calculating the difference in the callback rate (i.e., the fraction of invitations) to a job interview between the groups.

The advocates of correspondence studies argue that the method provides the most clear and convincing evidence of discrimination. Their main argument is that a carefully designed correspondence study can identify discriminatory treatment by employers since the signal of group belonging is randomized. This circumvents the problem with unobserved individual heterogeneity – a common problem in studies using administrative data.

The method's ability to identify discriminatory treatment by employers is certainly attractive, but it should be noted that correspondence studies cannot distinguish between preference-based (Becker, 1957) and statistical discrimination (Aigner and Cain, 1977; Arrow, 1973; Phelps, 1972). Somewhat simplified, preference-based discrimination is based on employer prejudice, while statistical discrimination arise when employers act

upon perceived group differences in the mean or variance of unobserved variables (i.e., variables not included in the job applications). However, the inability to separate between these two types of discrimination may not be a huge drawback unless the aim is solely to identify preference-based discrimination. In many countries, both preference-based and statistical discrimination against, e.g., ethnic minorities and women are illegal¹. Hence, the level of discrimination being measured by the standard correspondence study is an unbiased measure of the degree of discrimination as defined by the law for these countries.

More problematic is that, despite being an unbiased measure of what the law defines as discrimination, which includes the case when employers act upon perceived group differences in the variance of unobserved variables, it may not be very informative. The problem is that when employers perceive a group difference in the variance of unobserved variables, the degree of discrimination in a correspondence study depends on the level at which the experimenter standardizes the observed characteristics in the job applications². As a result, the standard correspondence study only tells the true level of discrimination against ethnic minority or female applicants who have similar qualifications as in the fictitious job applications. In the extreme case, a badly designed correspondence study may measure the degree of discrimination against a very atypical, or even non-existing, ethnic minority or female job applicant. In order to obtain an informative measure of discrimination in the market, the level of standardization must reflect the qualifications of a representative ethnic minority or female job applicant.

Although the idea of perceived group differences in the variance of unobservables has a long tradition in economics (e.g., Aigner and Cain, 1977), the issue has been essentially ignored in the empirical literature on correspondence studies until the appearance of the method proposed by Neumark (2012)³. In short, Neumark's method implies estimating the perceived relative variance in unobserved variables across groups, which then makes it possible to decompose the measured degree of discrimination into two parts. The first part captures discrimination in hiring due to employer preferences and/or a perceived group difference in the mean of unobserved variables, while the second part captures discrimination in hiring due to a perceived group difference in the variance of unobserved variables. In our study, the second part is of main interest, since it reveals to what extent the result of a particular correspondence study is affected by its design, i.e., the level of standardization of the qualifications included in the job applications. Neumark applies his method to the data in the seminal correspondence study conducted by Bertrand and Mullainathan (2004) and finds indicative evidence of that the degree of discrimination depends on perceived group differences in the variance of unobserved variables. Baert et al. (2013) also applies Neumark's method, but to their own data, and find a similar result.

In the current study, we use Neumark's method to analyze to what extent a perceived group difference in the variance of unobserved variables is an issue in a number of already published correspondence studies. To this end, we use data from three experiments conducted in the Swedish labor market between 2005 and 2007. In two of the experiments, our results indicate that the degree of discrimination depends on perceived group differences in the variance of unobservables, while in one experiment there is no evidence of a dependency⁴.

The next section explains the issue with perceived group differences in the variance of unobserved variables and the level of standardization (henceforth *the HS critique*, since it was first discussed in the seminal paper by Heckman and Siegelman, 1993). Section 3 explains Neumark's method in more detail, Section 4 describes the correspondence studies used to implement Neumark's method, Section 5 presents the main results, and Section 6 concludes.

2. The HS critique

This section aims at graphically explaining the intuition behind the HS critique. We first explain how an employer estimates the productivity of an applicant, and how an applicant's probability of being invited to a job interview is determined. Then we turn to the factors that determine the measured degree of discrimination in a correspondence study, where we focus on the level of standardization of the job applications. Much of the content of this section is inspired by Heckman and Siegelman (1993, henceforth HS)⁵. For readers that are interested in a more formal and detailed explanation we refer to HS's paper.

2.1 The productivity of a job applicant

To determine if a job applicant should be invited to a job interview, the employer estimates the productivity of the applicant. The productivity depends on deterministic observed variables X^{OBS} (variables that are included in the job application), unobserved random variables X^{LNOBS} (variables that are not included in the job application), and a discount factor γ that reflects employer preferences, which takes a negative value for applicants in the discriminated group and zero otherwise. Total productivity P for an applicant is then given by

$$P = \beta^{OBS} X^{OBS} + X^{LNOBS} + \gamma \quad (1)$$

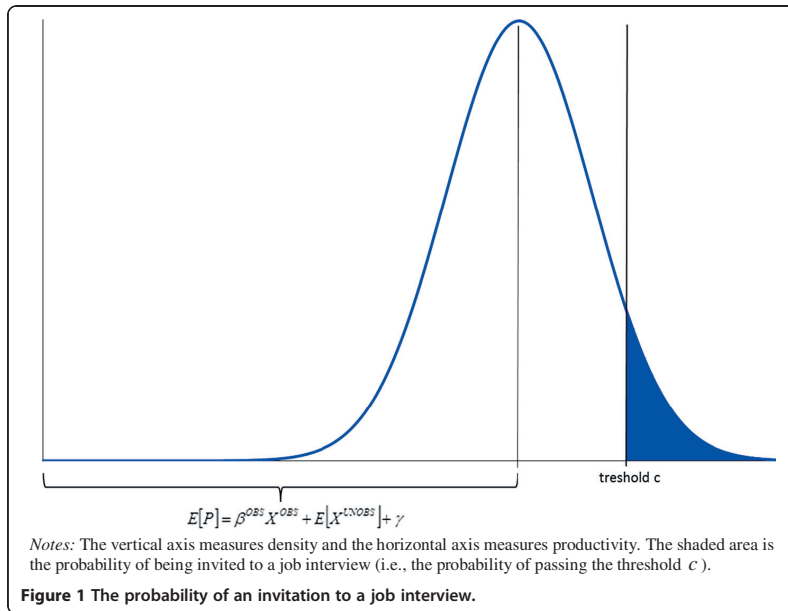
where β^{OBS} is the return to observed characteristics and the return to unobserved characteristics has been normalized to one.

2.2 The probability of a job interview

Job applicants that have productivity above a certain threshold are invited to a job interview. Since the productivity of a particular job applicant is unknown, the employer estimates the likelihood that the productivity of a job applicant passes the threshold; the only random factor in the estimation is X^{LNOBS} . If X^{LNOBS} follows a normal distribution, total productivity P is also normally distributed⁶. The mean of P is $E[P] = \beta^{OBS} X^{OBS} + E[X^{LNOBS}] + \gamma$, which depends on the employer's perception about the mean of X^{LNOBS} , and the variance of P is determined by the employer's perception about the variance of X^{LNOBS} . Both the mean and variance of P may vary between groups. In Figure 1, the probability of being invited to a job interview is graphically illustrated. The shaded area is the probability of being invited to a job interview, which corresponds to the probability of passing the threshold c ⁷.

2.3 Discrimination

The measure of discrimination in a correspondence study reflects the situation where there are two groups of job applicants with identical observed characteristics X^{OBS} , but



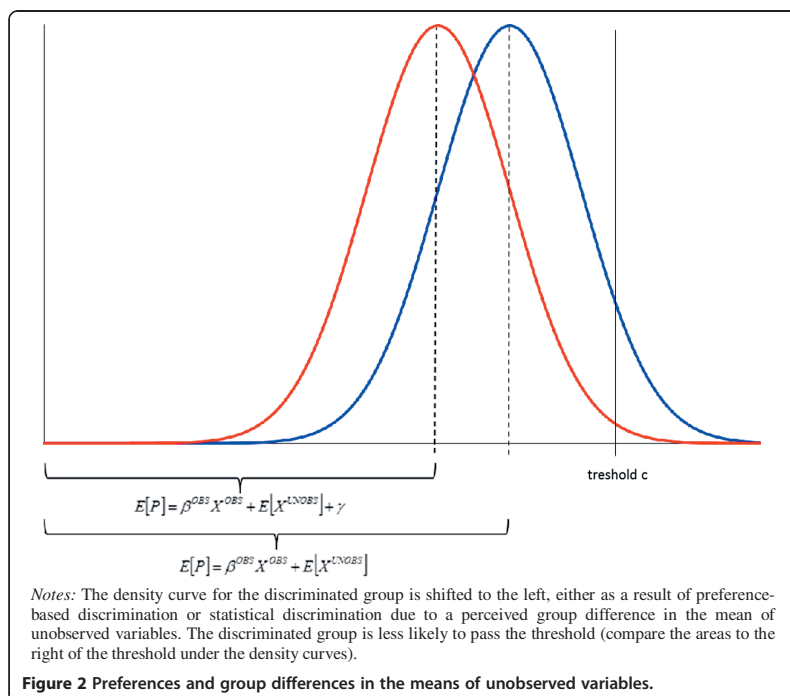
the likelihood of being invited to a job interview is higher for one group of applicants. This measure of discrimination can be decomposed into two parts. The first part, which we label *the effect through the level of discrimination*, reflects the combined effect on discrimination through preference-based discrimination and/or a perceived difference in the mean of unobserved variables. The second part, which we label *the effect through the variance of unobserved variables*, reflects the effect on discrimination through a perceived group difference in the variance of unobserved variables. The second part depends on the level of standardization, which we return to below.

The effect through the level of discrimination

Figure 2 illustrates the effect through the level of discrimination, i.e., a situation when there is no perceived group difference in the variance of unobserved variables. In this case, the expected productivity $E[P]$ is lower for applicants in the discriminated group for whom the density curve is shifted to the left. As a result, there is a lower probability of passing the threshold for these applicants. Note that $E[P]$ may be lower as a result of either preference-based discrimination ($\gamma < 0$) or statistical discrimination based on a perceived difference in the mean of unobserved variables ($E[X^{UNOBS}]$). Hence, correspondence studies cannot distinguish between preference-based discrimination and statistical discrimination based on a perceived difference in the mean of unobserved variables.

The effect through the variance

A more problematic case, which is the focus of this paper, is statistical discrimination due to perceived group differences in the variance of unobserved variables. This type of discrimination is problematic, since its magnitude depends on the level of standardization

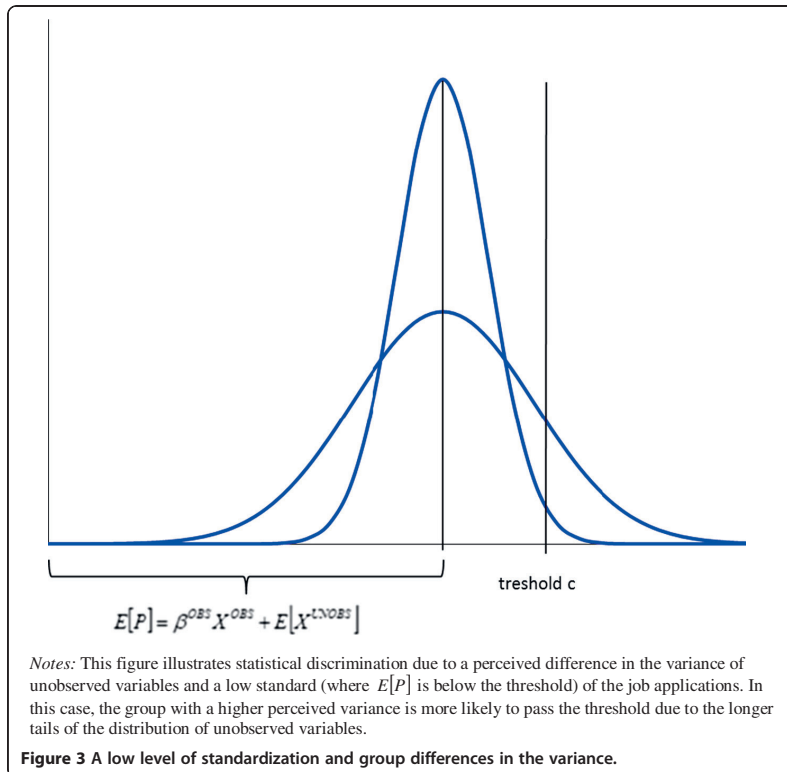


of the job applications. To provide the intuition to this issue, let us focus on the simplest case where there are no perceived group differences in the mean of unobserved variables and no preference-based discrimination (i.e., $\gamma = 0$).

Figure 3 illustrates a situation where the experimenter has set a low standard of the job applications. In this case, the expected productivity $E[P]$ is below the threshold for both groups of applicants. However, if one group of applicants has a higher variance of unobserved variables, then job applicants from this group are more likely to pass the threshold due to the longer tails of the distribution of unobserved variables.

Figure 4 illustrates the opposite case where the experimenter has set a high standard of the job applications. In this case, the expected productivity $E[P]$ is above the threshold for both groups of applicants. However, in this scenario the group with the higher variance of unobserved variables will now be less likely to pass the threshold due to the longer tails of the distribution of unobserved variables.

The cases illustrated in Figures 3 and 4 give the theoretical argument to why the results from a standard correspondence study may not be very informative about the degree of discrimination in the market. The cases show that the measured degree of discrimination depends on the level of standardization of the job applications if there is a perceived group difference in the variance of unobserved variables. Hence, if the level of the qualifications is not set to mirror a representative ethnic minority or female job applicant, the measured degree of discrimination may say little about the average degree of discrimination in the market⁸.

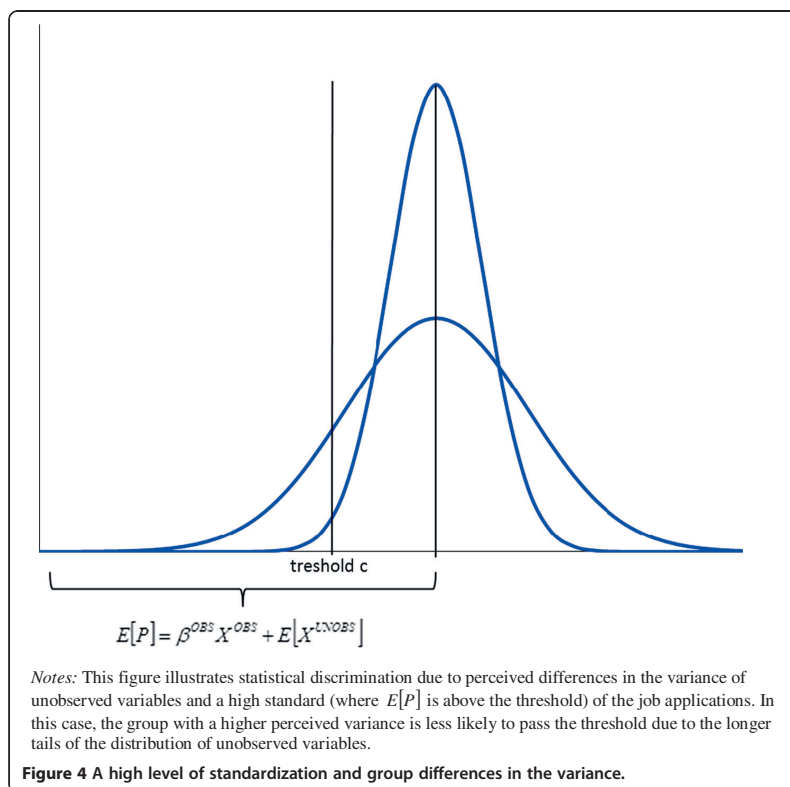


3. Neumark's method⁹

Neumark's insight is that the HS critique can be addressed in a two step procedure. In the first step, the degree of discrimination is estimated together with the group specific variance of unobserved variables. In the second step, the estimated degree of discrimination is decomposed into two parts: the effect of group belonging *through the level of discrimination* (see Figure 2) and the effect of group belonging *through the variance of unobserved variables* (see Figures 3 and 4). In the analysis, we have followed Neumark's two step procedure, which we have implemented using Stata 12¹⁰.

In the first step, Neumark uses the heteroskedastic probit model for estimation. Identification of the group specific variance in the heteroskedastic probit model requires data from a correspondence study that have random variation not only in the signal of group belonging, but also in some other observed productivity related variable(s) in the job applications. Importantly, there is an identifying assumption in the heteroskedastic probit model, which translates into an assumption of equal returns across groups to these additional productivity related variables. Below we return to whether this assumption is likely to hold.

In the second step, Neumark decomposes the marginal effect of group belonging in the heteroskedastic probit model into the two parts: the effect of group belonging *through the level of discrimination* and the effect of group belonging *through the*



variance of unobserved variables. The standard errors of the two parts are calculated using the delta method.

Returning to the identifying assumption of equal returns, this assumption is likely to hold for well designed correspondence studies, where there should be no group differences in the quality of the observed characteristics. E.g., in a written application the experimenter can easily choose not only the amount of schooling and work experience, but also similar schools and type of work experience so that the returns to those characteristics are the same across groups.

Moreover, the identifying assumption about equal coefficients can be tested. To implement the test, the first step is to estimate the probability of an invitation to a job interview separately for the two groups. In the second step, the residual standard deviations are normalized such that for one group the standard deviation is equal to unity while for the other group it is equal to the ratio of the group residual standard deviations. In the simplest case, with only one observed productivity related variable being varied in the job applications, a group difference in the coefficient of this variable can either arise because the identifying assumption does not hold or because the relative standard deviation is different from unity. However, with (at least) two observed productivity related variables that vary in the job applications, it becomes possible to test the null hypothesis of equal coefficients across groups of the observed applicant

characteristics. In the third step of the test, the ratios of the two coefficients are calculated separately for each group of applicants. It is the fact that the relative standard deviation cancels out for the second group, since this is a factor in both the dominator and numerator, which enables the test. In the final step, the null hypothesis of equal coefficients is tested by testing if the two ratios are equal across the groups¹¹. We apply this test in our empirical analysis.

4. Data

To implement Neumark's method we use data from three different correspondence studies conducted in the Swedish labor market, which investigate both ethnic and gender discrimination, and have random variation in applicant characteristics. These three data sets are labeled Experiment A, B, and C. Recall that Neumark's method requires observed applicant characteristics that have a significant effect on the probability of an invitation and that the effect is the same across groups. Since the set of variables that fulfills this requirement vary across the experiments, we use a different set of observed characteristics for each experiment.

4.1 Experiment A

In Experiment A, focus is on ethnic discrimination against applicants with Middle Eastern sounding names and the data was collected in a field experiment conducted between March and November 2007¹². This field experiment was designed for analyzing a number of research questions related to individual worker productivity and therefore has a large variation in productivity characteristics of the fictitious job applications. In principle, twelve different variables were randomly assigned to each application. However, not all of them were found to have an effect on the probability of a job interview or to have the same return across groups. In the end, we include five variables that fulfill these requirements in the analysis of the variance of unobservables, while the other variables are excluded from the regressions.

The first two variables regard the personality of the candidate, basically following the Big Five taxonomy using the two of its five categories - agreeableness and extroversion (see Borghans et al. 2008). Being an *agreeable* person has both a moral and social dimension. An agreeable applicant states that it is important to care about others and likes to work in a group. In contrast, an applicant that is not agreeable does not emphasize these qualities¹³. Considering the category *extrovert*, it was decided in the design of the experiment to focus on the lower level category *competence*. A competent applicant states that he or she is a hardworking person that puts a lot of effort on the job. In contrast, an applicant that is not competent does not emphasize these qualities¹⁴. Both these variables are coded as dummy variables in the empirical analysis.

The third variable captures the type of neighborhood the applicant lives, with a dummy variable that indicates if the applicant lives in a high income area (i.e., mean income in the area is above the average). The fourth variable gives the applicant's previous work experience), which varies between one and five years. In the empirical analysis, this variable is coded with dummies for each year of experience and with one year serving as the benchmark. Finally, the fifth variable measures whether the applicant is engaged in sport activities or not. Sport activities could be exercised at two

different effort levels: a recreational and a competitive level, and this variable is included as a dummy for each level of sport activity.

During Experiment A all employment advertisements in thirteen selected occupations found on the webpage of the Swedish employment agency were collected¹⁵. For these advertised jobs, 5,657 applications – 2,837 with a typical native Swedish sounding name and 2,820 with a typical Middle Eastern sounding name – were sent to 3,325 employers. Different job applications were used in each occupation in order to match the specific skills that are important in an occupation (this holds for Experiment B and C as well). All applications were sent by email; a clear majority of employers posting vacant jobs at this site accept applications by email. Jobs were applied to all over Sweden, but most advertisements were found in the two major cities of Sweden: Stockholm and Gothenburg. Callbacks for interview were received via telephone (voice mailbox) or e-mail.

4.2 Experiment B

Experiment B considers discrimination against female names. Within the same project as Experiment A, it is also possible to analyze gender discrimination, since additionally 2,830 applications with the same design but now with a native female name were sent to employers in the same occupations. Compared to Experiment A, we find much fewer individual variables that affect the probability of a job interview and which also have the same return for both men and women. However, there are variables that have a joint effect that fulfill these requirements. To this end, we construct two new combined variables based on the individual variables. We label these new variables *good labor market characteristics* and *good personal characteristics*, and both are simply two indicators. An applicant is defined as having good labor market characteristics if he or she has at least one of the following characteristics: the person has been abroad for one year during high school; the person has at least four years of experience; the person has experience from more than one previous employer; the person has employment at the moment. An applicant with good personal characteristics is defined as an individual that has at least one of the following characteristics: the person is extrovert or the person is agreeable. All other explanatory variables are excluded from the regressions.

4.3 Experiment C

Also Experiment C considers ethnic discrimination against applicants with Middle Eastern sounding names. Actually, Experiment C consists of observations from two different correspondence studies¹⁶. What justifies viewing them as a single experiment is that both studies have the same design and are conducted roughly during the same time period between 2005 and 2007. In both experiments, the job applicants are born in Sweden, have either a typical native Swedish or Middle Eastern sounding name, are on average 25–30 years old, have two to four years of work experience in the relevant occupation, and have obtained their education in the same type of school. Also, in both studies, the applications consist of a quite general biography on the first page and a detailed CV of education and work experience on the second page. Finally, in both studies a similar routine for receiving responses from the employers were used: email addresses and a telephone numbers (including an automatic answering service) were registered at a large Internet provider and a phone company.

Despite the similarities between these two studies there is one important distinguishing factor. For reasons unrelated to this paper, the applications in the second experiment were calibrated for six of the occupations relative to the characteristics in the first experiment; the quality of the applications in terms of labor market experience and skills were raised in three occupations and lowered in the other three occupations¹⁷. These six occupations contain 3,536 observations. This calibration generates the variation in the standards of the applications that we utilize in the current paper. However, since only one variable varies in Experiment C, we are not able to test the identifying assumption of equal returns to characteristics in this case.

5. Empirical analysis

In this section, we first use the standard probit model to provide a set of basic results for Experiment A-C. Then we turn to the main analysis, where we apply Neumark's method.

5.1 Basic results

Table 1 presents basic results for Experiment A-C. The purpose of this table is to report the estimate of the degree of discrimination that comes out of a standard correspondence study and to show that the observed variables in the job applications that we use to implement Neumark's method have significant effects – with expected signs – on the probability of receiving an invitation to a job interview¹⁸.

Experiment A

The first two columns of Table 1 report the basic results for Experiment A. In the top row of the first column, we find that the ethnic difference in the probability of a job interview is 9.4 percentage points. From the following seven rows of this column it is evident that applicants that are extrovert, agreeable, live in a high income area, or have more than one year of experience (the benchmark) have significantly higher probability of receiving an invitation to a job interview. Also, the next two rows in this column show that applicants that are engaged in sport activities have a (weakly significant) higher probability of an invitation to a job interview. This means that essentially all the observed application variables have a significant effect – with the expected sign – on the probability of a job interview. While the regression underlying the estimates in the first column does not include any other control variables, the second column includes all application attributes and occupational fixed effects. These additional controls include dummy indicators for whether the vacancy was located in Stockholm, Gothenburg, or in other parts of Sweden, the order the applications were sent, and the typeface and layout of the application.

Experiment B

The basic results for Experiment B are found in the third and fourth columns of Table 1. Again, in the top row we find the group difference in the probability of a job interview, now between male and female applicants, which is 2.6 percentage points in favor of female applicants. From the estimates further down in the table it is evident that applicants that have good labor market and personal characteristics have a significantly higher probability of an invitation to a job interview.

Table 1 Basic probit results

	Ethnicity		Gender		Ethnicity
	Experiment A		Experiment B		Experiment C
	(1)	(2)	(3)	(4)	(5)
Middle Eastern sounding/female Name	-.094*** [.009]	-.096*** [.009]	.026*** [.009]	.026** [.009]	-.128*** [.010]
Application characteristics					
<i>Extroversion/competence</i>	.03*** [.01]***	.04*** [.01]	-	-	-
<i>Agreeableness</i>	.03** [.01]	.02** [.01]	-	-	-
<i>Experience = 2</i>	.02 [.02]	.03 [.02]	-	-	-
<i>Experience = 3</i>	.06*** [.02]	.06*** [.02]	-	-	-
<i>Experience = 4</i>	.08*** [.02]	.08*** [.021]	-	-	-
<i>Experience = 5</i>	.03 [.02]	.04** [.02]	-	-	-
<i>High income area</i>	.02* [.01]	.02 [.01]	-	-	-
<i>Recreational sports</i>	.02 [.01]	.01 [.01]	-	-	-
<i>Competitive sports</i>	.03* [.02]	.03 [.02]	-	-	-
<i>Good labor market characteristics</i>	-	-	.03*** [.01]	.03*** [.01]	-
<i>Good personal characteristics</i>	-	-	.04*** [.01]	.04*** [.01]	-
<i>Increased quality application</i>	-	-	-	-	.04*** [.02]
Other application controls	No	Yes	No	Yes	-
Occupational fixed effects	No	Yes	No	Yes	Yes
Number of observations	5,636	5,636	5,662	5,662	3,536

Notes: The dependent variable is an indicator variable that takes the value one if the applicant was invited to a job interview and zero otherwise. The reported standard errors (in brackets) are clustered at the job advertisement level. ***, **, and * denote the 1, 5 and 10 percent significance levels, respectively.

Experiment C

The basic results for Experiment C are found in the last column of Table 1. This time the ethnic difference in the probability of a job interview is 12.8 percentage points, in favor of applicants with native Swedish sounding names. The row at the bottom of the table reveals that improved quality applications have a significantly higher probability of a job interview. Note that there is only one column of estimates for Experiment C. This is partly because in this experiment we do not have any useable information to construct other application controls other than the high quality variable. Moreover, in the case of Experiment C, it does not make sense to present the estimates without occupational fixed effects. The reason is that the quality of the applications are manipulated at the occupational level, which means that without occupational fixed effects the estimate of improved quality may also reflect occupation specific demand.

5.2 Main results

In this section, we use Neumark's method where we first estimate a heteroskedastic probit model, and then decompose the estimated degree of discrimination into two parts: the effect of group belonging through *the level of discrimination* and through *the variance of unobserved variables*. An issue is that Neumark's method often increases the standard errors of the decomposed marginal effects by a factor of 2.5, or more, compared to the standard error of the undecomposed marginal effect, which renders statistically insignificant estimates. Since we find the magnitude of the estimates to be

economically important we choose to view the decomposed marginal effects as still providing evidence.

To facilitate the interpretation of the results of the decomposition, we discuss the results for the first experiment (Experiment A) in detail, while the results of the remaining experiments are discussed more briefly.

Experiment A

The results of the decomposition for Experiment A (ethnic discrimination) are presented in the first column of Table 2. The estimate in panel A is for comparison and repeats the standard probit estimates of having a typical Middle Eastern sounding name in Table 1, while the estimate in panel B presents the estimated marginal effect obtained from the heteroskedastic probit model. If there is a group difference in the variance of unobservables, this would violate the standard probit model assumption of equal variances. Therefore, a first indication of a difference in the variance of unobservables would be if the estimate of the degree of discrimination differs between the standard probit and the heteroskedastic probit. However, in the case of Experiment A, the two estimates are very similar, which indicates that there is no perceived group difference in the variance of unobservables.

Next, the first two rows in panel C give the marginal effects of group belonging decomposed into *the effect through the level of discrimination* and *the effect through the variance of unobserved variables*, respectively. The key estimate of interest is *the effect through the variance*, since this estimate tells if the degree of discrimination depends on perceived group differences in the variance of unobserved variables. In Experiment A, the point estimate of *the effect through the variance* is small and insignificant. This implies that the estimate of discrimination in this experiment does not depend on the design of the experiment, i.e., the level of standardization being set by the experimenter.

Table 2 Decomposition

	Ethnicity Experiment A (1)	Gender Experiment B (2)	Ethnicity Experiment C (3)
<i>A. Basic probit</i>			
Middle Eastern sounding/female name	-.096*** [.009]	.026** [.009]	-.128*** [.010]
<i>B. Heteroskedastic probit</i>			
Middle Eastern sounding/female name	-.098*** [.009]	.029*** [.010]	-.135*** [.011]
<i>C. Decomposition</i>			
Marginal effect of name through level	-.088*** [.023]	.001 [.024]	-.090** [.029]
Marginal effect of name through variance	-.010 [.025]	.028 [.025]	-.044 [.033]
Relative standard deviation of unobserved variables	.96	1.13	.83
Wald test statistic, standard deviation == 1 (p-value)	.68	.29	.14
Wald statistic, ratios of coefficients are equal (p-value)	.67	.89	-
Other application controls	Yes	Yes	-
Occupational fixed effects	Yes	Yes	Yes
Number of observations	5,636	5,662	3,536

Notes: The dependent variable is an indicator variable that takes the value one if the applicant was invited to a job interview and zero otherwise. The reported standard errors (in brackets) are clustered at the job advertisement level. ***, **, and * denote the 1, 5 and 10 percent significance levels, respectively.

The next two rows (in panel C), respectively, present the point estimate of the relative standard deviation of the unobservables for applicants with typical native Swedish and Middle Eastern sounding names and the resulting p-value from testing the hypothesis that the relative standard deviation equals one. There is no evidence of a perceived difference in the variance of unobserved variables, since the relative standard deviation is close to one (.96) and the p-value is large (.68).

The last row in panel C contains the result from testing the identifying assumption of equal coefficients across groups for the observed applicant characteristics. The high p-value for the Wald statistic suggests that the data is consistent with the identifying assumption of equal coefficients.

Experiment B

The results of the decomposition for Experiment B (gender discrimination) are presented in the second column of Table 2. Here, the difference between the two estimates from the standard probit and the heteroskedastic probit is larger – at least in relative terms (compare the estimates in panel A and B). This indicates the existence of a perceived difference in the variance of unobserved variables across groups. Interestingly, when the estimated degree of discrimination is decomposed, we find that *the effect through the variance* is of the same magnitude, although insignificant, as the overall estimate of discrimination (second estimate in panel C), while *the effect through the level* is zero (see first estimate in panel C). This suggests that the measured degree of discrimination in this experiment depends on the level of the standardization of the job applications being set by the experimenter.

Although statistically insignificant, if we take the estimate of the relative standard deviation of unobservables at face value, the interpretation is that the standard deviation of the unobserved variables is 13 percent higher for females compared to males. This is consistent with a low standard (where $E[P]$ is below the threshold, see Figure 3) of the applications being set in the experiment and where the higher variance of unobserved characteristics benefits females.

The high p-value for the Wald statistic in the last row in panel C suggests that the data is consistent with the identifying assumption of equal coefficients for the observed applicant characteristics.

Experiment C

The results of the decomposition for Experiment C (ethnic discrimination) are presented in the third column of Table 2. Here, the difference between the two estimates from the standard probit and the heteroskedastic probit is larger than in both Experiment A and B, which suggests that employers have acted on perceived group differences in the variance of unobservables when hiring. As expected, when the estimated degree of discrimination is decomposed, *the effect through the variance* is quite large, but statistically insignificant. Taking the point estimate at face value suggests that the level of standardization plays a role for the estimated degree of discrimination in this experiment.

Although the relative standard deviation of unobservables in experiment C is not different from one in a statistical sense (the p-value is .14), the interpretation of the point estimate is that the standard deviation of the unobserved variables for

applicants with typical Middle Eastern sounding names is only .83 of the standard deviation for applicants with native Swedish sounding names. Similarly as in Experiment B, this is consistent with setting a low standard (where $E[P]$ is below the threshold, see Figure 3) of the applications in the experiment where applicants with typical Middle Eastern sounding names are suffering from their lower variance of the unobserved variables.

In this experiment, it is not possible to test the identifying assumption of equal returns to observed applicant characteristics, since, in addition to the Middle Eastern sounding name dummy, there is only variation in one explanatory variable.

6. Concluding remark

It can be argued that correspondence studies provide the most clear and convincing evidence of discrimination since the signal of group belonging in these studies is randomized, which circumvents the problem with unobserved individual heterogeneity. However, the results in HS show that the measured degree of discrimination in a correspondence study may still not be very informative if the level of qualifications of the fictitious job applications do not match up with the representative job applicant in the discriminated group. The reason is that when employers act upon perceived group differences in the variance of unobserved variables the degree of discrimination depends on how the experimenter sets the level of qualifications in the job applications. This so called HS critique has essentially been ignored in the empirical literature on correspondence studies until the appearance of the methodology proposed by Neumark (2012).

We use Neumark's method to reexamine a number of already published standard correspondence studies, which do not take into account the level of standardization. We find suggestive evidence that the results of discrimination depend on the level of standardization and hence, perceived group differences in the variance of unobserved variables may be important, not just as a theoretical argument, but also for the empirical design when conducting correspondence studies.

What are the implications of our findings? We believe our results are sufficiently strong to suggest that correspondence studies cannot continue to ignore the issue raised by HS. In our opinion, future correspondence studies should try to invest more effort in the design of the job applications, aiming for a level of standardization that reflects the representative ethnic minority or female job applicant in the population (and ideally for each type of job). This requires information on what qualifications real job applicants have and a challenge may be that, at least historically, such information has been difficult to obtain. However, today there exist large databases with such information as a result of online job search: Job applicants put their CVs online to make them available for employers searching for workers. We believe that a natural way forward would be to use information from such databases when designing the job applications to obtain a truly informative measure of the degree of discrimination in the labor market. Also, we believe that future correspondence studies should be designed to satisfy the requirements necessary to implement Neumark's method, which makes it possible to, at least in retrospect, analyze to what extent the measured degree of discrimination depends on the level of standardization.

Endnotes

¹Under European law, which applies to the member countries of the European Union, of which Sweden is a member, discrimination in employment situations based on, e.g., nationality, race, ethnic origin, and gender is considered a crime. Discrimination under European law includes both preference-based and statistical discrimination in employment situations by covering general situations where “one person is treated less favorably in a comparable situation” (European Union Agency for Fundamental Rights, 2011). Similar legislation is found in many other countries, including the U.S. (Riach and Rich, 2002).

²This idea was originally formulated by Heckman and Siegelman (1993) stating that, if perceived group differences in unobserved variables exists, *preference based* discrimination is unidentified in a correspondence study. Heckman (1998) also discusses this issue.

³This issue is also discussed in Neumark (2013), but his method is applied in Neumark (2012).

⁴These experiments are found in Carlsson and Rooth (2007), Carlsson (2010), Rooth (2010), and Eriksson and Rooth (2014).

⁵See also Heckman (1998).

⁶As HS argue, the results shown here hold for all distributions in the family of bell shaped distributions.

⁷A subtle issue, which Neumark (2012) points out, is that it should actually be deterministic who is invited to a job interview and who is not, if all employers make the same probability calculation and invite applicants based on the probability of passing the threshold. Obviously, this is not the pattern we see in reality. However, it is straightforward to incorporate a random component into the framework that describes the employers’ decision making. One way is to assume firm specific thresholds that are, e.g., normally distributed.

⁸Note that nothing essential changes in our conclusions if we also allow for preference-based discrimination and/or perceived differences in the mean of unobserved variables. This would only affect the probability to be hired for the discriminated group, or both groups, either counteracting or reinforcing the effect from the level of standardization through perceived differences in the variance.

⁹Much of the content of this section is taken from Neumark (2012). For a more detailed explanation of the issues involved in this section the reader should turn to Neumark (2012).

¹⁰The Stata code is available upon request.

¹¹As Neumark points out, failing to reject the null hypothesis of equal coefficients does not decisively rule out the alternative hypothesis of unequal coefficients. On the other hand, with a large number of varying variables, failing to reject a false null hypothesis becomes less likely.

¹²Details of this experiment are found in Eriksson and Rooth (2014), Carlsson and Rooth (2012), and Rooth (2011).

¹³The text (translated from Swedish) for *agreeable* is “My friends and former colleagues would probably state that I am a warm and social person who gets along great with others. Also, I think it is important to ensure people’s needs, and not just focusing on the economic side. I have a strong empathy with people who are less fortunate than

myself and I am active in the Red Cross relief work”, while the text for the opposite is “I usually do not sit and keep my opinions to myself but rather instead say what I think. Some of my former colleagues would probably call me a bit stubborn, but I believe it is important to be correctly understood and to get the job done. I do not mind working alone, since it is then sometimes easier to concentrate on the job task”.

¹⁴The text for *competence* is “I am used to put great effort into work and I always try to do my best. I strive to be as precise as possible so the work tasks need not to be repeated. My old work mates would probably say that I am a person who always manage to get the job done. In addition, I would describe myself as a hardworking and tenacious (sw: *uthållig*) person who withstand stress”, while for the opposite it is “I really like to work but at the same time I think it is important to keep a balance between work and leisure. The best days are the ones when I feel I have done my job and yet have energy to be active in my spare time. It is not important for me to be the best at work and my colleagues would probably describe me as a pretty relaxed”.

¹⁵The included occupations were accountants, business sales assistants, cleaners, computer professionals, construction workers, language teachers in upper compulsory school, math/science teachers in upper compulsory school, mechanics, motor-vehicle drivers, nurses, restaurant workers, shop sales assistants, and teachers in secondary school.

¹⁶The first one is Carlsson and Rooth (2007) and the second one is Carlsson (2010).

¹⁷The quality was raised in the following occupations: accountants, restaurant workers, and shop sales assistants. The quality was lowered in the following occupations: business sales assistants, construction workers, and motor-vehicle drivers.

¹⁸The estimates in Table 1 are obtained using the `dprobit` command in Stata 12.

Competing interests

The IZA Journal of Migration is committed to the IZA Guiding Principles of Research Integrity. The authors declare that they have observed these principles.

Acknowledgements

The authors thank David Neumark, an anonymous referee, participants at the EALE 2013 conference in Turin and at the EBES 2012 conference in Warsaw and seminar participants at Linnaeus University for useful comments and suggestions. Research grants from the Swedish Council for Working Life and Social Research (FAS) and Handelsbanken's Research Foundation are gratefully acknowledged.

Responsible editor: Denis Fougère.

Author details

¹Centre for Labour Market and Discrimination Studies, Linnaeus University, Kalmar, Växjö, Sweden. ²IZA, Bonn, Germany. ³CREAM, London, UK

Received: 8 November 2013 Accepted: 28 May 2014

Published: 23 Jun 2014

References

- Aigner DJ, Cain GG (1977) Statistical theories of discrimination in labor markets. *Ind Labor Relat Rev* 30(2):175–187
- Arrow K (1973) The theory of discrimination. In: Ashenfelter O, Rees A (eds) *Discrimination in labor markets*. Princeton University Press, Princeton
- Baert S, Cockx B, Gheyle N, Vandamme C, Omev E (2013) Do Employers Discriminate Less if Vacancies are Difficult to Fill? Evidence from a Field Experiment. Discussion Paper Series #7145, IZA, Bonn
- Becker GS (1957) The economics of discrimination. University of Chicago press, Chicago
- Bertrand M, Mullainathan S (2004) Are Emily and Greg More Employable than Lakisha and Jamal? A field experiment on labor market discrimination. *Am Econ Rev* 94(4):991–1013
- Borghans L, Duckworth AL, Heckman JJ, Ter Weel B (2008) The economics and psychology of personality traits. *J Hum Resour* 43(4):972–1059
- Carlsson M (2010) Experimental evidence of discrimination in the hiring of first-and second-generation immigrants. *Labour* 24(3):263–278
- Carlsson M, Rooth DO (2007) Evidence of ethnic discrimination in the Swedish labor market using experimental data. *Labour Econ* 14(4):716–729

- Carlsson M, Rooth DO (2012) Revealing taste-based discrimination in hiring: a correspondence testing experiment with geographic variation. *Appl Econ Lett* 19(18):1861–1864
- Eriksson S, Rooth D-O (2014) Do employers use unemployment as a sorting criterion when hiring? Evidence from a field experiment. *Am Econ Rev* 104:1014–1039
- European Union Agency for Fundamental Rights (2011) Council of Europe, Handbook on European non-discrimination law., http://www.echr.coe.int/Documents/Handbook_non_discr_law_ENG_01.pdf. (Accessed 24 February 2014)
- Heckman JJ (1998) Detecting discrimination. *J Econ Perspect* 12(2):101–116
- Heckman JJ, Siegelman P (1993) The Urban Institute audit studies: Their methods and findings. In: Fix M, Struyk R (eds) *Clear and convincing evidence: Measurement of discrimination in America*. Urban Institute Press, Washington DC, pp. 187–258
- Neumark D (2012) Detecting discrimination with audit and correspondence studies. *J Hum Resour* 47(4):1128–1157
- Neumark D (2013) Ethnic Hiring. In: Constant AF, Zimmermann KF (eds) *International Handbook on the Economics of Migration*. Edward Elgar, Cheltenham, UK, and Northampton, USA, pp 193–213
- Phelps ES (1972) The statistical theory of racism and sexism. *Am Econ Rev* 62(4):659–661
- Riach PA, Rich J (2002) Field experiments of discrimination in the market place. *Econ J* 112(482):F480–F518
- Rooth DO (2010) Automatic associations and discrimination in hiring: real world evidence. *Labour Econ* 17(3):523–534
- Rooth DO (2011) Work out or out of work — the labor market return to physical fitness and leisure sports activities. *Labour Econ* 18(3):399–409

10.1186/2193-9039-3-11

Cite this article as: Carlsson et al.: Does the design of correspondence studies influence the measurement of discrimination? *IZA Journal of Migration* 2014, 3:11

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com

III

Does Labor Market Tightness Affect Ethnic Discrimination in Hiring?*

Magnus Carlsson[†]

Luca Fumarco[‡]

Dan-Olof Rooth[§]

Abstract. In this study, we investigate whether ethnic discrimination depends on labor market tightness. While ranking models predict a negative relationship, the prediction of screening models is ambiguous about the direction of the relationship. Thus, the direction of the relationship is purely an empirical issue. We utilize three (but combine into two) correspondence studies of the Swedish labor market and two distinctly different measures of labor market tightness. These different measures produce very similar results, showing that a one percent increase in labor market tightness increases ethnic discrimination in hiring by 0.5-0.7 percent, which is consistent with a screening model. This result stands in sharp contrast to the only previous study on this matter, Baert et al. (forthcoming), which finds evidence that supports a ranking model.

Keywords: Hiring discrimination, ethnic discrimination, labor market tightness, field experiments, ranking models, screening models.

JEL-codes: C93, J15, J21, J71.

* We appreciate the valuable comments from Marianne Bitler, Patrick Button, David Neumark, Paul Nystedt and seminar participants at the Centre for Labor Market and Discrimination Studies at Linnaeus University, the PSI Talks at the University of California, Irvine, and the SWEGPEC Workshop in 2013.

[†] Corresponding author. Linnaeus University Center for Labor Market and Discrimination Studies, Linnaeus University, e-mail: Magnus.Carlsson@lnu.se

[‡] Linnaeus University, e-mail: Luca.Fumarco@lnu.se

[§] Linnaeus University Center for Labor Market and Discrimination Studies, Linnaeus University, e-mail: Dan-Olof.Rooth@lnu.se

1. Introduction

In many Western countries, there are substantial ethnic employment gaps, and research has shown that these gaps can be attributed, at least partly, to ethnic discrimination in the labor market.¹ These ethnic employment gaps also tend to widen during recessions and narrow during booms, suggesting that ethnic discrimination varies with the business cycle.² However, this body of research uses survey or administrative data in which ethnic discrimination is unidentified, and therefore, it is impossible to separate the effect of ethnicity from the effect of unobserved individual characteristics. As a result, it is not possible to conclude that there is a relationship between ethnic discrimination and labor market tightness.

In this study, we employ correspondence studies³ to investigate whether ethnic discrimination depends on labor market tightness.⁴ Correspondence studies solve the problem

¹ A number of recent correspondence studies show that ethnic minorities are discriminated against in hiring. For example, Carlsson and Rooth (2007) study the Swedish labor market and find that job applicants with a Swedish-sounding male name have a fifty-percent higher probability of receiving a callback for a job interview compared to job applicants with a Middle Eastern-sounding male name. Similar results have been found for ethnic minorities in the Australian, Belgian, Norwegian, UK, and US labor markets, to name a few (Booth et al. (2012), Baert et al. (forthcoming), Kaas and Manger (2012), Drydakis and Vlassis (2010), Midtbøen (forthcoming), Fibbi et al. (2006), Woods et al. (2009), Riach and Rich (2002), and Bertrand and Mullainathan (2004)).

² See Biddle and Hamermesh (2013), Bratsberg, Barth and Raaum (2006), and Dustman, Glitz and Vogel (2010). This research mainly finds that the ethnic wage/employment gap is increasing in unemployment.

³ In correspondence studies, researchers send fictitious written job applications to employers with a job vacancy. The job applications are designed to be qualitatively identical except for the applicant's name, which is randomly assigned and chosen to signal ethnicity. Discrimination is then measured as the difference in the share of job interview invitations between the majority and minority.

⁴ A common definition of labor market tightness, which we also use in this paper, is the vacancy-unemployment ratio, where a higher ratio means a tighter labor market. In a tight labor market, there are many vacancies and few unemployed workers looking for jobs, meaning that vacancies are difficult to fill. In a slack labor market,

of unobserved individual variables by introducing random assignment of the signal of ethnicity, thereby solving the identification problem associated with analysis of administrative data.⁵

There are (at least) two categories of theories that help explain why and how ethnic discrimination may depend on labor market tightness (a more detailed explanation is given in Section 2). Ranking models (e.g., Blanchard and Diamond, 1994) predict a negative relationship between the degree of ethnic discrimination and labor market tightness, while in screening models (e.g., Vishwanath, 1989), the prediction could go in both directions, depending on differences in the distribution of unobserved characteristics between majority and minority workers. Hence, determining the direction and the strength of the relationship between the degree of ethnic discrimination and labor market tightness is purely an empirical issue.

One previous study, Baert et al. (forthcoming; henceforth, BCGV), is especially relevant to our study.⁶ BCGV is a correspondence study of the Belgian labor market, which focuses on how the hiring of ethnic minorities is affected by labor market tightness. They find evidence that is supportive of a ranking model because ethnic discrimination is lower when labor market tightness is higher. In fact, when they divide their occupations into two categories

the opposite is true, i.e., there are few vacancies and many unemployed workers looking for jobs, meaning that vacancies are easy to fill.

⁵ Although a correspondence study identifies the level of discrimination in the market, Heckman and Siegelman (1993) and Heckman (1998) show that correspondence studies cannot distinguish between preference-based and statistical discrimination.

⁶ Another study that is related to ours is Kroft et al (2013), which (without considering ethnic dimension) analyzes whether the general hiring behavior of employers varies with labor market tightness. They find that the signaling value of an unemployment spell is stronger in a tighter labor market, which is consistent with a screening model.

defined by the degree of labor market tightness, they find no evidence of discrimination in the tighter labor market, which suggests that labor market tightness explains the entire ethnic gap in hiring in the Belgian labor market.

However, a closer look at the results of the BCGV study puts into question their empirical measure of labor market tightness and, hence, their main result. Intuitively, a necessary property of a valid measure of labor market tightness is that it has a positive general effect on the likelihood of finding a job: the probability of finding a job should be higher when vacancies are difficult to fill. However, the measure of labor market tightness in the BCGV study lacks such a main effect. In fact, the callback rate of natives is even lower in the tighter labor market.⁷ Hence, it is uncertain whether their results truly arise from labor market tightness.

Our main contribution is to test the relationship between ethnic discrimination and labor market tightness in a situation where the measure of labor market tightness has the necessary property in terms of a general positive effect on the probability of finding a job. An additional contribution is that we put great effort into measuring labor market tightness and employ two quite different measures. The most obvious measure of labor market tightness is the vacancy-unemployment ratio. However, such a measure is not available at the local level. Instead, our

⁷ The authors realize this shortcoming and address it in two ways. First, as a sensitivity check of their main analysis, they discard potentially problematic data, excluding almost half of their observations in *bottleneck* occupations (an occupation for which vacancies are difficult to fill, as established by the Flemish Public Employment Service), but even then, the native callback rate in bottleneck occupations is not significantly higher than for non-bottleneck occupations. Second, they aggregate their data into twelve occupation groups and then adopt an alternative approach similar to ours and to the approach of Kroft et al. (2013), calculating the correlation between the occupation callback rate and level of discrimination. However, also this analysis appears uncertain since the data have at least one extreme outlier, which potentially drives their result (see their Figure A1).

first measure of labor market tightness is taken from an additional correspondence study in which we sent job applications with a native female name, and we simply use the callback rate of applicants with a female name as the measure of labor market tightness. This measure should closely mimic the degree of occupation specific labor market tightness, especially because other studies have shown that females are not discriminated against in hiring in the Swedish labor market (see, e.g., Carlsson, 2011, and Eriksson and Lagerström, 2012). For the second measure, we collect the actual number of applicants for each job vacancy in the correspondence testing experiment through a telephone survey. Importantly, both these measures are found to have a general effect on the callback rate because they have a strong association with the callback rate for native Swedish men. For example, if we simply divide the occupations into tight and slack labor markets around the median of labor market tightness using the female callback rate, then the callback rate of native Swedish males is 75 percent higher in the tighter labor market.

Our study also makes an important contribution in terms of arriving closer to a causal interpretation of the results. Because labor market tightness is not randomly assigned, it could potentially pick up other characteristics related to labor market tightness and ethnic differences in the callback rate, leading to biased inferences. Due to the data at hand, we are able to address such potential omitted variable bias by incorporating occupation fixed effects into the regression model. A remaining issue, which also could lead to biased inferences, is whether there are unobserved variables that are correlated with the change in the occupational labor market tightness over time and the change in the ethnic callback rate gap. To address this problem, we add control variables at the firm level.

The next section discusses the theoretical background with a focus on ranking and screening models. Section 3 describes the correspondence studies and construction of the labor market tightness variables. Section 4 presents the main results, and section 5 concludes.

2. Theoretical background

There are (at least) two categories of theories stating that labor market tightness should affect ethnic difference in callback rates, providing different expectations.

The first category of models is ranking models (e.g., Blanchard and Diamond, 1994), which do not consider an ethnic dimension but can be applied to our case with a slight modification. In these models, employers consider the length of the unemployment spell as a signal of low unobserved productivity, and employers then rank the job applicants according to the length of their unemployment spells. One can then exchange the long-term unemployed signal with the ethnic minority signal. Now, the ethnic minority name sends a signal about low unobserved productivity. In a tight labor market, there is less competition for jobs, meaning that the ethnic minority is ranked higher. As a result, the ethnic difference in callbacks is lower in a tight labor market. The opposite pattern holds in a slack labor market.

The second category of models is screening models (Vishwanath, 1989; Lockwood 1991). Screening models also consider the relationship between long-term unemployment and hiring without an explicit ethnic dimension. In these models, it is assumed that some productive characteristics of a worker are unobserved to the employer. However, the employers learn from their own experience of hiring and from other firms, that long-term unemployed workers tend to have worse unobserved skills. Rational employers then use long-term unemployment as a signal of unobserved productivity in the hiring decision. As a result, the probability of being hired will be lower for long-term unemployed workers, implying a duration dependency.

In screening models, duration dependency varies by labor market tightness. Duration dependency will be stronger in a tight labor market, where mostly low-skilled workers are long-term unemployed, and, as a result, unemployment duration is a stronger signal of

productivity. In contrast, duration dependency will be weaker in a slack labor market, where both low- and high-skilled workers are long-term unemployed, and, as a result, unemployment duration is a weaker signal of productivity

Similarly, variation in duration dependency over the business cycle could depend on ethnic group belonging. Intuitively, if minority and majority workers have different skill distributions, the skill composition of long-term unemployed workers may vary differently for the two groups over the business cycle. As a result, the strength of the long-term unemployment signal varies differently for minority and majority workers, and we observe a cyclical pattern in the degree of discriminatory treatment by ethnicity.

To illustrate this cyclical pattern in a more formal way, imagine a bell-shaped distribution of unobserved skills where the variance is relatively larger for ethnic minority workers (while the mean is the same as for the majority). In a tight labor market, only low-skilled minority and majority workers, located far to the left in the skill distribution, are long-term unemployed. The left tail of the skill distribution is more stretched out for the minority group, and this results in an ethnic difference in skill composition among the long-term unemployed, implying that job seeking minority and majority workers in this situation send signals about their group's productivity of different strengths.

Now, imagine that the labor market weakens. Consequently, slightly better skilled workers, who were previously on the margin of being long-term unemployed, become long-term unemployed. Due to the ethnic difference in the variance of the skill distribution, the density of workers at the margin is different for minority and majority workers. As a result, the compositional skill change among long-term unemployed workers changes differently for minority and majority workers. This is why the strength of the long-term unemployment signal varies differently for minority and majority workers over the business cycle, and why the degree of discriminatory treatment changes with labor market tightness.

Notice that the direction of the change in discriminatory treatment depends on the shape of the skill distribution for majority and minority workers. In our example, we assumed that the variance of the skill distribution is relatively larger for minority workers, but the opposite can be true as well, leading to reversed expectations.

Finally, if we find a negative empirical relationship between ethnic discrimination and labor market tightness, we cannot distinguish between ranking and screening models. However, a positive relationship in which ethnic discrimination increases when the labor market tightens can only be explained by screening models.

3. Data

This section describes the correspondence studies and both measures of labor market tightness that we use to test the relationship between the degree of ethnic discrimination and labor market tightness.⁸

3.1 The correspondence studies

The data are taken from three correspondence studies, Carlsson and Rooth (2007), Carlsson (2010), and Rooth (2011). These experiments were conducted between 2005 and 2007 in the Swedish labor market.⁹ In Carlsson and Rooth (2007), we applied to 1,552 job advertisements; in Carlsson (2010), to 1,314 job advertisements; and in Rooth (2011), to 3,821 job advertisements. These studies posed somewhat different research questions but were similar enough to allow us to pool the data. The most important similarity is that all three experiments studied ethnic discrimination against applicants with Middle Eastern-

⁸ We also analyzed an additional measure of labor market tightness: the local unemployment rate. However, this analysis does not produce a statistically significant main effect, that is, the association between the local unemployment rate and the native callback rate, and is therefore not reported.

⁹ The interested reader should turn to these sources for further detail.

sounding names. The experiments are also similar in that, for the most part, the same occupations were included and the same procedures for applying to jobs and receiving responses were used.

The occupations included shop sales assistants, construction workers, computer professionals, motor vehicle drivers, business sales assistants, teachers (math/science and language teachers in upper level compulsory school and secondary school teachers), accountants, restaurant workers, and nurses.¹⁰ These are among the most common occupations in the Swedish labor market and include skilled, semiskilled and unskilled occupations employing varying shares of immigrants (see Eriksson and Rooth, 2014).

The job applications were designed to be realistic but not represent real persons. In addition, because competition from other applicants was considerable, the fictitious applications were constructed to signal well-qualified applicants. The applications consisted of quite general biographies on the first page and detailed CVs listing education and work experience on the second page. To signal ethnicity, common Swedish- and Middle Eastern-sounding male names were randomly assigned to resumes.

In all three experiments, the same procedures were used to apply for jobs and measure callbacks for job interviews. All vacancies in the chosen occupations found on the webpage of the Swedish employment agency were collected. The majority of vacancies were found in the two major cities of Sweden: Stockholm and Gothenburg. Callbacks for job interviews were received via telephone or email.

¹⁰ In addition, Carlsson and Rooth (2007) included preschool teachers, and Rooth (2011) included cleaners and mechanics. However, as we explain below, we pool the data from these experiments in the analysis and utilize variation within occupations between experiments, which implies that the observations in these occupations do not contribute to identification. Therefore, these occupations are excluded from the beginning. In addition, Carlsson (2010) included preschool teachers. In this case, we keep the observations for preschool teachers because, here, we utilize variation between firms (see details below).

In Table A1 in the Appendix, we report the main result for ethnic discrimination observed in the experiments. When the individual probability of receiving a callback for a job interview is regressed on the ethnic minority indicator, we find that ethnic discrimination in hiring is on average about ten percentage points in all three experiments.

Beyond the data on callbacks, we have access to a number of firm characteristics, such as firm size, sex ratio of the employees, and sex of the recruiter.

3.2 Measures of labor market tightness

As stated previously, the most obvious measure of labor market tightness is the vacancy-unemployment ratio, but this measure is not available at the local level. Instead, we employ two other measures that are expected to reflect labor market tightness: the native female callback rate at the occupational level and the actual number of job applicants to each vacancy in the correspondence testing experiment.

Occupational-level female callback rate

Our first measure of labor market tightness is the callback rate for native Swedish female applicants. It is possible to construct this measure because, in connection to the experiments in Carlsson and Rooth (2007) and Rooth (2011), additional applications with a native female name were sent to the employers. The resulting data enable us to construct occupation-specific native female callback rates, which constitute our measure of labor market tightness. The distribution of the female callback rate is shown in Figure A1 in the Appendix. From this figure, it is evident that the female callback rate varies between occupations and within occupations over time.

The native female callback rate is our preferred measure of labor market tightness for several reasons. First, it measures labor market tightness directly and precisely in the

occupations for which we study ethnic discrimination. Second, the female callback rate is likely to be a fair measure of labor market tightness because there is convincing evidence that this group is not discriminated against in hiring in the Swedish labor market (see, e.g., Carlsson, 2011 and Eriksson and Lagerström, 2012). To see that this is an important requirement, imagine that female applicants were discriminated against in the same occupations as the ethnic minority. Then, our measure of labor market tightness would be endogenous because it also captures the degree of ethnic discrimination. Third, as we will show in the empirical section, this measure of labor market tightness has the important property that it shows a strong general positive effect on the callback rate of native men.

Number of job applicants

Our second measure of labor market tightness is the actual number of job applicants per vacancy. The number of job applicants per vacancy is available for a subsample of firms in the Carlsson (2010) data and was collected through a short telephone survey. This measure is available for 196 applications sent to 98 firms. The frequencies for the number of applicants per job vacancy are shown in Figure A2 in the appendix. There is a fair amount of variation in the received number of job applications per vacancy, varying between 3 and 80 applications, with a mean of 31.

In theory, the number of job applications is a sensible measure of labor market tightness because it should be closely related to the vacancy-unemployment ratio in the local labor market in which the firm operates. In contrast to the previous measure of labor market tightness, which varies by occupation, this measure varies across firms that operate in different local labor markets defined by geography and by the type of job. As we will show in the empirical section, this measure also shows a strong general positive effect on the callback rate of native men.

Although this measure seems ideal in theory, the manner in which it was collected creates concern. Due to insufficient resources at the time of the data collection, we only collected information on the number of applicants for a subsample of firms, and hence, there may be a problem with selective participation. Of the 1,314 firms that were part of the correspondence experiment, 824 were sampled to participate. After trying to reach the firms over a period of two weeks, we had only contacted 402 of these firms. Unfortunately, in the end, only 98 firms had this information readily available and could report the number of job applicants for the vacancy, which yielded a sample of 196 applications for the empirical analysis.¹¹ Table A2 in the appendix shows that the average callback rate is much higher in the survey sample, which is mainly due to a higher response rate in the survey for occupations characterized by a high callback rate (teachers, pre-school teachers and nurses). When comparing the occupational distributions of the original and survey samples, the χ^2 is statistically significant at the one percent level ($\chi^2(10)=26.8$). However, an ocular inspection indicates quite small differences, and the statistic is mainly driven by math/science teachers and restaurant workers and is insignificant at the five-percent level when these occupations are excluded ($\chi^2(8)=13.6$). In addition, as the analysis in section 3 will reveal, the measured degree of ethnic discrimination for the firms in the survey sample is approximately eight percentage points and is similar to the corresponding estimate for the full experiment (i.e., ten percentage points).

Another issue of concern is that the number of applicants per vacancy is self-reported, and hence, we cannot exclude the possibility of classical measurement error in the data.

¹¹ There could be several reasons for not reporting this number. Some employers stated that they did not remember that particular vacancy. In other cases, we could not locate the person responsible for recruitment. In others, the employer did not want to participate. A significant number of employers reported an imprecise number, such as “many.” All non-number answers were discarded.

However, as previously mentioned, we find that this measure has a strong main effect on the callback rate.

4. Results

In this section, we first introduce the model specification and then proceed with the empirical analysis using both measures of labor market tightness.

4.1 Model Specification

To test the relationship between ethnic discrimination and labor market tightness, we estimate the following simple model as our main specification:

$$Callback_i = \beta_0 + \beta_1 Minority_i + \beta_2 Tightness_i + \beta_3 [Minority_i * Tightness_i] + \varepsilon_i$$

$Callback_i$ is a dummy variable indicating whether application i resulted in a job interview, $Minority_i$ is a dummy variable indicating whether application i was assigned an ethnic minority name, and $Tightness_i$ is a continuous (mean adjusted) variable measuring either of our two measures of labor market tightness. The constant β_0 gives the callback rate for the ethnic majority, while β_1 is the difference in the callback rate between minority and majority job applicants measured at the mean value of labor market tightness. β_2 is the general effect of labor market tightness on the callback rate of native men. For any sensible measure of labor market tightness, β_2 is positive because an increase in labor market tightness should increase the probability of being invited for a job interview. β_3 is the main parameter of interest, measuring whether the ethnic minority is differently affected by labor market tightness compared to native men. A negative coefficient β_3 says that ethnic discrimination increases with labor market tightness, meaning that an increase in labor market tightness does not help

the minority as much as the majority in getting a callback for a job interview, while a positive coefficient says that ethnic discrimination decreases with labor market tightness. In light of the theoretical discussion in the introduction, and in section 2, on how labor market tightness impacts upon ethnic discrimination, we have ambivalent expectations about the sign of this coefficient.

While the ethnic dummy is exogenous by design, this is not true for labor market tightness. Therefore, we are worried that the estimate of labor market tightness captures a spurious correlation due to omitted variables. To address this potential problem, and to arrive at an estimate with a causal interpretation, we include occupation fixed effects in the model. For our preferred measure of labor market tightness (the female callback rate), which is measured at the occupational level, this is possible because this measure is available for two separate correspondence studies conducted at different times and containing the same occupations. For the second measure of labor market tightness (the number of job applications), this is possible because the measure is collected at the firm level. Adding occupation fixed effects neutralizes the influence of unobserved factors that relate to labor market tightness, ethnic discrimination and the callback rate for a job interview operating at the occupational level. Hence, if employers in certain occupations are discriminating more, or less, in unobserved ways, then this will be controlled for. A remaining issue that may cause bias is that the interaction term may be correlated with some time-variant unobserved characteristics of the occupations. To address this potential bias, we include a number of firm characteristics that vary within occupations across experiments.

4.2 Results: occupational female callback rate

In this section, we focus on our first measure of labor market tightness – the occupation specific female callback rate. We start by regressing the callback dummy on the ethnic

minority indicator and the measure of labor market tightness. The results in the first column of Table 1 reveal that applications assigned a Middle Eastern sounding male name have a 9.5 percentage points lower probability of receiving a callback for a job interview compared to applications assigned a Swedish sounding male name. The results also show that the return to labor market tightness is positive and statistically significant at the one-percent level. An increase of one percentage point in the occupation specific callback rate of native females increases the probability of receiving a callback for a job interview by approximately 0.8 percentage points for native men.

Next, we add the main variable of interest, the interaction between the measure of labor market tightness and the ethnic minority indicator (see column 2). Interestingly, we find that the estimate is negative and significant at the one percent level, which is opposite the result found in BCGV. This result implies that when labor market tightness increases, the probability of receiving a callback for a job interview increases more slowly for the ethnic minority relative to the majority, that is, ethnic discrimination increases. The interpretation of the estimate is that a ten percentage point increase in the callback rate for female applicants is associated with a 1.5 percentage point increase in ethnic discrimination. If we instead express this result in terms of standard deviations, then the level of discrimination increases from ten to twelve percentage points (that is, by 20 percent) when labor market tightness increases by one standard deviation, which amounts to a 13 percentage point increase (that is, a 43 percent increase from a mean of .30) in the female callback rate.¹² The estimate can also be expressed as an elasticity of approximately 0.5 ($0.20/0.43$), that is, a one-percent increase in labor market tightness is associated with a 0.5-percent increase in ethnic discrimination.

¹² The estimated increase in discrimination by 2 percentage points is obtained multiplying the estimate of the interaction effect by .13 (the standard deviation).

*** Table 1 about here ***

In column 3, we add occupation fixed effects to approach a causal interpretation by relying on variation within occupations over time. This is an important extension of BCGV, who rely on cross sectional variation. Interestingly, despite using a quite different type of variation, we arrive at almost identical estimates, as reported in the previous column.

Although the occupation fixed effects analysis captures many of the most likely biases, we cannot rule out that the characteristics of firms are different across experiments (within occupations). Such differences could lead to biased results if the firm characteristics are correlated with labor market tightness and also affect discrimination. In an attempt to address such potential bias, we control for the firm characteristics that are available from the experiment: firm size, female share at the firm, and sex of the recruiter¹³. However, the estimates in column 4 are unaffected by these control variables.

4.3 Results: Number of job applicants

Before turning to the results in this section, notice that if the number of job applicants per vacancy increases, then labor market tightness decreases because it becomes easier for firms to fill vacancies. Therefore, we must multiply the number of job applications by (-1) to obtain a measure of labor market tightness, which we do for all regressions in this section. This transformation also makes the interpretation of the labor market tightness parameter consistent with the previous section.

¹³ For a small number of firms, information on these variables is missing. We do not exclude these firms from the regressions. Instead, we include a dummy variable that indicates whether the information is missing. For continuous variables, we also impute the mean value for variables with missing information.

We start by establishing that our measure of labor market tightness has a general positive effect on the callback rate. As in the previous section, we first regress the callback dummy on the ethnic minority indicator and the measure of labor market tightness. The first column of Table 2 reveals a very similar picture to the previous analysis. First, we find that applications assigned a Middle Eastern-sounding male name have an approximately eight percentage point lower probability of receiving a callback for a job interview compared to applications assigned a Swedish-sounding male name. Moreover, this second measure of labor market tightness provides a sensible estimate because the return to labor market tightness is positive and statistically significant at the five-percent level. The estimate reveals that a decrease of ten job applications to a vacancy increases the probability of receiving a callback for a job interview by five percentage points.

**** Table 2 about here ****

In the second column, we add the interaction between the measure of labor market tightness and the ethnic minority indicator, which produces the estimate of main interest. Again, we find that the return to labor market tightness is lower for the ethnic minority than for the majority, and, hence, ethnic discrimination is larger when the labor market is tighter. The interpretation of the estimate is that an increase of ten job applications to a vacancy is associated with an approximately 1.7-percentage point increase in discrimination. In terms of standard deviations, the estimate means that if labor market tightness increases by one standard deviation, corresponding to a 71-percent increase in the number of job applications from the mean, then the level of discrimination increases from eight to twelve percentage

points (or by 50 percent).¹⁴ In this case, the elasticity of ethnic discrimination with respect to labor market tightness is approximately 0.7 (0.50/0.71), that is, a one-percent increase in labor market tightness is associated with a 0.7-percent increase in ethnic discrimination.

Next, we add firm-level characteristics to the regression model. These variables control for firm size, which we expect to be correlated with the number of applicants for a job if a vacancy at a large firm attracts more job applicants, share of women at the firm, and sex of the recruiter¹⁵. The estimates are unaffected by their inclusion (see column 3).

As stated previously, we would ideally also control for occupation and only use variation in the number of job applications to a vacancy within occupations. However, after adding occupation fixed effects to the regression reported in column 3, the positive association between the callback rate and labor market tightness no longer exists (see column 4). Although the estimate of the interaction variable is unaffected, which is still negative and statistically significant, we believe that this specification is likely asking too much of the data and should be interpreted with caution.

To sum up, we obtain similar results in this section compared to the previous section, and the elasticities of discrimination with respect to labor market tightness are found to be similar, at 0.5 and 0.7, for the first and second measures of labor market tightness, respectively. Hence, two rather different measures of labor market tightness used in two different

¹⁴ The increase in ethnic discrimination of -0.037, or four percentage points, is produced by multiplying the estimate of the interaction effect by 2.2 (the standard deviation is 22 and the estimate reported in the table is multiplied by 10). The mean number of job applications is 31, which facilitates the 71-percent increase in labor market tightness.

¹⁵ For these variables, data are missing for a small number of cases. As before, we do not exclude these firms from the regressions. Instead, we include a dummy variable that indicates whether the information is missing. For continuous variables, we also impute the mean value for variables with missing information.

correspondence studies produce similar results for how labor market tightness effects ethnic discrimination in hiring.

5. Conclusion

The strength of correspondence studies compared to other methods of measuring ethnic discrimination is their ability to identify discrimination. However, it has proven difficult to understand some of the more subtle patterns in the results of a typical correspondence study, such as variation in ethnic discrimination across occupations. A potential explanation for variation in ethnic discrimination across occupations is that ethnic discrimination depends on labor market tightness, which may vary by occupation.

Using two quite different measures of labor market tightness, we find that in Sweden, ethnic discrimination in hiring increases with labor market tightness, that is, an improving labor market produces more job opportunities for natives than for the ethnic minority. This result stands in sharp contrast to the only previous study that has investigated this issue, Baert et al. (forthcoming). On a more general note, our results hint that a screening model of hiring better explains the data than a ranking model, which is consistent with the results reported in Kroft et al (2013).

Finally, our result points to an important fact for policy makers – ethnic discrimination increases when the labor market improves, which conflicts with research on ethnic employment gaps and the business cycle. Therefore, one piece of advice for policy makers is to maintain measures to reduce ethnic discrimination even when the business cycle improves.

References

- Baert, S., Cockx, B., Gheyle, N., & Vandamme, C. (Forthcoming). Is there Less Discrimination in Occupations where Recruitment is Difficult? *Industrial and Labor Relations Review*.
- 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.
- Biddle, J. E., & Hamermesh, D. S. (2013). Wage discrimination over the business cycle. *IZA Journal of Labor Policy*, 2(1), 1-19.
- Blanchard, O. J., & Diamond, P. A. (1994). Ranking, Unemployment Duration, and Wages. *Review of Economic Studies*, 61(3), 417-434.
- Booth, A. L., Leigh, A., & Varganova, E. (2012). Does ethnic discrimination vary across minority groups? Evidence from a field experiment. *Oxford Bulletin of Economics and Statistics*, 74(4), 547-573.
- Bratsberg, B., Barth, E. & Raaum, O. (2006). Local Unemployment and the Relative Wages of Immigrants: Evidence from the Current Population Surveys. *Review of Economics and Statistics*, 88(2): 243–263.
- Carlsson, M. (2010). Experimental Evidence of discrimination in the Hiring of First- and Second-generation Immigrants. *Labour*, 24 (3), 267-278.
- Carlsson, M. (2011). Does hiring discrimination cause gender segregation in the Swedish labor market? *Feminist Economics*, 17(3), 71-102.
- Carlsson, M., & D.-O., R. (2007). Evidence of Ethnic Discrimination in the Swedish Labor Market Using Experimental Data. *Labor Economics*, 14 (4), 716-729.

- Drydakis, N., & Vlassis, M. (2010) Ethnic Discrimination in the Greek Labour Market: Occupational Access, Insurance Coverage and Wage Offers. *The Manchester School*, 78(3), 201-218.
- Dustman, C., Glitz, A. & Vogel, T. (2010) "Employment, Wages, and the Economic Cycle: Differences between Immigrants and Natives", *European Economic Review*, 54: 1–17.
- Eriksson, S., & Lagerström, J. (2012). Detecting discrimination in the hiring process: evidence from an Internet-based search channel. *Empirical Economics*, 43(2), 537-563.
- Eriksson, S., & Rooth, D.-O. (2014). Do employers use unemployment as a sorting criterion when hiring? Evidence from a field experiment. *The American Economic Review*, 104(3), 1014-1039.
- Fibbi, R., Lerch, M., & Wanner, P. (2006). Unemployment and discrimination against youth of immigrant origin in Switzerland: when the name makes the difference. *Journal of International Migration and Integration/Revue de l'integration et de la migration internationale*, 7(3), 351-366.
- Heckman, J.J. & Siegelman P. (1993). The Urban Institute audit studies: Their methods and findings. In: Fix M., Struyk R. (eds.) *Clear and convincing evidence: Measurement of discrimination in America*, 187-258. Washington DC: The Urban Institute.
- Heckman, J. J. (1998). Detecting Discrimination. *The Journal of Economic Perspectives*, 12(2), 101-116.
- Kaas, L., & Manger, C. (2012). Ethnic Discrimination in Germany's Labour Market: A Field Experiment. *German Economic Review*, 13(1), 1-20.
- Kroft, K., Lange, F., Notowidigdo, M. J., & Katz, L. F. (2013). Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment. *The Quarterly Journal of Economic*, 128(3), 1123-1167.

- Lockwood, B. (1991). Information externalities in the labour market and the duration of unemployment. *The Review of Economic Studies*, 58(4), 733-753.
- Midtbøen, A. H, (forthcoming), Discrimination of the Second Generation: Evidence from a Field Experiment in Norway. *Journal of International Migration and Integration/Revue de l'integration et de la migration internationale*.
- Riach, P. A., & Rich, J. (2002). Field Experiments of Discrimination in the Market Place. *Economic Journal*, 112(483), F480-F518.
- Rooth, D.-O. (2011). Work Out or Out of Work - The Labor Market Return to Physical Fitness and Leisure Sports Activities. *Labour Economics*, 18 (3), 399-409.
- Vishwanath, T. (1989). Job Search, Stigma Effect, and Escape Rate from Unemployment. *Journal of Labor Economics*, 7(4), 487-502.
- Woods, M., Hales, J., Purdon, S., Sejersen, T., & Hayllar, O. (2009). A test for racial discrimination in recruitment practice in British cities. Department for Work and Pensions. Research Report No 607.

Tables:

Table 1. The probability of a job interview. Labor market tightness measured by the female callback rate.

	(1)	(2)	(3)	(4)
Constant	.275*** (.015)	.275*** (.014)	.270*** (.016)	.250*** (.024)
Minority	-.095*** (.011)	-.095*** (.010)	-.095*** (.010)	-.095*** (.010)
Labor market tightness	.831*** (.070)	.905*** (.081)	.772*** (.093)	.786*** (.104)
Labor market tightness * Minority	-	-.150*** (.035)	-.150*** (.035)	-.150*** (.036)
Occupation FEs	No	No	Yes	Yes
Firm characteristics	No	No	No	Yes
Number of observations	8,514	8,514	8,514	8,514

Notes: The dependent variable is an indicator of whether the applicant was invited to a job interview. Labor market tightness is measured by the occupation specific callback rate of native females. The regressions that include occupation fixed effects also include an experiment fixed effect. The firm characteristics are firm size, share of females at the firm, and sex of the recruiter. The estimates are from a linear probability model. Standard errors are clustered at the occupational level in all models. *** p <.01, ** p <.05, * p <.1.

Table 2. The probability of a job interview. Labor market tightness measured by applicants per vacancy.

	(1)	(2)	(3)	(4)
Constant	.592*** (.048)	.592*** (.048)	.461*** (.138)	.288* (.164)
Minority	-.082*** (.028)	-.082*** (.028)	-.082*** (.029)	-.082*** (.029)
Labor market tightness (estimate multiplied by 10)	.054*** (.020)	.062*** (.021)	.051** (.023)	.026 (.023)
Labor market tightness * Minority (estimate multiplied by 10)	-	-.017** (.009)	-.017* (.009)	-.017* (.009)
Firm characteristics	No	No	Yes	Yes
Occupation FEs	No	No	No	Yes
Number of observation	196	196	196	196

Notes: The dependent variable is an indicator of whether the applicant was invited to a job interview. Labor market tightness is measured by the number of job applications. To facilitate the interpretation of the estimates of labor market tightness, the measure has been multiplied by 10, and is interpreted as the change in the probability of a callback when labor market tightness increases by ten job applicants. The firm characteristics are firm size, share of females at the firm, and sex of the recruiter. The estimates are from a linear probability model. Standard errors are clustered at the job level in all models. *** p <.01, ** p <.05, * p <.1.

Appendix

Table A1. The probability of a job interview across experiments.

	Carlsson and Rooth (2007)	Carlsson (2010)	Rooth (2011)
Minority	-.102*** (.026)	-.098*** (.011)	-.094*** (.008)
Number of observations	3,104	2,628	5,636

Notes: The number of observations is the number of job advertisements multiplied by two, which is the number of applications sent to each job advertisement. In columns 1 and 2, the number of observations is $2 \times 1,552 = 3,104$ and $1,314 \times 2 = 2,628$, respectively. In Rooth (2011), only one job application was sent to a job advertisement in some instances, which explains why we in this case have fewer observations than 7,642 ($2 \times 3,821$). The dependent variable is an indicator of whether the applicant was invited to a job interview. The estimates are from a linear probability model. *** $p < .01$, ** $p < .05$, * $p < .1$.

Table A2. The distribution of occupations for the original field experiment and the sub-sample survey data and the mean callback rate.

	Carlsson (2010)	Survey data
Shop sales assistants	.114	.051
Construction workers	.036	.031
Computer professionals	.059	.071
Motor-vehicle drivers	.059	.041
Business sales assistants	.196	.214
Math/science teachers	.023	.061
Secondary school teachers	.037	.071
Accountants	.105	.061
Restaurant workers	.114	.031
Nurses	.114	.153
Pre-school teachers	.143	.215
Mean callback rate	.325	.551
Number of observations	2,628	196

Notes: For information on the full sample, see Carlsson (2010).

Figure A1. The callback rate for female applicants by occupation

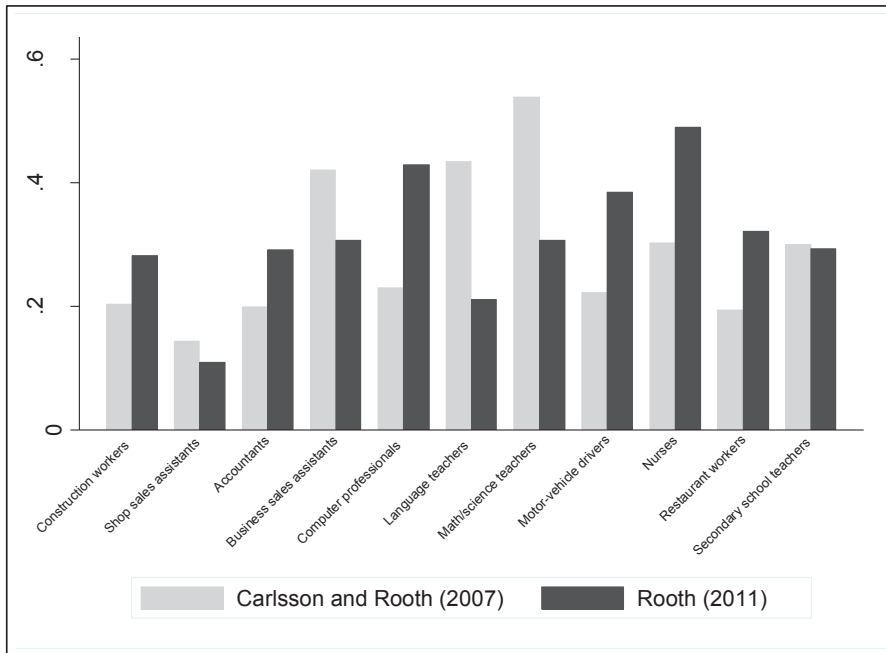
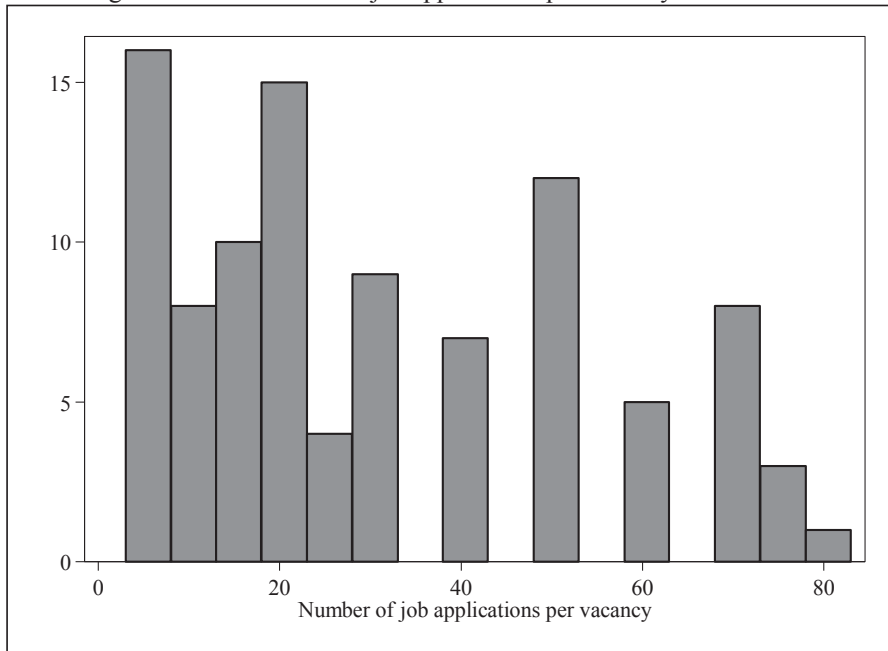


Figure A2. The number of job applications per vacancy. N=98 firms.



IV

Relative Age Effect on Labor Market Outcomes for High-Skilled Workers – Evidence from Soccer

by

Luca Fumarco[†]

Abstract. In sports and education contexts, children are divided into age groups that are arbitrary constructions based on admission dates. This age-group system is thought to determine differences in maturity between pupils within the same group, that is, relative age (RA). In turn, these within-age-group maturity differences produce performance gaps, that is, relative age effects (RAEs), which might persist and affect labor market outcomes. I analyze the RAE on labor market outcomes using a unique dataset of a particular group of high-skilled workers: soccer players in the Italian major soccer league. In line with previous studies, evidence on the existence of an RAE in terms of representativeness is found, meaning that players born relatively early in an age group are over-represented, while players born relatively late are under-represented, even accounting for specific population trends. Moreover, players born relatively late in an age group receive lower gross wages than players born relatively early. This wage gap seems to increase with age and in the quantile of the wage distribution.

JEL-Classification: J24, J31, J71, L83, M53

Keywords: Relative age, labor markets in sports

[†] Linnaeus University, e-mail: luca.fumarco@lnu.se

Acknowledgements. I would like to thank my supervisors, Dan-Olof Rooth and Magnus Carlsson, for guidance and many useful suggestions; Giambattista Rossi, for the provision of most of the dataset; Simone Scarpa, Peter Karlsson, Håkan Locking and PhD fellows from the Department of Economics and Statistics of the Linnaeus University, for valuable feedback; participants at the seminars series of the Centre for Labour Market and Discrimination Studies, at the Linnaeus University; participants at the AIEL conference, 2014, in Pisa; and participants at the SWEGPEC PhD workshop, 2014, in Jönköping.

1 Introduction

Extensive empirical evidence shows that children born late in the education and sports admission year are systematically disadvantaged throughout childhood up to the late teens. Scholars from different disciplines explain this evidence through the existence of the so-called relative age effect. This concept has recently even gained popularity outside academia (e.g., Gladwell, 2008; Dubner & Levitt, 2010).

The relative age effect is produced by similar complex mechanisms in education and sports. In both contexts, age groups are formed using arbitrary admission dates that determine some children to be older than others within the same age group. This chronological difference, called relative age (henceforth RA), is responsible for early differences in maturity (e.g., Bedard & Dhuey, 2006; Musch & Hay, 1999), which cause a performance gap, that is, the relative age effect (henceforth RAE),¹ and affect children's achievements. Because of its nature, this effect is expected to dissipate with age and eventually disappear. However, it might persist, and even widen, because of certain characteristics of the human capital accumulation process that lead to "path dependence" (Bryson et al., 2014, p.12), which means that children born early in the admission year are more likely to be perceived as talented (e.g., Allen & Barnsley, 1993), and thus they are given more chances to develop their skills (e.g., teachers and parents motivate them more, or children could be provided with superior educational quality).

Although a significant consensus exists regarding the negative RAE on relatively young children's achievements, no equivalent consensus exists on the RAE on labor market

¹ Consider the case where all children who turn 6 in a given calendar year are expected to start the first grade of primary school in that year (i.e., the admission date is the 1st of January; note that the beginning of the school year is irrelevant). In the same class, there might be children who turn 6 in January and children who turn 6 in December; the relatively older pupils born in January are 17% older than the relatively younger pupils born in December. This chronological difference is the RA, which causes differences in terms of maturity, leading to a performance gap; this performance difference is the RAE.

outcomes (Ponzo & Scoppa, 2014). Whether there is such a long-run effect is a compelling economic question (e.g., Allen & Barnsley, 1993; Bedard & Dhuey, 2006).

One of the possible reasons for this lack of consensus is the presence of two important confounders that affect scholars' analyses: "season-of-birth effects" and heterogeneous ages within age groups. The "season-of-birth effects" are confounding factors because they are unrelated to within-age-group maturity differences and are due to climatic, environmental, sociocultural and biological factors (Musch & Grondin, 2001). Season of birth explains performance gaps between children born in the *same calendar year* based on the *position of their birthdates within the calendar year*, whereas the RAE explains performance gaps between children born in the *same admission year* with the *maturity gap caused by the relative position of their birthdates within the admission year*. On the one hand, when the beginning of the admission year coincides with a period of the calendar year that conveys advantages due to seasonal effects, the estimate of the RAE is likely to be upward biased.² On the other hand, the estimate could be downward biased if later months of the selection year coincide with a period of the calendar year that conveys advantages to children born within that period.^{3,4} In addition, the presence of heterogeneous ages within age groups may bias the results of RAE analyses. Consider the education context, where children born late in the admission year, that is, relatively young children, are held back one year, that is, they either repeat a grade or they enter primary school one year later. These children end up in an age group where the typical children are younger; thus, they become relatively older children in

² Consider the case where the school admission year coincides with that of sport, and the researcher was interested only in the RAE from either education or sport, not their combined effect. The estimates would be biased (Musch & Hay, 1999; Helsen et al., 2012).

³ If the admission date was shifted by a few months, e.g., shift the admission date in the example in Footnote 2 by 6 months, the estimated RAE would be downwardly biased. Additionally, if households with high socioeconomic status tend to give birth in months that do not coincide with the beginning of the admission year, as in the US (Bound & Jaeger, 2001) and Sweden (Carlsson et al., forthcoming), the estimate of RAE from education would be downward biased.

⁴ The seminal paper by Angrist and Kruger (1992) might be interpreted as a particular case where the estimate of the RAE could be downward biased because of the school-leaving age. The authors find that pupils born at the beginning of the admission year attain less schooling than their younger peers because they are legally allowed to leave school before graduation.

their new age group (Bedard & Duhey, 2006). In this case, the estimate of the RAE might be downward biased.⁵ Moreover, as Bedard and Duhey (2006) suggest, in countries where pre-school institutions are not free, the possibility of redshirting, that is, entering primary school one year later, might also affect RAE estimates via socioeconomic status. In countries such as the US, high socioeconomic status parents are more likely to be able to afford one extra year of pre-school. In this case, an estimate of the RAE is likely to be even more downward biased.

The goal of this study is threefold. First, this paper adds to the existing economic literature by investigating different aspects of the RAE on labor market outcomes, including the long-run RAE. The focus is on a particular group of high-skilled workers: professional soccer players from the Italian major league, that is, Serie A. Second, this paper aims to provide a descriptive general framework of the RAE by bringing articles from different disciplines to the reader's attention. The literature review in this article stresses the importance of different mechanisms and different evidence on the RAE that is sometimes neglected in studies in economics. Third, this article proposes the use of a quantile regression to obtain more insights into the long-run economic RAE.

What is the reason for analyzing soccer players? The first reason for studying this particular group of workers is that season-of-birth effects seem to play a minor role in the soccer domain. There is evidence that seasonal effects have only an attenuated—if not null—effect on the mechanisms leading to an RAE in professional soccer. Munch and Hay (1999) explain that, at the end of the 1980s in the major soccer leagues of Germany, Brazil, Australia and Japan, soccer players born early in the admission year were consistently over-represented. This result is consistent with the RAE: throughout the years of sports activity, more early-born soccer players were considered more talented and thus reached the top leagues. This

⁵ In a similar manner, the RAE in sport might be nullified. As documented by Parent-Harvey et al. (2013) and Böheim and Lackner (2012), when the selection of athletes into professional competition is based on a draft system, relatively younger athletes might delay entry into professional sport by one year to overcome developmental differences.

result is obtained despite a number of differences between these four countries: admission dates, reversed seasons, typical climate, biological characteristics, and socio-cultural factors. In addition, studies on the effect of a shift in the admission date provide results that are consistent with the RAE, ruling out alternative explanations. Munch and Hay (1999) show that a shift of the Australian admission date by a few months led to a corresponding reduction in the players' birthrate for the early months under the previous admission date. Helsen et al. (2000) study the effect of a similar shift in the admission date in Belgium and find a corresponding adjustment. Seasonal effects may hardly offer an explanation for performance gaps between players born in two adjacent months, where one month is before and one after the admission date (e.g., Barnsley & Thompson, 1988; Ponzio & Scoppa, 2014).

A second reason is that the presence of age groups with heterogeneous ages is limited in soccer. In Italy, which is the context of this analysis, the age-group system for soccer is very strict, so that the bias given by heterogeneous ages within age groups should be less of an issue.⁶ Moreover, related to this aspect, the effect of household socioeconomic status via redshirting is avoided a priori because redshirting is not possible; additionally, there are reasons to believe that such a practice would not matter anyway. For what reason would someone assume that only households with high socioeconomic status can afford to have their children start to play soccer later? In conclusion, no particular identification strategy must be adopted to address the bias caused by age groups with heterogeneous ages.⁷

A third reason to study the RAE in the soccer players' labor market is the quality of the available data. As stated by Kahn (2000), data are very detailed within the sports field. For instance, data on employees' performance and compensation are accessible, and the data on

⁶ According to rules set by the Italian Football Federation (FIGC), a team may deploy one overage player in regular matches only in the last juvenile category, and in only one intermediate category may a team deploy underage players.

⁷ For instance, because of the possibility of postponing or accelerating entry into school, Bedard and Duhey (2006) and Ponzio and Scoppa (2014) adopt an instrumental variable estimation strategy where they instrument the students' actual age with their so-called *expected age*, that is, the age children should have at the moment their performance is measured based on both their month of birth and the admission month used in the schooling system.

the employees can be easily matched to those of their employers throughout the career and are often more accurate and detailed than usual microdata (Kahn, 2000).

Based on the previous literature, the first hypothesis tested in this paper concerns the presence of the RAE in terms of representativeness. In the presence of the RAE, the observed number of Italian players born at the beginning of an admission year should be larger than the expected number based on the birthrate of the general population; the contrary should be true for those players born at the end of the admission year.⁸ The RAE mechanism suggests that relatively older players are often perceived as talented in their early ages, they are (more or less formally) streamed (Allen & Barnsley, 1993), and they reach Serie A more frequently than their relatively younger peers.

The results provide evidence for the existence of the RAE in terms of representativeness in Serie A among Italian players. Moreover, an additional analysis suggests the presence of a specific trend that is explainable through the RAE: the over-representation decreases and turns into under-representation as the end of an admission year is approached.

The RAE in terms of wage gaps is also analyzed. The RA framework suggests three different possible results. Traditionally, the RA suggests that on average, relatively older players should perform better (Allen & Barnsley, 1993) and thus should receive higher wages because they have had a relative advantage throughout the pre-labor market period. However, the opposite result is illustrated by Ashworth and Heyndels (2007), Gibbs et al. (2012), and Bryson et al. (2014). Positive selection and peer effects could positively affect relatively younger players' performances and lead to higher wages. The best relatively younger children manage to overcome these difficulties and eventually benefit from learning and training with

⁸ Moreover, in the presence of the RAE, players born in January would be over-represented in the sample, players born in March would still be over-represented but to a lower extent, players born in October would be under-represented, and players born in December would be the most under-represented. This result would hold even when trends in the general population birthrate are accounted for.

stronger peers.⁹ In addition, recent studies suggest that the performance gap disappears in the labor market because the discriminatory streaming criteria that affect pre-labor market achievements cease to be relevant (Crawford et al., 2013). To the best of my knowledge, only three other studies investigate the RAE on wages for high-skilled workers: Kniffin and Hanks (2013) for PhD students, Böheim and Lackner (2012) for American football players, and Ashworth and Heyndels (2007) for German soccer players.

The main results provide statistically significant evidence that relatively younger players earn lower wages, supporting the theory according to which the RAE also negatively affects performance in the long run (Allen & Barnsley, 1993). Additional analyses suggest that this wage gap might be the largest at entry into the labor market, while in the remainder of the career the wage gap is smaller, although it tends to increase toward the end of the career. This particular development of the wage gap could be due to players' career choices. As a further contribution to the economic literature, this paper analyzes whether the RAE on wages differs by wage quantile. To the best of my knowledge, none of the existing studies analyzes the wage gap using a quantile regression. This analysis is important when investigating a labor market characterized by a strongly positively skewed wage distribution, and when the researchers hypothesize the existence of peer effects or positive selection. The results point to the possibility that the wage gap could increase in the quantile of the wage distribution; in turn, this result could imply the absence of positive peer effects and selection for relatively younger players.

The remainder of the paper proceeds as follows. Section II presents a summary of the literature review on the RAE in education and sport; Section III discusses the data and

⁹ Alternatively, Williams (2010) hypothesizes that in the long run, relatively younger players might outperform their relatively older peers because relatively younger athletes experience a more complete training, while their relatively older peers place less emphasis on skills development because they are primarily selected based on their physical attributes.

presents descriptive statistics; Section IV presents the empirical methodology; Section V illustrate the results; and Section VI concludes.

2 The Relative Age Effect: Mechanisms and Evidence

2.1 Mechanisms

RAEs in education and sports contexts appear to be similar in their mechanisms and consequences for people's achievements. The similarities between these two contexts are emphasized when a competitive streaming process takes place.

In education, the RAE is initially caused by differences in children's cognitive development. These differences trigger misjudgments of pupils' talent and, eventually, more or less flexible streaming (Bedard & Dhuey, 2006). In a case of formal streaming, some children are assigned to vocational schools and others to academic schools, or they are divided into ability-based reading groups (Bedard & Dhuey, 2006). When there is no formal streaming, social interactions between children, parents, and educators play a prominent role (Hancock et al., 2013) because stronger students are encouraged to progress, while weaker students are allowed to lag behind (Bedard & Dhuey, 2006). An example of the social interaction effect is the Pygmalion effect, which predicts that teachers, trainers, and parents' expectations regarding children's ability trigger self-fulfilling prophecies (Musch & Grondin, 2001; Hancock et al., 2013). Another example is the Galatea effect, which predicts that children's expectations of themselves trigger self-fulfilling prophecies (Hancock et al., 2013).

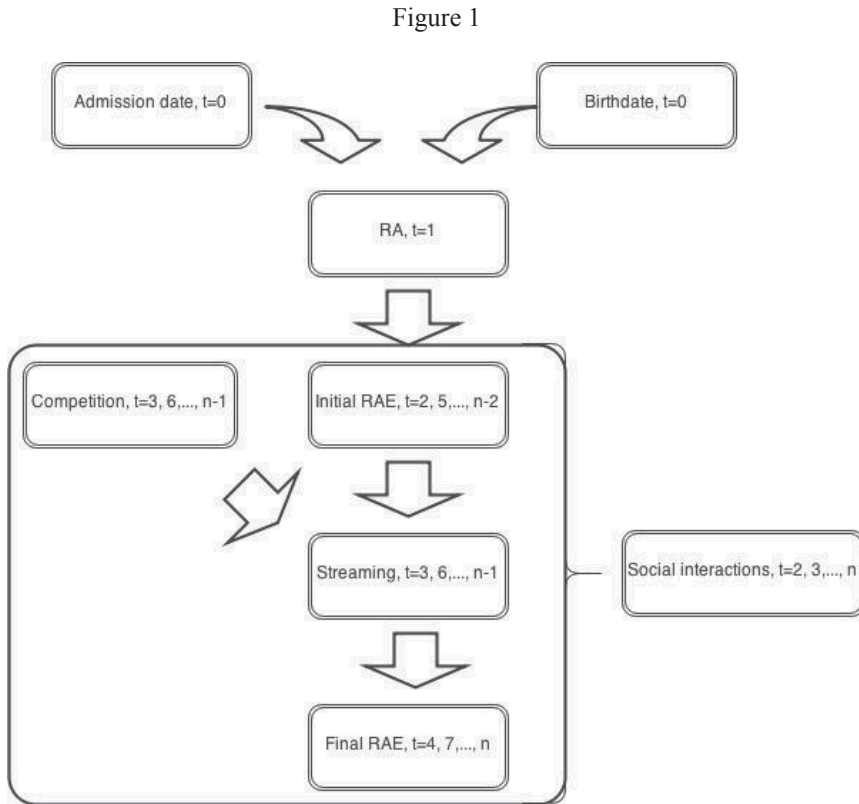
The RAE in sports differs from that in education with respect to at least three aspects. First, the RAE in sports is caused by initial differences in children's cognitive *and physical* development (Allen & Barnsley, 1993), conveying an additional edge to relatively older children. Second, competition might be tougher from the early stages of youth sports. The competition level is determined by a number of factors, such as the number of teams within a region, the number of available spots per team, and the number of children who can

eventually compete regularly (Allen & Barnsley, 1993; Musch & Grondin, 2001). Considering the general case in soccer, where the number of teams per region and the number of available spots per team are not binding, children join a team by simply paying a fee. Only a limited number of children per team eventually get to play often in regular matches; the children in the starting team plus the substitutes who actually enter the pitch will accumulate experience and skills more rapidly. Because relatively older children are more mature, they perform better and improve more rapidly. In a case where the number of teams and the number of available spots per team were binding, competition might be fiercer, and teams could select children based on their perceived talent, increasing the effect of competition, for instance, in national youth summer camps (Glamser & Vincent, 2004) or youth national teams (Williams, 2010).¹⁰ Third, in sports, children may drop out (Barnsley & Thompson, 1988; Helsen et al., 1998).¹¹ While school is compulsory at early ages, and it is only possible to drop out during later years of high school or while attending university, sports are based on voluntary participation (Musch & Hay, 1999; Musch & Grondin, 2001).

¹⁰ To the best of my knowledge, only one study investigates the RAE in soccer academies (Carling et al., 2009). The authors find that the relative age effect might not always determine significant performance gaps. However, the study analyzes only the physical components of young players' performance.

¹¹ This option does not mean that children drop sports activities in general; they could simply change sports, opting for one in which the admission date either has lower or no importance (Williams, 2010) or one that provides them with a positive RAE (Thompson et al., 1999), thus contributing to the RAE in that sport.

Overall, the RAE mechanism found in education and sports might be summarized by Figure 1.¹²



At $t=0$, there is a given admission date and a given birthdate that cause the RA at $t=1$. The RA at $t=1$ creates the initial RAE at $t=2$, and then there is a (more or less formal) streaming process that is affected by competition, which generates the final RAE. After the final RAE is created, the cycle begins all over again with a new initial RAE. In all periods from $t=2$ onward, the initial RAE, the streaming process, the final RAE and competition affect and are affected by social interactions. Note that the initial RAE and the final RAE might differ

¹² I produced this original flowchart based on the theories illustrated by articles from different disciplines.

because of the social interactions, even in the absence of a formal streaming process and competition.

The mechanism that leads professional athletes into the labor market is similar to the mechanism that leads high-skilled workers into the labor market. Although initially they differ somewhat, in the last stages they share a number of characteristics: in both education and sports, there is more or less formal streaming, participation in training / education is voluntary (e.g., in the last stages of secondary education and in all of tertiary education) and there is high competition (e.g., in education, there is competition for scholarships and for spots in programs with limited number of seats).

2.2 Evidence from Prior Literature

The short-run evidence on the RAE from education and sports can be reconciled. In education, for example, late-born children are more likely to be retained for an additional year in the same grade or to be assigned to remedial classes (Dixon, Horton, & Weir, 2011); they are more likely to be diagnosed with learning disability (Dhuey & Lipscomb, 2009); they are more likely to be diagnosed with attention-deficit/hyperactivity disorder and to be prescribed ad hoc stimulants (Zoëga et al., 2012); they are characterized by lower performance (Plug, 2001; Bedard & Dhuey, 2006; Ponzio & Scoppa, 2014),¹³ and they have a lower school attendance rate (Cobley et al., 2009). The sports context differs in terms of the type of evidence provided for the existence of the RAE. While in education RAE is predominantly measured in terms of actual performance, in sports it is measured in terms of representativeness. In fact, because of the tougher competition and the possibility of dropping out, early-born athletes in each age group are over-represented, and late-born athletes are under-represented with respect to the general population. This result is similar to that from the

¹³ Two articles find opposite results. Fredriksson and Öckert (2005) find that absolute age when starting school, in lieu of relative age, is responsible for different school performance. Cascio and Schanzenbach (2007) find that positive peer effects benefit relatively younger pupils.

education context: because the best performers continue practicing their sport (e.g., they do not drop out or are selected into higher tiers), and a larger percentage of these performers are born early in an admission year, it follows that relatively older children should on average outperform relatively younger children.

Conclusions regarding the long-run RAE are ambiguous in both education and sports, however. At the university level, the RAE might turn in favor of relatively younger students in terms of academic performance, although at the cost of lower social skills (Pellizzari & Billari, 2012).¹⁴ However, relatively younger students seem to earn a Ph.D. at the same age as relatively older students and seem to earn the same salary in postdoc positions (Kniffin & Hanks, 2013). In the general labor market, other studies provide evidence for a null RAE in terms of wages. Perhaps different performances reflect only chronological age differences (Larsen & Solli, 2012) so that overall there is a null RAE on life earnings. There might even be no wage gap at all if employers reward employees' productivity irrespective of their educational achievements but biased in favor of relatively older students (Crawford et al., 2013). Du et al. (2012) instead find a negative RAE in terms of representativeness in the labor market; they study a sample of the CEOs of the S&P 500 firms and find that relatively older CEOs are over-represented. Muller-Daumann and Page (2014) find an equivalent result among US congressmen. Finally, Black et al. (2011) and Plug (2001) find a wage gap in favor of relatively older workers. In sports, Ashworth and Heyndels (2007) find a reverse RAE in terms of wages, with relatively younger athletes receiving higher wages, and an RAE in terms of representativeness, with relatively older athletes being over-represented. Additionally, a reverse RAE in terms of representativeness among the very best hockey and soccer players has been found,¹⁵ with relatively younger players being over-represented (Gibbs et al., 2011; Bryson et al. 2014). Usually, however, over-representation of relatively older players is found

¹⁴ Examples of lower social skills are leadership skills (Dhuey & Lipscomb, 2008), self-esteem (Thompson et al., 2004), and less satisfactory social lives (Pellizzari & Billari, 2012).

¹⁵ Players selected for all-star teams and for Olympic team rosters in hockey and team captains in soccer.

among other professional athletes, for example, in soccer (e.g., Musch & Hay, 1999), tennis (Edgar & O'Donoghue, 2005), in both the summer and winter Olympic Games (Joyner, et al., 2013), and in the NFL (Böheim & Lackner, 2012).

In conclusion, on the one hand, the literature shows that in both contexts, the short-run RAEs on children's achievements are qualitatively similar. This finding comes as no surprise because the RAE is generated through similar mechanisms in sports and education. On the other hand, the evidence on long-run RAE is mixed in both contexts.

3 Institutional Context and Data

The empirical setting for our analysis is the Italian soccer major league, called Serie A. It is currently composed of 20 teams, but these teams do not play permanently in the major league because Italian soccer has a tiered structure, with promotions and relegations at the end of each season. The last three teams in the ranking are relegated to the second national division, that is, Serie B, which is composed of 22 teams, while the top three teams from this second league are promoted to Serie A.

In Italy, the age-group system for soccer is strictly regulated. January 1 is the relevant admission date applied to each age group, although specific age groups have slightly different rules. There are seven age groups in youth competitions; some are one-year age groups, while others are two-year age groups. In the latter case, children of different ages might play in separate games, despite training together, if the rules so specify. In general, children must train and play with their assigned age group.¹⁶ The minimum age requirement to play for a

¹⁶ The lowest age category is for children from 5 to 7 years of age; they are placed in the same age group, called "Piccoli Amici" (i.e., Small Friends), for both training and competition. In the next two categories, children of different ages are still grouped together for training, but they are divided based on year of birth for competitions. These categories are "Pulcini" (i.e., Chicks), for children under 11 years of age, and "Esordienti" (i.e., Newcomers), for children under 13 years of age. Up to three underage players may play in "Esordienti" matches. In the next categories, teenagers of different ages are placed together for both training and competition. These categories are "Giovanissimi" (i.e., Very Young), for players under 15 years of age; "Allievi" (i.e., Cadets), for players under 17 years of age; and finally "Primavera" (i.e., Spring), for players between 15 and 20 years of age. In all of these categories except the last one, no overage players are allowed; in the "Primavera", only one overage player per team may participate in the matches. The rules do not seem to set restrictions on whether

professional soccer team is 14 (art. 33, Internal Organization Rules FIGC);¹⁷ however, it is only possible to sign a contract with a team in a professional league at 16 years of age (art. 33, Internal Organization Rules FIGC).

The dataset contains information on players from seven Serie A seasons, 2007-08 to 2013-14. There are observations on 508 Italian soccer players who played for at least one Serie A team over the seven seasons of analysis.¹⁸ In total, the unbalanced panel data contain 1,704 Italian soccer-season observations. Most soccer players appear in our dataset for one or two seasons, 139 and 100 players respectively; 56 and 48 players are present for 3 and 4 seasons, respectively; 53 and 45 players are present for 5 and 6 seasons, respectively; and 45 players are present in all 7 seasons. Players may leave the dataset either permanently or temporarily: some players play for teams that are eventually relegated and may or may not be re-promoted to Serie A; players may be sold / lent to a Serie A team; players may be sold / lent to foreign teams or to teams in lower leagues and may or may not be transferred back to Serie A teams; and some players may retire.¹⁹

The empirical analyses use information on the players' wage, age, quarter and month of birth, current team, soccer season and role on the pitch.²⁰ The description of the variables and descriptive statistics are presented in Appendix A, Table A.1 and Table A.2, respectively.

Figure 2 illustrates the histogram for Italian players' birthrate per quarter.²¹ The division of the admission year into quartiles is a convention adopted within the relative age

children are free to train with an age group that is different from the one to which they are assigned and eventually to play official games with their assigned age group. More information on the Italian players' age-grouping system can be found on the official website of the Italian Football Game Federation (FIGC).

¹⁷ Norme Organizzative Interne Della FIGC.

¹⁸ The focus is on Italian players so as to analyze a set of players who have trained under the same admission date. Moreover, research on the admission dates in other countries is a complex task: the admission dates might differ between countries and within countries for different youth categories and different regions or states.

¹⁹ The impact of all of these player movements on the estimates of the RAE is not clear. Good Serie A players might move abroad because they are attracted by better contractual conditions and / or greater visibility. However, such movement may also occur for poorer performers who may want to play more often. This aspect is noted also by Ashworth and Heyndels (2007), who suggest that the number of foreign players in the domestic league might affect the estimate of the RAE through increased competition.

²⁰ The information on wages is obtained from annual reports compiled by Italy's sports-dedicated newspapers, whereas the other information is collected on sports-dedicated websites.

research (Wattie et al., 2015). The black rhombuses represent the average birthrate per quarter in the Italian population between 1993 and 1998; Italians' birthrates for previous years are unavailable. Appendix B reports the number of births per month and per year in the general population. This figure suggests the presence of an RAE in terms of representativeness, that is, the relatively younger players born toward the end of an admission year are under-represented, while relatively older players are over-represented. Moreover, there seems to be a specific trend: Serie A players' birthrate decreases with distance from the admission date.

Figure 2

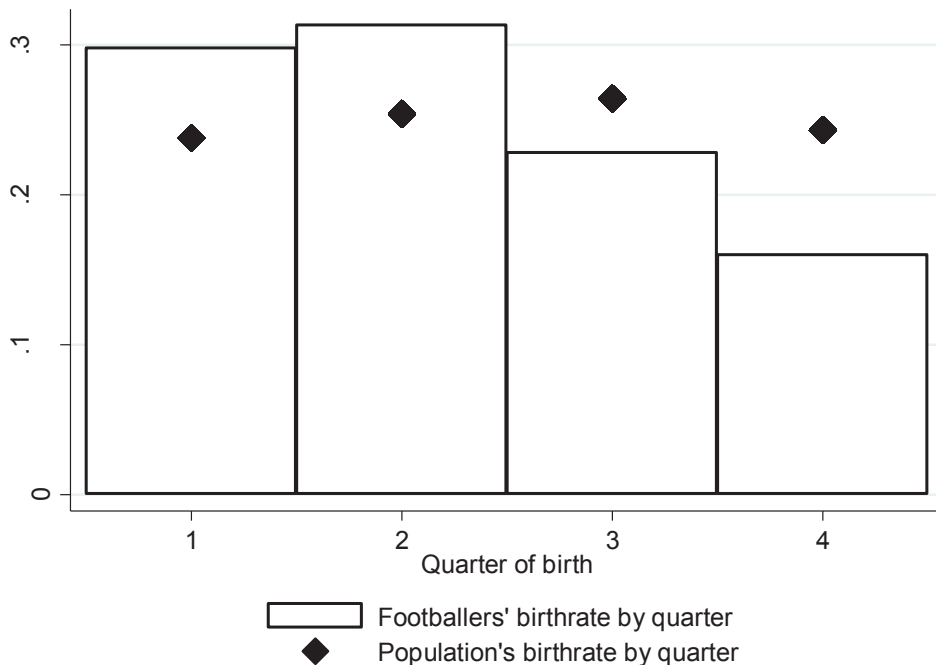
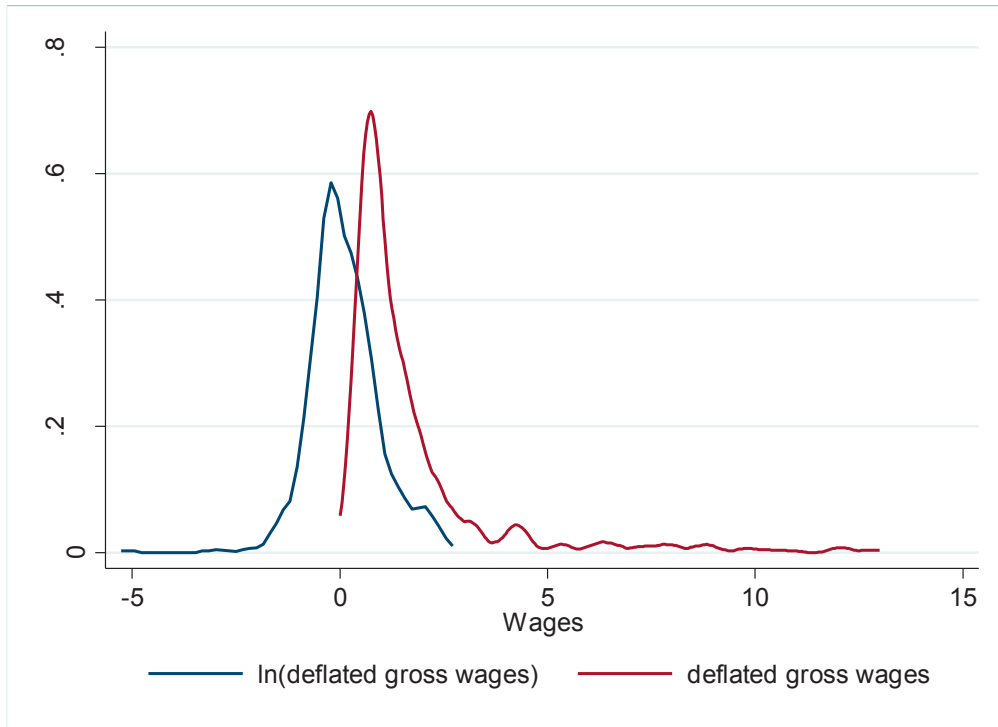


Figure 3 illustrates the players' wage distribution. The wages are measured before taxation and do not include bonuses, image rights or other deals, and they are deflated at the 2013

²¹ Where 1 is the quarter for January-March, 2 is the quarter for April-June, 3 is the quarter for July-September, and 4 is the quarter for October-December.

price level, with the annual coefficients provided by the Italian National Institute for Statistics (ISTAT).

Figure 3



The original distribution of the gross wages is highly unequal with a substantial positive skewedness, as expected in labor markets characterized by the presence of superstars (Lucifora & Simmons, 2003).²² The transformation of gross wages into natural logarithms returns a somewhat normal distribution.

Initial insights on possible wage gaps can be obtained by comparing the distribution of the gross wages for relatively older and younger players. Figure 4 compares the kernel density

²² Superstar is the term used to refer to extreme wage outliers (e.g., Bryson et al., 2014; Kleven et al., 2013; Lucifora & Simmons, 2003; Adler, 1985; Rosen, 1981). These outliers are such that in a labor market there appears to be a convex relationship between wages and skills (Lucifora & Simmons, 2003). There are two main competing, yet not mutually exclusive, superstar theories: Rosen (1981) suggests that superstars enjoy huge salaries because of a scarcity of talent, so that a little additional talent translates into large additional earnings, whereas Adler (1985) suggests that huge salaries are caused by positive network externalities that create additional popularity, even in the absence of talent.

distributions of the gross wages for Italian players divided by quarter of birth. Because players can change teams during the season, they are assigned the gross wage they receive from the team with which they start the season.²³

Figure 4

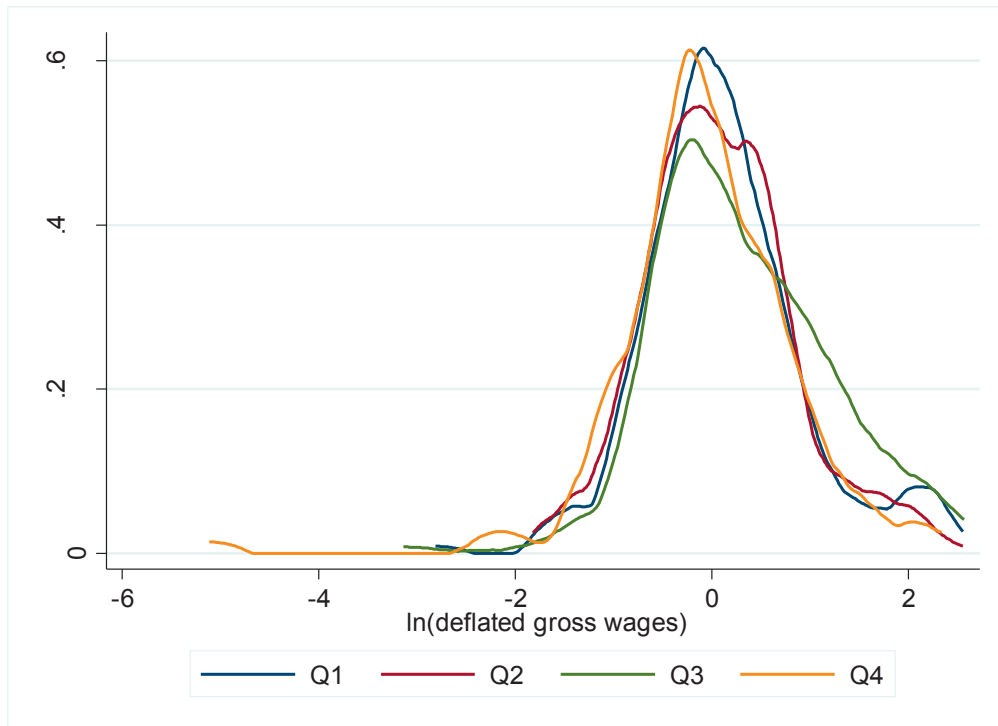
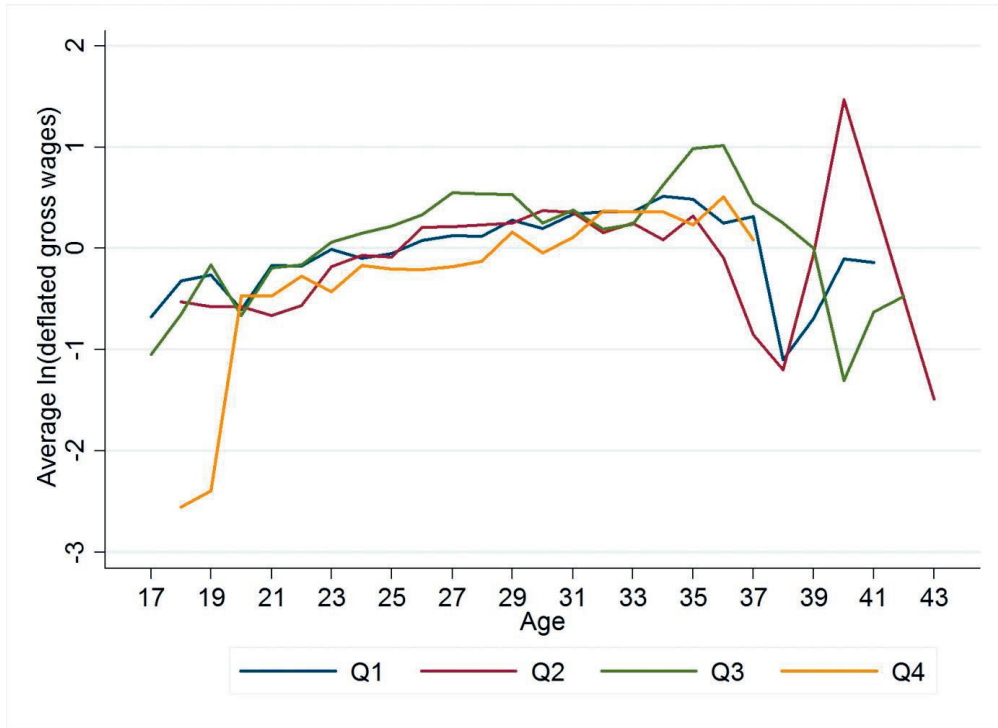


Figure 4 shows that the wage distribution of players born in the last quarter of the year (yellow line, Q4) has a longer and thicker left tail than the distributions for players born in the other quarters. This result suggests that there might be a wage gap penalizing players born in the last quarter. A somewhat puzzling result is that players born in the third quarter (green line, Q3) tend to more frequently earn wages in the top percentiles of the wage distribution.

²³ Teams trade players during two main market sessions: in the summer, which separates different soccer seasons, and during the Christmas break and January, which is toward the end of the first half of the season. Players who change teams during the latter session and come from another Serie A team are assigned the gross wage that they received from the team with which they started the season. Players who join a Serie A team during the Christmas break and come either from abroad or from Serie B—or from other lower domestic leagues—are assigned the new wage.

Additional insights on the nature of the wage gap might be gained with the investigation of its dynamics through players' careers. For this purpose, Figure 5 plots the average natural logarithm of the gross wages against age for the four groups of players.

Figure 5



For players born in the last quarter of the year (yellow line, Q4), there is an important entry wage gap that disappears in the early twenties; afterward, a wage gap again appears that then disappears only toward the end of a player's career. The entry wage gap is smaller for players born in the third quarter (green line Q3);²⁴ these players also seem to enjoy higher wages in the prime of their career and toward its end. At approximately 40 years of age, a gap in favor of players born in the first quarter (red line, Q1) appears. This figure should be considered

²⁴ The entry wage gap might be explained through a physical development gap that disappears at 20 years of age (see the WHO growth charts for children from 5 to 19 years of age on the WHO website and US growth charts in Kuczarski et al, 2002), while the cognitive development gap disappears between 20 and 25 years of age (Salthouse et al., 2004).

carefully; the graph illustrates averages, and there are few observations for extreme ages,²⁵ so outliers may drive the results, which is particularly important given the presence of superstars.

In conclusion, a visual inspection of the data suggests that players born early in an admission year are over-represented and that the players' birthrate by quarter decreases with distance from the admission date. The visual inspection also suggests that players born in the last quarter of an admission year receive a lower entry wage than players born in other quarters; they also earn lower wages throughout their career.

4 Methods and Results

4.1 RAE in Terms of Representativeness

Regarding the presence of RAE in terms of representativeness, the observed distribution of the quarter of birthrate should differ from the expected distribution. Players born at the beginning of an admission year should be over-represented, while players born at the end of an admission year should be under-represented. Moreover, there should be a specific birthrate trend: the birthrate should decrease with the distance from the admission date.

A chi-square goodness-of-fit test is used to compare the difference between the observed and expected number of players across birth quarters (e.g., Sims & Addona, 2014; Helsen et al., 2012), and the observations from the seven seasons are pooled. The expected number of players is based on the average birthrate by quarter in Italy between 1993 and 1998; data for previous years are unavailable. A uniform birthrate distribution is not assumed because of the seasonality of birth demonstrated for the general Italian population (e.g., Rizzi & Dalla Zuanna, 2007; Prioux, 1988). Table 1 shows the results from this analysis.

²⁵ There are 27 player-season observations at 18 years of age, or less, and 33 player-season observations at 37 years of age, or more.

Table 1. Chi-squared goodness-of-fit by quarter of birth.

Quarter	Observed counts	Expected counts	Difference
Q1 (January-March)	508	406.989	101.011
Q2 (April-June)	534	433.769	100.231
Q3 (July-September)	389	450.194	-61.194
Q4 (October-December)	273	414.048	-141.048
χ^2 (3)		105.813	
P-value		0.000	

Note: “Expected counts” is the expected number of players born in each quarter. “Observed counts” is the observed number of players born per quarter. “Difference” provides the differences between the observed and expected counts, which are used to compute the chi-squared statistics.

The table confirms the insights provided by the descriptive statistics: Serie A is characterized by the presence of RAE in terms of representativeness. The distribution of the observed quarters of birthrate for Italian Serie A players is statistically significantly different from its expected distribution. This result is in line with other studies that analyze the RAE in Serie A (e.g., Salinero, Pérez, Burillo & Lesma, 2013; Helsen et al., 2012).

Furthermore, the column “Difference” suggests the existence of a specific trend in the players’ birthrate by quarter. In fact, in the presence of the RAE, players born at the beginning of an admission year are over-represented, and the extent of this over-representation decreases with distance from the admission month. The trend then turns into under-representation, which increases moving toward the end of the admission year. The formal analysis on the existence of this specific birthrate trend is implemented using the Spearman-rank correlation coefficient (for a similar application of this test, see Musch & Hay, 1999, and Ashworth & Heyndels, 2007).

The Spearman-rank correlation coefficient is computed between two measurement variables that are converted to ranks (McDonald, 2014). One variable is the *months’ representativeness* in Serie A, and it is based on the differences between the expected number of players based on the Italian population’s birthrate by month and the observed number of players born in each month. When the difference between the expected and the observed

number of observations is negative, the players born in that month are over-represented; vice-versa, a positive difference signals under-representation. First place in the ranking is assigned to the most under-represented month, while the last place is assigned to the most over-represented month. The measurement variable represents the *admission date distance*, which is based on the distance of the month from the admission date (i.e., January has the first position in the ranking, whereas December has the last position); it measures the RAE.

If the Spearman rank-correlation coefficient was simply computed between the ranking based on the observed counts—in lieu of the differences between the expected and observed number of players—and the admission date distance, the possible presence of trends in the birthrate by month for the underlying general population would not be taken into account. If in the general population the birthrate by month increased with distance from the admission date, the results could provide artifactual evidence of the RAE.

How to interpret the output of the Spearman-rank correlation coefficient test? On one hand, if the estimate of the correlation was negative and statistically significant, there would be evidence of a specific trend characterizing the RAE, that is, players born in early months would be over-represented and, with an increase in the distance of the month of birth from the admission date, the over-representation would decrease until eventually players born toward the end of the year would be under-represented. On the other hand, if the correlation was positive and statistically significant, the trend would proceed in the opposite direction, and there would be evidence of a reverse RAE. In this case, we would observe under-representation in the early months, and the birthrate by month would tend to increase with distance from the admission date, such that players born toward the end of the year would be over-represented. Based on the previously discussed RAE mechanism, we expect to find a negative correlation between the order of the months and admission date distance. Thus, the

H_0 of the Spearman-rank correlation coefficient test is that the correlation between the two rankings is zero.²⁶

Table 2. Correlation between months' representativeness and admission date distance.

Month	Months' representativeness				Admission date distance
	Expected counts	Actual counts	Difference	Ranking (1)	Ranking (2)
January	138.826	234	-95.174	12	1
February	126.411	120	7.411	7	2
March	140.188	154	-13.812	9	3
April	138.077	139	-0.923	8	4
May	149.192	206	-56.808	11	5
June	145.925	189	-43.075	10	6
July	154.599	138	16.599	5	7
August	146.515	136	10.515	6	8
September	149.011	115	34.011	3	9
October	144.797	111	33.797	4	10
November	132.585	89	43.585	2	11
December	136.873	73	63.873	1	12
Spearman	-0.860				
	(0.000)				

Note: The shaded areas include the figures of interest. "Expected counts" is the expected number of players born in each month, based on the monthly average Italian birthrate from 1993 to 1998. "Actual counts" is the observed number of players born per month. "Difference" provides the differences between observed and expected counts. "Ranking (1)" is the month ranking based on the differences between the expected and actual counts: first place is assigned to the most under-represented month, i.e., the month with the largest positive difference; last place is assigned to the most over-represented month, i.e., the month with the largest negative difference. "Ranking (2)" is the distance of the month of birth from the admission date; this ranking is directly established by the "Admission date distance;" for example, January, February and March receive ranks 1, 2, 3. "Spearman" is the estimate of the Spearman-rank correlation coefficient; the corresponding P-value is in parentheses.

Table 2 presents a highly statistically significant and negative correlation between months' representativeness and admission date distance, providing evidence for a downward trend in the soccer players' birthrate. This result reinforces the evidence of an RAE in terms of representativeness.

²⁶ The usage of the two-tailed test is motivated by the possibility of observing a reverse RAE as well.

4.2 RAE in Terms of Wages

The sign of the RAE in terms of wage gaps might differ from the sign of the RAE in terms of representativeness. It is possible to have an RAE in terms of representativeness yet a reverse RAE in terms of wages. Relatively younger players enjoy higher, equal or lower wages depending on several characteristics of the streaming process. As discussed in the introduction, relatively younger players might have been exposed to positive peer effects or might have been positively selected (e.g., Ashworth & Heyndels, 2007; Gibbs et al., 2012; Bryson et al., 2014), which eventually benefits them in terms of wages. Alternatively, it is possible that path dependence might have increased original performance differences (Allen & Barnsley, 1993), which would eventually disadvantage younger players in terms of wages. Finally, discriminatory criteria that cause original differences in achievement might cease to be relevant in the labor market (Crawford et al., 2013); thus, in this case, there could be an equalization of wages. The analysis in this section focuses on the empirical sign of the RAE in terms of wages.

The descriptive statistics suggest that players born in the last quarter receive gross wages in the bottom percentiles of the wage distribution more frequently than players born in other quarters. However, other characteristics that determine wages are not controlled for, so it is not possible to obtain clear insights.

The empirical investigation of wage gaps proceeds with a standard methodology used in economics of sports: the pooled OLS regression.²⁷ The model is the following:

$$\ln(w_i) = \beta_0 + \beta_1 age_i + \beta_2 age_i^2 + \beta RA_i + \varepsilon_i$$

²⁷ See, for instance, Bryson et al. (2014) and Frank and Nüesch (2012).

The natural logarithm of the deflated gross wage for player i is the outcome variable. The set of control variables includes age_i and age_i^2 , both of which refer to player i 's age; age is computed in years instead of in months as in Ponzio and Scoppa (2014) because there is no information on the date of signing the contract. This variable is rescaled; it ranges from 0 to 26 because in our dataset, the minimum registered age for Italian players is 17, whereas the maximum is 43, so the estimate of the constant is directly interpretable. The variable for squared age captures the decreasing return to age. \mathbf{RA}_i is a vector of dummy variables for quarter of birth, where the first quarter of the admission year (i.e., January to March) is the reference quarter. The estimated coefficients represent the estimates of the RAE for different quarters and reflect differences in both sheer maturity and productivity. The RAE is unbiased if two assumptions hold true: i) date of birth is unrelated to innate ability; this assumption is also called the “nonastrology assumption” (Allen & Barnsley, 1993, p. 654); and ii) season-of-birth is unrelated to players' performance, which also implies the absence of a relationship between household socioeconomic status and birthdate.²⁸ Finally, some may argue that the estimates capture the combined RAE from sports and school because soccer and school admission years overlap; however, previous literature suggests that educational achievements do not affect returns from playing soccer at the professional level (Barros, 2001).

As Ashworth and Heyndels (2007) explain, the inclusion of other variables that are normally used in studies of economics of sports—for instance, measures of players' performance—would cause multicollinearity. For the same reason, controls for players'

²⁸ The existence of a correlation between date of birth and socioeconomic status seems to differ by country. For instance, some studies suggest the presence of such a correlation in the US (Bound & Jaeger, 2001; Cascio & Lewis, 2005; Buckels & Hungerman, 2013), Sweden (Carlsson et al., forthcoming), and Austria (Doblhammer & Vaupel, 2001). However, a recent study suggests that in Norway there is no such correlation (Black et al., 2011). Moreover, for Italy, this correlation has not been systematically analyzed, but a previous study on RAE on school performance suggests that it is non-existent (Ponzio & Scoppa, 2014). Bound and Jaeger (2001) note the presence of additional season-of-birth effects on health, e.g., mental health, but these results can be ignored because they involve only a small number of individuals among the whole population (Plug, 2001).

experience are not included. There is no problem of collinearity with age,²⁹ but there might be collinearity with the RA variable: in the presence of an RAE, relatively older players also benefit from more playing time, which increases their experience more rapidly.³⁰

As a robustness check, the analyses are re-run while adding variables for specific effects on wage heterogeneity. There is one vector for teams and one for season fixed effects.³¹ The favorite model is that including all of the fixed effects.

As an additional robustness check, the analyses are re-run on a discontinuity sample (e.g., Ponzio & Scoppa, 2014; Black et al. 2011). This strategy consists of focusing on the narrower sample of footballers born in either January or December; these are two adjacent months, with one after the admission date (January 1) and the other immediately before it. As suggested by Barnsley and Thompson (1988), in the analysis of a discontinuity sample, season-of-birth effects should be eliminated because the two months are in the same season. All of the fixed effects are included in this analysis.

Because repeated observations of individual players are not likely to be independent, standard errors are clustered on footballers in all of the analyses.

Although all of the estimates are reported, the focus of the analyses is on the comparison between footballers born in the fourth and first quarters, or in December and January.

²⁹ In European sports, athletes may enter professional competitions at different ages. Therefore, the usage of both age and experience does not create multicollinearity. Conversely, in studies of US sports, the introduction of both variables would create multicollinearity because the draft system is such that athletes enter professional leagues at a somewhat uniform age (Lucifora & Simmons, 2003).

³⁰ Ashworth and Heyndels (2007) do not find such a correlation, so they also include experience in the model.

³¹ This paper cannot investigate the effect of the increased competition from foreign players because the number of foreign players changes constantly throughout the seven seasons under examination; hence, season fixed effects would be perfectly collinear with this measure of competition.

The estimates and the robustness checks are reported in Table 3.

Table 3. Wage gap between relatively older and younger players.

Variables	(1)	(2)	(3)	(4)
Age	0.227*** (0.039)	0.225*** (0.039)	0.224*** (0.028)	0.270** (0.105)
Age ²	-0.008*** (0.002)	-0.008*** (0.002)	-0.008*** (0.001)	-0.010** (0.004)
Q2 (April-June)	-0.062 (0.093)	-0.061 (0.093)	-0.053 (0.055)	-
Q3 (July-September)	0.181 (0.114)	0.179 (0.114)	0.045 (0.069)	-
Q4 (October-December)	-0.203* (0.114)	-0.207* (0.114)	-0.177** (0.0752)	-
December	-	-	-	-0.369* (0.186)
Constant	-1.284*** (0.225)	-1.215*** (0.249)	-1.366*** (0.215)	-1.841** (0.732)
F-test, quarters of birth (p-value)	0.022	0.022	0.033	-
Season F.E.	N	Y	Y	Y
Team F.E.	N	N	Y	Y
R-square	0.142	0.151	0.560	0.658
Observations	1,598	1,598	1,598	282

Note: Standard errors clustered on footballers are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Column (1) reports the results from the OLS without fixed effects; column (2) reports the results from the OLS with season fixed effects; column (3) reports the results from the OLS with players' team and season fixed effects. The sample size for the analyses in column (1)-(2)-(3) is smaller than 1,704 observations because of missing values for wages. Column (4) reports the results obtained with the discontinuity sample, i.e., only players born in January or December; in addition, this analysis is conducted with OLS and includes players' team and season fixed effects. "F-test, quarters of birth" provides the p-value from the F-test on the joint significance of the quarters of birth estimates.

This table provides evidence for an RAE on wage gaps. The results displayed in column (3) provide statistically significant evidence that, ceteris paribus, a player born in the fourth quarter of the admission year receives a wage that is approximately 19% lower than that earned by footballers born in the first quarter.³² The p-value from the F-test on the joint significance of the estimates for the quarter of birth coefficients suggests that these three coefficients considered together are statistically significant. The estimates in column (4) are

³² The wage gap in percentage terms is computed as $[\exp(0.177)-1]*100= 19.3\%$.

obtained from the analysis of the discontinuity sample; an even larger and statistically significant wage gap is obtained.

These results are robust to two robustness checks. First, the results are confirmed after winsorization at 3% and 97%; see Appendix C, Table C.1. This analysis is conducted to verify whether the evidence of an RAE on wages is led by wage outliers. Second, the results are also confirmed when players in their first season in the dataset and players who have changed teams are dropped from the dataset. This robustness check is conducted because the signing date of the contract is not known, and it is therefore not possible to distinguish between players whose first wage refers to the whole season or only to a few months (e.g., a player signs a contract for a new team in January, so his first season wage could refer only to the period from January to the end of the season, which is May or June).

5 Heterogeneity Analyses on Wage Gaps

In this section, heterogeneity analyses on the RAE in terms of wage gaps are examined. The wage gap might differ across ages as the sheer maturity differential decreases³³ and additional player selections occur throughout one's career.³⁴ Additionally, the wage gap might differ across quantiles of the wage distribution; such differences might be particularly important in light of the positive skewedness of the wage distribution. While the main analysis conducted with OLS focuses on the sign of the RAE in terms of wages, these additional analyses might provide additional insights.

Does the wage gap change over footballers' careers? Figure 5 suggests that the answer might be affirmative. To formally investigate this research question, the analyses are re-run for different age categories: footballers younger than 21 (players who can still compete in the last youth category, "Primavera," and might present physical development differentials),

³³ The introduction notes that there might still be a maturity differential due to RA until players' mid-twenties.

³⁴ Some players leave Serie A (permanently or temporarily) during their career for different reasons; other players can start their careers in Serie A when they are older. These changes might alter the wage gap at different ages; the direction of this change depends on the performance of the players who leave or join Serie A.

footballers between 21 and 25 (players for whom complete cognitive maturity is still to be reached), between 26 and 30 (career prime), and older than 30 (retirement period). This approach of using age category to study the evolution of RAE is also used by Black et al. (2011).³⁵ For brevity, only the estimates obtained using season and players' team fixed effects are reported in Table 4.

Table 4. Wage gap between relatively older and younger players, by age group.

Variables	< 21	21-25	26-30	> 30
	(1)	(2)	(3)	(4)
Age	0.825 (0.665)	0.181 (0.166)	0.206 (0.232)	0.330** (0.143)
Age ²	-0.049 (0.183)	-0.059 (0.013)	-0.007 (0.010)	-0.011*** (0.004)
Q2 (April-June)	-0.344 (0.608)	-0.092 (0.084)	-0.038 (0.069)	-0.099 (0.089)
Q3 (July-September)	-0.866* (0.509)	-0.013 (0.107)	0.076 (0.093)	-0.092 (0.114)
Q4 (October-December)	-1.129 (0.906)	-0.195* (0.108)	-0.240*** (0.092)	-0.134 (0.110)
Constant	-3.517** (1.402)	-1.011** (0.532)	-1.357 (1.249)	-1.975 (1.238)
F-test, quarters of birth (p-value)	0.291	0.223	0.018	0.213
Season F.E.	Y	Y	Y	Y
Team F.E.	Y	Y	Y	Y
R-square	0.841	0.491	0.550	0.635
Observations	51	403	629	515

Note: Standard errors clustered on players are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Column (1) reports the results from the OLS with players' team and season fixed effects for players younger than 21 years of age; column (2) reports the results from the OLS with players' team and season fixed effects for players between 21 and 25 years of age; column (3) reports the results from the OLS with players' team and season fixed effects for players between 26 and 30 years of age; and column (4) reports the results from the OLS with players' team and season fixed effects for players older than 30 years of age. "F-test, quarters of birth" provides the p-value from the F-test on the joint significance of the quarters of birth estimates.

These results confirm the presence of an RAE in terms of wage gaps and suggest that wage gaps might evolve during players' careers. The estimates for players born in the second and third quarters confirm the results of the main analysis. The players born in these two quarters

³⁵ The difference with Black et al. (2011) is that those authors use single ages from 24 to 35, whereas here, four age groups are used; the choice of multiple age categories is due to the sample size.

do not suffer from any statistically significant wage gap at any age group. The p-value from the F-test on the joint significance of the estimates for the quarter of birth coefficients suggests that these three coefficients considered together are statistically significant only in the age group between 26 and 30 (career prime). These results are also similar after winsorization at 3% and 97%; see Appendix C, Table C.2.

Although the dataset does not allow for further investigation of the wage gap dynamics through players' careers, it is possible to interpret these results based on the existing literature. Players born in the fourth quarter suffer from the largest wage gap when they are under 21 years of age; this estimate is not statistically significant (there are a very limited number of observations for this age group). Once in the labor market, the performance gap between these new, relatively younger players and their older peers might be relevant. Therefore, between 21 and 25 years of age, the worst performers among the relatively younger players might decide to leave Serie A to gain experience in lower leagues or abroad, and consequently the wage gap decreases; however, this gap is not statistically significant. During this period away from Serie A, the part of the RAE due only to the sheer maturity differential is eliminated.³⁶ When these players re-enter Serie A between 26 and 30 years of age,³⁷ they are still worse performers than their older peers on average: the sheer maturity gap is now filled, but tangible and intangible skills might still differ because of the RAE. Therefore, the wage gap increases again,³⁸ even if it is now smaller compared with the entry wage gap that they suffered at the beginning of their careers. Finally, after 30 years of age, the wage gap might widen even more because the characteristics of the players who decide to retire might differ by quarter of birth. For example, if the best players among those who are born in the first quarter tend to retire late, while the best players from the fourth quarter retire

³⁶ Note that the same occurs if the best players among the relatively older players leave Serie A to gain more experience.

³⁷ Similarly, Parent-Harvey et al. (2013) and Böheim and Lackner (2012) suggest that relatively younger athletes might delay entry into professional sports to wait for the maturity gap to be filled.

³⁸ The wage gap might also increase because the best relatively older players who left Serie A are now back.

early, the average performance gap diverges, even if on average these two groups of players retire at the same age.

Does the wage gap differ over different quantiles of the wage distribution? This question makes empirical sense. Because the distribution of players' wages is positively skewed and is characterized by the presence of superstars, the normality assumption does not hold. Moreover, the analysis at the median, that is, the 50th percentile of the wage distribution, is more robust to wage outliers—including superstars—than the OLS. The OLS estimates describe the relationship between the regressors and the conditional mean of the outcome variable, whereas the quantile regression describes the relationship between the regressors and the conditional quantile of the outcome variable; therefore, the quantile regression provides a more comprehensive picture of the relationship. The results from the quantile regression also help to determine whether there are positive peer effects for and selections of relatively younger players; this aspect is illustrated below and is important to understanding the RAE mechanism. To the best of my knowledge, this study represents the first time that the quantile regression has been used to analyze the RAE on wages. The analyses are implemented at the 25th, 50th, 75th and 90th percentiles of the wage distribution.³⁹ For brevity, only the estimates obtained using season and players' team fixed effects are reported in Table 5.

³⁹ Because repeated observations on individual players are not likely to be independent, standard errors are clustered on players. The employed method is that suggested by Parente and Silva (2013).

Table 4. Wage gap between relatively older and younger players.
Quantile regression at 25%, 50%, 75% and 90% of the wage distribution.

Variables	25%	50%	75%	90%
	(1)	(2)	(3)	(4)
Age	0.258*** (0.030)	0.193*** (0.032)	0.163*** (0.027)	0.132*** (0.036)
Age ²	-0.009*** (0.001)	-0.007*** (0.001)	-0.006*** (0.001)	-0.004*** (0.001)
Q2 (April-June)	-0.029 (0.055)	-0.030 (0.052)	-0.057 (0.078)	-0.172** (0.072)
Q3 (July-September)	0.025 (0.066)	0.061 (0.071)	0.096 (0.103)	0.067 (0.072)
Q4 (October-December)	-0.128 (0.085)	-0.097 (0.063)	-0.113 (0.091)	-0.158 (0.109)
Constant	-2.055*** (0.349)	-1.222*** (0.227)	-0.355 (0.221)	1.096*** (0.242)
F-test, quarters of birth (p-value)	0.317	0.172	0.310	0.000
Season F.E.	Y	Y	Y	Y
Team F.E.	Y	Y	Y	Y
Pseudo R-square	0.537	0.544	0.538	0.505
Observations	1,598	1,598	1,598	1,598

Note: Standard errors clustered on players are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The samples for the analyses use less than 1,704 observations because of missing values for wages. Column (1) reports the results from the quantile regression at the 25th percentile of the wage distribution; column (2) reports the results from the quantile regression with players' team and season fixed effects at the 50th percentile of the wage distribution; column (3) reports the results from the quantile regression with players' team and season fixed effects at the 75th percentile of the wage distribution; and column (4) reports the results from the quantile regression with players' team and season fixed effects at the 90th percentile of the wage distribution. "F-test, quarters of birth" provides the p-value from the F-test on the joint significance of the quarters of birth estimates.

Two important aspects of Table 5 deserve some attention. On the one hand, the analysis at the median, column (3), reports a quite lower wage gap than that obtained using the OLS for players born in the fourth quarter.⁴⁰ On the other hand, the wage gap seems to increase in the quantile of the wage distribution. If these results were to be taken seriously, there would be two implications. First, there might not be positive peer effects and selection for relatively younger Serie A players; otherwise, a reduction of the wage gap, or even its reversal, would have been observed in the top quantiles of the wage distribution rather than its increase.⁴¹

⁴⁰ Interpretation of the estimated coefficients is similar to that for OLS estimates (Hao & Naiman, 2007).

⁴¹ Gibbs et al. (2011) show that the RAE in terms of representativeness could reverse among the very best players; therefore, a reversal of the RAE in terms of wages among the very best players is also plausible.

Second, the RAE on wages seems to be driven by the effect on wages in the top quantiles of the distribution; this tendency is not captured by the analysis with winsorized wages (see Table C.1), which is a methodology that is frequently adopted for robustness checks.

It is worth mentioning that the RAE might also differ based on players' roles. The different effects might depend on at least three factors: i) innate ability might be more important and to some extent compensate for differences in maturity at an early age for some roles, for instance, for forwards and their "tor-instinct" (Ashworth & Heyndels, 2007, p. 368); ii) the identification of role-specific skills might be easier for some roles, for instance, for forwards (Ashworth & Heyndels, 2007); and iii) the importance of physical characteristics is greater for goalkeepers and defenders (Salinero et al., 2013). Because the reasons that lead to possible differences in the RAE by players' role are peculiar to sports disciplines, the results are reported and commented on more extensively in Appendix D. In summary, the RAE seems to be larger for goalkeepers and midfielders, while it seems to be non-existent for forwards.

6 Discussion and Conclusions

Chronological differences between individuals within the same age group, that is, relative age (RA), determine maturity gaps during childhood, both in school and in sports contexts. These differences translate into a performance gap (e.g., Musch & Hay, 1999; Musch & Grondin, 2001; Bedard & Dhuey, 2006; Dhuey & Lipscomb, 2009), that is, the relative age effect (RAE), which should disappear over time. However, because of streaming, competition and social interactions, the performance gap might extend to the long run (e.g., Allen & Barnsley, 1993; Bedard & Dhuey, 2006), thus affecting labor market outcomes.

This paper focuses on two aspects of the RAE on the labor market for a particular group of high-skilled workers: professional soccer players. First, it investigates the existence of the RAE in terms of representativeness, that is, whether players born at the beginning of an

admission year are over-represented. Second, it investigates the presence of the RAE in terms of wage gaps. Heterogeneity analyses on the evolution of this possible wage gap across ages and quantiles of the wage distribution are also carried out.

This study provides statistically significant evidence for the RAE in terms of representativeness, with relatively older players being over-represented in Serie A (e.g., Barnsley & Thompson, 1988; Musch & Hay, 1999; Musch & Grondin, 2001; Böheim & Lackner, 2012). The analyses also suggest the existence of a specific trend in the birthrate distribution, further reinforcing the evidence of the RAE: the over-representation decreases and turns into under-representation as the end of the admission year is approached.

This paper also provides statistically significant evidence of the RAE on players' wages, with relatively younger players, that is, players born late in an admission year, earning on average lower gross wages. The size of the wage gap caused by the RAE is economically important; in particular, it is important for players born in December, which is the last month of the admission year. Furthermore, this wage gap appears to increase with age after an initial reduction. I speculate that this trend could be a result of players' mobility into and out of Serie A, which affects the distribution of the characteristics of the players who do not move. The analyses on different wage quantiles suggest that wage gaps tend to increase in the quantiles of the wage distribution; therefore, the estimates obtained with the OLS are driven by differences in the top quantiles of the wage distribution. These results seem to rule out the presence in Serie A of positive peer effects and selection in favor of relatively younger players.

These analyses should be considered carefully. Because the admission date for the players in this sample does not vary, it is not possible to definitively rule out any season-of-birth effect on wages.

Additionally, these results on wage gaps are somewhat different from those in the only other study on the RAE in the soccer player labor market, which examined the case of the German major league (Ashworth & Heyndels, 2007). However, this difference might be due to the type of data and the different characteristics of the Italian and German institutional contexts.⁴²

Future analyses of the RAE on wages should include a quantile regression. This approach is particularly important when the wage distribution is expected to be positively skewed, for instance, in the case of studies on the RAE among CEOs (e.g., Malmendier & Tate, 2009; Frank & Nüesch, 2012). The use of a quantile regression is also important to investigate positive peer effects and selection, which might characterize relatively younger workers.

Further work is required to increase the knowledge on the short- and long-run RAE. Future studies on wage gaps should exploit variations in the admission date to obtain clearer evidence of the RAE. This variation can be obtained in cross-country analyses, similar to the studies by Bedard and Dhuey (2006) and Munch and Hay (1999). Moreover, different aspects, such as career promotions, retirement and migration decisions, have yet to be analyzed.

Additional evidence in favor of the RAE would call for a revision of the age-grouping system. Age groups could be shortened, for instance, to 6 or 9 months (e.g., Pellizzari & Billari, 2012; Barnsley & Thompson, 1988) instead of 12 months, so that within-age-group performance differences would be reduced, potentially decreasing the wage gaps in the labor market. Alternatively, the admission date could rotate in different ways so that the RAE

⁴² The Italian and German youth category systems might be different. Moreover, Ashworth and Heyndels (2007) use data from 1997-98 and 1998-99, immediately after the Bosman ruling in 1995 imposed by the European Court of Justice. This ruling affected the soccer players' labor market because it banned limitations on the number of players from EU countries and introduced free agency. The data used in this paper are almost a decade more recent, so a longer period has passed, allowing the Bosman ruling to fully affect the soccer player labor market. During the seven seasons of the analysis, the number of foreign players constantly increased, although until 2008, the number of foreign players in Italy was among the lowest in EU (Bryson et al., 2014). Because the Bosman ruling affects players' competition, it could also affect the RAE on wages; however, the expected sign of this effect is uncertain. It depends on many factors, such as the top tax rates, which determine the quality of the players who move (Kleven et al., 2013).

would not consistently provide advantages to people born in a given month (Barnsley & Thompson, 1988; Wattie et. al, 2015).

Bibliography

- Adler, M., (1985). Stardom and talent. *The American Economic Review.* , 75(1), 208-212.
- Allen, J., & Barnsley, R. (1993). Streams and Tiers: The Interaction of Ability, Maturity, and Training in Systems with Age-Dependent Recursive Selection. *Journal of Human Resources*, 28(3), 649-659.
- Angrist, J. D., & Krueger, A. B. (1992). The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples. *Journal of the American Statistical Association*, 87(418), 328-336.
- Ashworth, J., & Heyndels, B. (2007). Selection Bias and Peer Effects in Team Sports The Effect of Age Grouping on Earnings of German Soccer Players. *Journal of Sports Economics*, 8(4), 355-377.
- Barnsley, R. H., & Thompson, A. H. (1988). Birthdate and Success in Minor Hockey: The Key to the NHL. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 20(2), 167-176.
- Barros, C. P. (2001). Economic Return on Schooling for Soccer Players. *Journal of Sports Economics*, 2(4), 369-378.
- Bedard, K., & Dhuey, E. (2006). The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects. *The Quarterly Journal of Economics*, 121(4), 1437-1472.
- Black, S. E., Devereux, P. J., & Salvanes, K. G. (2011). Too Young to Leave the Nest? The Effects of School Starting Age. *The Review of Economics and Statistics*, 93(2), 455-467.
- Böheim, E., & Lackner, M. (2012). Returns to Education in Professional Football. *Economics Letters*, 114(3), 326-328.

Bound, J., & Jaeger, D. A. (2001). Do Compulsory School Attendance Laws Alone Explain the Association Between Quarter of Birth and Earnings? *Research in Labor Economics*, 19, 83-108.

Bryson, A., Rossi, G., & Simmons, R. (2014). The Migrant Wage Premium in Professional Football: A Superstar Effect? *Kyklos*, 67(1), 12-28.

Bryson, A., Gomez, R., & Zhang, T. (2014). All-star or benchwarmer? Relative age, cohort size and career success in the NHL. *IZA Discussion Paper*, 8645.

Buckles, K. S., & Hungerman, D. M. (2013). Season of Birth and Later Outcomes: Old Questions, New Answers. *Review of Economics and Statistics*, 95(3), 711-724.

Carling, C., le Gall, F., Reilly, T., & Williams, A. M. (2009). Do Anthropometric and Fitness Characteristics Vary According to Birth Date Distribution in Elite Youth Academy Soccer Players? *Scandinavian Journal of Medicine & Science in Sports*. 19(1), 3-9.

Carlsson, M., Dahl, G., Öckert, B., & Rooth, D.-O. (forthcoming). The Effect of Schooling on Cognitive Skills. *The Review of Economics and Statistics*.

Cascio, E. U., & Lewis, E. G. (2006). Schooling and the Armed Forces Qualifying Test: Evidence from School-Entry Laws. *The Journal of Human Resources*, 41(2), 294-318.

Cascio, E. U., & Schanzenbach, D. (2007). First in the Class? Age and the Education Production Function. *Discussion Paper 13663, NBER*.

Cobley, S., McKenna, J., Baker, J., & Wattie, N. (2009). How Pervasive are Relative Age Effects in Secondary School Education? *Journal of Educational Psychology*, 101(2), 520-528.

Crawford, C., Dearden, L., & Greaves, E. (2013). *The Impact of Age Within Academic Year on Adult Outcomes*. Institute for Fiscal Studies, 1-46.

Dixon, J., Horton, S., & Weir, P. (2011). Relative Age Effects: Implications for Leadership Development. *International Journal of Sport & Society*, 2(2), 1-15.

Dhuey, E., & Lipscomb, S. (2009). What makes a leader? Relative Age and High School Leadership. *Economics of Education Review*, 27(2), 173-183.

Doblhammer, G., & Vaupel, J. W. (2001). Lifespan Depends on Month of Birth. *Proceedings of the National Academy of Sciences*, 98(5), 2934-2939.

Du, Q., Gao, H., & Levi, M. D., (2012). The Relative-Age Effect and Career Success: Evidence from Corporate CEO. *Economics Letters*, 117(3), 660-662.

Dubner, S. J., & Levitt, S. D. (2010). *Superfreakonomics: Global Cooling, Patriotic Prostitutes and Why Suicide Bombers Should Buy Life Insurance*. Penguin UK.

Edgar, S., & O'Donoghue, P. (2005), Season of Birth Distribution of Elite Tennis Players. *Journal of sports sciences*, 23(10), 1013-1020.

Frank, E., & Nüesch, S. (2012). Talent and/or Popularity: What Does it Take to Be a Superstar? *Economic Inquiry*, 50(1), 202-216.

Fredriksson, P., & Öckert, B. (2006). Is Early Learning Really More Productive? The Effect of School Starting Age on School and Labor Market Performance. *Working Paper, IFAU-Institute for Labour Market Policy Evaluation*. 12, 1-52.

Gibbs, B. G., Jarvis, J. A., & Dufur, M. J. (2012). The Rise of the Underdog? The Relative Age Effect Reversal Among Canadian-Born NHL Hockey Players: A Reply to Nolan and Howell. *International Review for the Sociology of Sport*, 47(5), 644-649.

Gladwell, M. (2008). *Outliers: The story of success*. Penguin UK.

Glamser, F. D., & Vincent, J. (2004). The Relative Age Effect Among Elite American Youth Soccer Players. *Journal of sport Behavior*, 27(1), 31-38.

Hancock, D. J., Adler, A. L., & Côté, J. (2013). A Proposed Theoretical Model to Explain Relative Age Effects in Sport. *European journal of sport science*, 13(6), 630-637.

Hao, L., & Naiman, D. Q. (2007). *Quantile regression* (149). Sage.

Helsen, W. F., Starkes, J. L., Van Winckel, J. (1998). The Influence of Relative Age on Success and Dropout in Male Soccer Players. *American Journal of Human Biology*, 10(6), 791-798.

Helsen, W. F., Starkes, J. L., & Van Winckel, J. (2000). Effect of a Change in Selection Year on Success in Male Soccer Players. *American Journal of Human Biology*, 12(6), 729-735.

Helsen, W. F., Baker, J., Michiels, S., Schorer, J., Van Winckel, J., Williams, A. M. (2012). The Relative Age Effect in European Professional Soccer: Did Ten Years of Research Make any Difference? *Journal of Sports Science*, 30(15), 1665-1671.

Joyner. P. W., Mallon, W. J., Kirkendall, D. T., & Garret Jr. W. E. (2013) Relative Age Effect: Beyond the Youth Phenomenon. *The Duke Orthopaedic Journal*, 3(1), 74-79.

Kahn, L. M. (2000). The Sports Business as a Labor Market Laboratory. *The Journal of Economic Perspectives*, 14(3), 75-94.

Kleven, H. J., Landais, C., & Saez E. (2013). Taxation and International Migration of Superstars: Evidence from the European Football Market. *American Economic Review*, 103(5), 1892-1924.

Kniffin, K. M., & Hanks, A. S. (2013). Revisiting Gladwell's Hockey Players: Influence of Relative Age Effects upon Earning the PhD. *CHERI Working Paper*, 157.

Kuczmariski, R. J., Ogden, C. L., Guo, S. S., Grummer-Strawn, L. M., Flegal, K. M., Mei, Z., Wei, R., Curtin, L. R., Roche, A. F., & Johnson, C. L. (2002). 2000 CDC Growth Charts for the United States: Methods and Development. *Vital and health statistics. Series 11, Data From the National Health Survey*, 246, 1-190.

Larsen, E. R., & Solli, I. F. (2012) *Born to run Behind? Persisting Relative Age Effects on Earnings*. University of Stravenger, Faculty of Social Sciences.

Lucifora, C., & Simmons, R. (2003). Superstar effects in sport evidence from Italian soccer. *Journal of Sports Economics*, 4(1), 35-55.

Malmendier, U., & Tate, G. (2009). Superstar CEOs. *The Quarterly Journal of Economics*, 124(4), 1593-1638.

McDonald, J.H. 2014. Handbook of Biological Statistics. Sparky House Publishing, Baltimore, Maryland.

Muller-Daumann, D., & Page, D. (2013). Political Selection and the Relative Age Effect. *QuBE Working Papers 009, QUT Business School*.

Musch, J., & Grondin, S. (2001). Unequal Competition as an Impediment to Personal Development: A Review of the Relative Age Effect in Sport. *Development Review*, 21(2), 147-167.

Musch, J., & Hay, R. (1999). The Relative Age Effect in Soccer: Cross-Cultural Evidence for a Systematic Distribution Against Children Born Late in the Competition Year. *Sociology of Sport Journal*, 16(1), 54-6.

Parent-Harvey C. I., Desjardins C. & Harvey, E. J. (2014). Factors Affecting the Relative Age Effect in NHL Athletes. *Canadian Journal of Surgery*, 57(3), 157-161.

Parente, P. M. D. C., & Santos Silva, J. M. C. (2013). Quantile Regression with Clustered Data. *University of Exeter, Economics Department Discussion Papers Series*, 13(5), 1-24.

Pellizzari, M., & Billari, F. C. (2012). The Younger, the Better? Age-Related Differences in Academic Performance at University. *Journal of Population Economics*, 25(2), 697-739.

Plug, E. J. S. (2001). Season of Birth, Schooling and Earnings. *Journal of Economic Psychology*, 22(5), 641-660.

Ponzo, M., & Scoppa, V. (2014). The Long-Lasting Effects of School Entry Age: Evidence from Italian Students. *Journal of Policy Modeling*, 36(3), 578-599.

Prioux, F. (1988). Mouvement Saisonnier des Naissances: Influence du Rang et de la Légitimité dans Quelques Pays d'Europe Occidentale. *Population*, 3, 587-609.

Rizzi, E. L., & Dalla-Zuanna (2007). The Seasonality of Conception. *Demography*, 44(4), 705-728.

Rosen, S. (1981). The economics of superstars. *The American Economic Review*. 71(5), 845-858.

Salinero, J. J., Pérez, B., Burillo, P., & Lesma, M. L. (2013). Relative Age Effect in European Professional Football. Analysis by position. *Journal of Human Sport & Exercise*, 8(4), 966-973.

Salthouse, T. A., Schroeder D. H., & Ferrer, E. (2004). Estimating Retest Effects in Longitudinal Assessments of Cognitive Functioning in Adult between 18 and 60 Year of Age *Developmental Psychology*, 40(5), 813-822.

Sims, J., & Addona, V. (2014). Hurdle Models and Age Effects in the Major League Baseball Draft. *Journal of Sports Economics*, 1-16.

Thompson, A. H., Barnsley, R. H., & Battle, J. (2004). The Relative Age Effect and the Development of Self-Esteem. *Educational Research*, 46(3), 313-320.

Thompson, A. H., Barnsley, R. H., & Dyck, R. J. (1999). A New Factor in Youth Suicide: The Relative Age Effect. *Canadian journal of psychiatry*, 44(1), 82-85.

Wattie, N., Schorer, J., & Baker, J. (2015). The Relative Age Effect in Sport: A Developmental Systems Model. *Sports Medicine*, 45(1), 83-94.

Williams J., H. (2010). Relative Age Effect in Youth Soccer: Analysis of the FIFA U17 World Cup Competition. *Scandinavian Journal of Medicine and Science in Sports*, 20(3), 502-508.

Zoëga, H., Valdimarsdóttir, U. A., & Hernández-Diaz, S. (2012). Age, Academic Performance, and Stimulant Prescribing for ADHD: A Nationwide Cohort Study. *Pediatrics*, *130*(6), 1012-1018.

Appendix A

Table A.1 Variables description.

Variable	Description
Ln(w)	Natural logarithm of the gross wage deflated at 2013 for player i.
Age	Age of player i. This variable is rescaled; it ranges from 0 to 26 (0 corresponds to 17, while 26 corresponds to 43).
Age ²	Squared age of player i.
Quarter dummies	Relative age in terms of quarter of birth for player i. The reference quarter is January-March.
December dummy	Relative age for players i born in December. The reference month is January.
Team dummies	Dummies for team; Udinese is the reference team.
Season dummies	Dummies for season; 2013-14 is the reference season.

Table A.2 Descriptive statistics for wage and age, by quarter of birth.

Variable	Whole sample	Q1	Q2	Q3	Q4
Ln(w)	0.126 (0.835)	0.128 (0.813)	0.078 (0.753)	0.310 (0.894)	-0.048 (0.894)
Age	28.046 (4.469)	27.777 (4.483)	28.095 (4.392)	28.082 (4.596)	28.399 (4.406)

Note: Mean Ln(w) and age; standard deviation in parentheses.

Appendix B

Table B.1 reports the number of births in the Italian general population from 1993 to 1998. It contains the statistics used to compute the Italian quarterly and monthly birthrates.

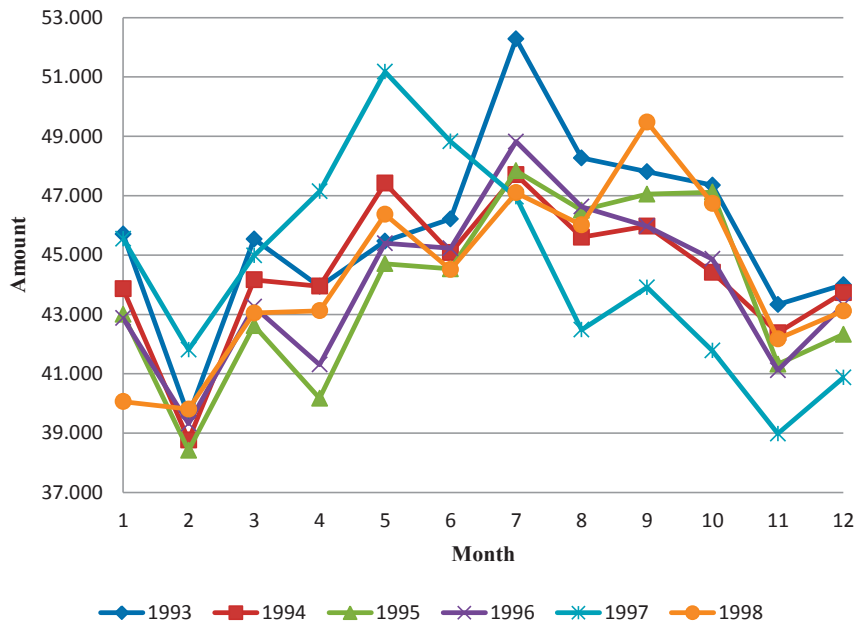
Table B.1 Number of births in the Italian general population from 1993 to 1998.

Period	1993	1994	1995	1996	1997	1998	Average counts	Average rate
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
January	45,704	43,859	42,996	42,872	45,549	40,063	43,507.167	0.082
February	39,578	38,767	38,420	39,318	41,809	39,807	39,616.500	0.074
March	45,536	44,163	42,618	43,258	44,982	43,047	43,934.000	0.082
April	43,942	43,948	40,172	41,306	47,143	43,124	43,272.500	0.081
May	45,465	47,419	44,708	45,398	51,174	46,371	46,755.833	0.088
June	46,215	45,082	44,533	45,228	48,824	44,510	45,732.000	0.086
July	52,284	47,703	47,839	48,822	46,955	47,099	48,450.333	0.091
August	48,270	45,599	46,507	46,630	42,482	46,014	45,917.000	0.086
September	47,807	45,976	47,050	45,971	43,910	49,480	46,699.000	0.087
October	47,353	44,419	47,118	44,866	41,778	46,737	45,378.500	0.085
November	43,330	42,382	41,321	41,118	38,981	42,176	41,551.333	0.078
December	44,000	43,733	42,327	43,316	40,874	43,120	42,895.000	0.080
Year	549,484	533,050	525,609	528,103	534,461	531,548	533,709.167	100

Note: Columns (1)-(6) report the number of births in the Italian general population per month and per year from 1993 to 1998. Column (7) reports the average number of births per month and per year from 1993 to 1998. Column (8) reports the average monthly birth rate (monthly average / yearly average).

Figure B.1 plots the number of births per month by year for the general Italian population.

Figure B.1



The majority of the births are concentrated in the summer and the beginning of fall, while the lowest levels of births occur in the winter.

Appendix C

Main analysis - winsorization at 3% and 97%.

Table C.1 Wage gap between relatively older and younger players. Winsorization at 3% and 97%.

Variables	(1)	(2)	(3)	(4)
Age	0.170*** (0.029)	0.168*** (0.029)	0.168*** (0.019)	0.149*** (0.045)
Age ²	-0.006*** (0.001)	-0.005*** (0.001)	-0.006*** (0.001)	-0.005*** (0.002)
Q2 (April-June)	-0.061 (0.090)	-0.060 (0.090)	-0.056 (0.051)	-
Q3 (July-September)	0.179* (0.108)	0.178* (0.108)	0.052 (0.065)	-
Q4 (October-December)	-0.162 (0.103)	-0.165 (0.103)	-0.131** (0.063)	-
December	-	-	-	-0.176** (0.078)
Constant	-0.953*** (0.163)	-0.855*** (0.174)	-1.028*** (0.166)	-1.025*** (0.356)
F-test, quarters of birth (p-value)	0.026	0.026	0.043	-
Season F.E.	N	Y	Y	Y
Team F.E.	N	N	Y	Y
R-square	0.112	0.123	0.566	0.700
Observations	1,598	1,598	1,598	282

Note: Standard errors clustered on footballers are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The samples for the analyses use less than 1,704 observations because of missing values for wages. Column (1) reports the results from the OLS without fixed effects; column (2) reports the results from the OLS with season fixed effects; column (3) reports the results from the OLS with players' team and season fixed effects; and column (4) reports the results obtained using the discontinuity sample strategy, i.e., only players born in January or December, which are obtained from the OLS with players' team and season fixed effects. "F-test, quarters of birth" provides the p-value from the F-test on the joint significance of the quarters of birth estimates.

Column (3), with season and players' team fixed effects, provides estimates similar to those displayed in Table 3, whereas the estimate for the December coefficient from column (4) is almost halved.

Analysis by age group - winsorization at 3% and 97%.

Table C.2 Wage gap between relatively older and younger players by age group.
Winsorization at 3% and 97%.

Variables	< 21 (1)	21-25 (2)	26-30 (3)	> 30 (4)
Age	0.480* (0.270)	0.173 (0.155)	0.217 (0.228)	0.280** (0.128)
Age ²	-0.070 (0.075)	-0.006 (0.012)	-0.008 (0.010)	-0.010*** (0.004)
Q2 (April-June)	-0.352 (0.344)	-0.073 (0.079)	-0.036 (0.068)	-0.091 (0.084)
Q3 (July-September)	-0.575* (0.338)	0.014 (0.103)	0.072 (0.090)	0.092 (0.106)
Q4 (October-December)	0.116 (0.235)	-0.190* (0.103)	-0.221** (0.0880)	-0.121 (0.101)
Constant	-1.257** (0.559)	-0.986** (0.496)	-1.430 (1.230)	-1.593 (1.128)
F-test, quarters of birth (p-value)	0.263	0.229	0.019	0.197
Season F.E.	Y	Y	Y	Y
Team F.E.	Y	Y	Y	Y
R-square	0.889	0.502	0.556	0.640
Observations	51	403	629	515

Note: Standard errors clustered on players are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The samples for the analyses use less than 1,704 observations because of missing values for wages. Column (1) reports the results from the OLS with players' team and season fixed effects for players younger than 21 years of age; column (2) reports the results from the OLS with players' team and season fixed effects for players between 21 and 25 years of age; column (3) reports the results from the OLS with players' team and season fixed effects for players between 26 and 30 years of age; and column (4) reports the results from the OLS with players' team and season fixed effects for players older than 30 years of age. "F-test, quarters of birth" provides the p-value from the F-test on the joint significance of the quarters of birth estimates.

The estimates are similar to those displayed in Table 4.

Appendix D

In this appendix, the main analysis is repeated on different sub-samples based on the players' role. Unfortunately, the sample sizes for the analyses by players' role mostly lead to inconclusive results, except for the subsample of midfielders. Therefore, any interpretation of the estimates in Table D.1, D.2, D.3, and D.4 should be considered carefully.

The analysis on the goalkeepers' sub-sample is reported below in Table D.1.

Table D.1 Wage gap between relatively older and younger players, only goalkeepers.

Variables	(1)	(2)	(3)	(4)
Age	0.286*** (0.089)	0.286*** (0.095)	0.201*** (0.063)	0.347*** (0.099)
Age ²	-0.010*** (0.003)	-0.0098*** (0.003)	-0.007*** (0.002)	-0.0125*** (0.003)
Q2 (April-June)	-0.527 (0.387)	-0.513 (0.419)	-0.295** (0.122)	-
Q3 (July-September)	-0.042 (0.345)	-0.038 (0.351)	0.032 (0.147)	-
Q4 (October-December)	-0.571 (0.361)	-0.556 (0.368)	-0.389 (0.339)	-
December	-	-	-	0.187 (0.320)
Constant	-1.706*** (0.483)	-1.600** (0.601)	-1.623** (0.616)	-2.461*** (0.176)
F-test, quarters of birth (p-value)	0.181	0.221	0.048	-
Season F.E.	N	Y	Y	Y
Team F.E.	N	N	Y	Y
R-square	0.208	0.219	0.729	0.864
Observations	163	163	163	33

Note: Standard errors clustered on footballers are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. These analyses are run on the sub-sample of Italian goalkeepers, n=175; the missing values for wages are such that the analyzed sub-sample is smaller. Column (1) reports the results from the OLS without fixed effects; column (2) reports the results from the OLS with season fixed effects; column (3) reports the results from the OLS with players' team and season fixed effects; and column (4) reports the results obtained using the discontinuity sample strategy, i.e., only players born in January or December, which are obtained from the OLS with players' team and season fixed effects. "F-test, quarters of birth" provides the p-value from the F-test on the joint significance of the quarters of birth estimates.

The estimates for quarters of birth are rarely statistically significant. The analysis on the goalkeepers' sub-sample seems to suggest the presence of an RAE and, as expected, that the RAE is stronger for goalkeepers; the literature suggests that this result is due to the

importance of physical characteristics in this role (Salinero et al., 2013). However, the large and statistically significant RAE for players born in the second quarter, see column (3), is puzzling. Moreover, column (4) suggests a reverse RAE for goalkeepers born in December.

The analysis on the defenders' sub-sample is reported below in Table D.2.

Table D.2 Wage gap between relatively older and younger players, only defenders.

Variables	(1)	(2)	(3)	(4)
Age	0.196** (0.081)	0.195** (0.079)	0.227*** (0.059)	0.727*** (0.240)
Age ²	-0.006* (0.003)	-0.006* (0.003)	-0.008*** (0.002)	-0.028*** (0.010)
Q2 (April-June)	-0.041 (0.144)	-0.039 (0.143)	-0.006 (0.084)	-
Q3 (July-September)	0.037 (0.197)	0.035 (0.197)	-0.078 (0.104)	-
Q4 (October-December)	-0.108 (0.182)	-0.113 (0.183)	-0.135 (0.135)	-
December	-	-	-	-0.233 (0.293)
Constant	-1.266*** (0.443)	-1.338*** (0.506)	-1.632*** (0.376)	-5.557*** (1.684)
F-test, quarters of birth (p-value)	0.896	0.888	0.675	-
Season F.E.	N	Y	Y	Y
Team F.E.	N	N	Y	Y
R-square	0.139	0.155	0.586	0.817
Observations	582	582	582	83

Note: Standard errors clustered on footballers are reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. These analyses are run on the sub-sample of Italian defenders, $n=604$; the missing values for wages are such that the analyzed sub-sample is smaller. Column (1) reports the results from the OLS without fixed effects; column (2) reports the results from the OLS with season fixed effects; column (3) reports the results from the OLS with players' team and season fixed effects; and column (4) reports the results obtained using the discontinuity sample strategy, i.e., only players born in January or December, which are obtained from the OLS with players' team and season fixed effects. "F-test, quarters of birth" provides the p-value from the F-test on the joint significance of the quarters of birth estimates.

The estimates for quarters of birth are never statistically significant. The analysis on the defenders' sub-sample seems to suggest the presence of an RAE. However, the estimates are smaller than those obtained in the main analysis; according to the literature (e.g., Salinero et al., 2013), the estimates are expected to be larger in the presence of an RAE.

The analysis on the midfielders' sub-sample is reported below in Table D.3.

Table D.3 Wage gap between relatively older and younger players, only midfielders.

Variables	(1)	(2)	(3)	(4)
Age	0.264*** (0.050)	0.262*** (0.050)	0.213*** (0.038)	0.198*** (0.070)
Age ²	-0.010*** (0.002)	-0.010*** (0.002)	-0.007*** (0.002)	-0.00598* (0.003)
Q2 (April-June)	0.090 (0.145)	0.087 (0.146)	0.024 (0.075)	-
Q3 (July-September)	0.120 (0.165)	0.109 (0.165)	-0.064 (0.083)	-
Q4 (October-December)	-0.331** (0.133)	-0.343** (0.135)	-0.176** (0.079)	-
December	-	-	-	-0.177** (0.085)
Constant	-1.485*** (0.273)	-1.403*** (0.289)	-1.438*** (0.233)	-1.426*** (0.444)
F-test, quarters of birth (p-value)	0.009	0.008	0.074	-
Season F.E.	N	Y	Y	Y
Team F.E.	N	N	Y	Y
R-square	0.170	0.183	0.634	0.804
Observations	572	572	572	126

Note: Standard errors clustered on footballers are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. These analyses are run on the sub-sample of Italian midfielders, n=612; the missing values for wages are such that the analyzed sub-sample is smaller. Column (1) reports the results from the OLS without fixed effects; column (2) reports the results from the OLS with season fixed effects; column (3) reports the results from the OLS with players' team and season fixed effects; and column (4) reports the results obtained using the discontinuity sample strategy, i.e., only players born in January or December, which are obtained from the OLS with players' team and season fixed effects. "F-test, quarters of birth" provides the p-value from the F-test on the joint significance of the quarters of birth estimates.

The analysis on the midfielders' sub-sample provides statistically significant evidence of the RAE, which is also confirmed by the estimate of the RAE for midfielders born in December; see column (4). The preferred model, which is the model that includes seasons and team fixed effects, provides estimates for the RAE that are close to those obtained in the corresponding model in the main analysis.

The analysis on the forwards' sub-sample is reported below in Table D.4.

Table D.4 Wage gap between relatively older and younger players, only forwards.

Variables	(1)	(2)	(3)	(4)
Age	0.217** (0.0968)	0.229** (0.0975)	0.234*** (0.053)	0.0324 (0.089)
Age ²	-0.007 (0.004)	-0.007* (0.004)	-0.008*** (0.002)	-0.003 (0.004)
Q2 (April-June)	-0.092 (0.211)	-0.096 (0.214)	-0.001 (0.130)	-
Q3 (July-September)	0.319 (0.209)	0.327 (0.210)	0.212* (0.126)	-
Q4 (October-December)	0.0511 (0.277)	0.0601 (0.282)	0.037 (0.152)	-
December	-	-	-	-
Constant	-0.992* (0.514)	-0.964* (0.527)	-1.297*** (0.371)	0.781 (0.650)
F-test, quarters of birth (p-value)	0.234	0.224	0.335	-
Season F.E.	N	Y	Y	Y
Team F.E.	N	N	Y	Y
R-square	0.209	0.224	0.664	0.906
Observations	281	281	281	40

Note: Standard errors clustered on footballers are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. These analyses are run on the sub-sample of Italian forwards, n=310; the missing values for wages are such that the analyzed sub-sample is smaller. Column (1) reports the results from the OLS without fixed effects; column (2) reports the results from the OLS with season fixed effects; column (3) reports the results from the OLS with players' team and season fixed effects; and column (4) reports the results obtained using the discontinuity sample strategy, i.e., only players born in January or December, which are obtained from the OLS with players' team and season fixed effects. "F-test, quarters of birth" provides the p-value from the F-test on the joint significance of the quarters of birth estimates.

The estimates for quarters of birth are rarely statistically significant. The analysis on the forwards' sub-sample does not seem to suggest the presence of an RAE. No estimate for the RAE is displayed in column (4) because there is only 1 forward born in December for the period under consideration.

Linnaeus University Dissertations

Below please find a list of recent publications in the series Linnaeus University Dissertations. For a full list and more information: Lnu.se

170. Oscarsson, Jan 2014. *Strength grading of structural timber and EWP laminations of Norway spruce – development potentials and industrial applications* (byggteknik/building technology) ISBN: 978-91-87427-84-8.

171. Gil de la Iglesia, Didac 2014. *A Formal Approach for Designing Distributed Self-Adaptive Systems* (mediateknik/media technology) ISBN: 978-91-87427-85-5.

172. Svensson, Henric 2014. *Characterization, toxicity and treatment of wood leachate generated outdoors by the wood-based industry* (miljövetenskap/environmental science) ISBN: 978-91-87427-86-2.

173. André Salomão 2014. *Occurrence and ecotoxicity of endocrine disruptor chemicals in aquatic environment and sewage treatment systems* (miljöteknik/environmental technology) ISBN: 978-91-87427-87-9.

174. Lars Eliasson 2014. *Some aspects on quality requirements of wood for use in the industrial manufacture of single-family timber houses* (skog och träteknik/forestry and wood technology) ISBN: 978-91-87427-88-6.

175. Jonas Lundberg 2014. *Fast and Precise Points-to Analysis* (datavetenskap/computer science) ISBN: 978-91-87427-89-3.

176. Sepideh Olausson 2014. *Intensivvårdsrummets betydelse för vårdande och välbefinnande – patienters närståendes och vårdpersonalens erfarenheter* (vårdvetenskap/caring science) ISBN: 978-91-87427-90-9.

177. Margaretha Larsson 2014. *Tonårsflickors hälsa – att stödja och stärka* (vårdvetenskap/caring science) ISBN: 978-91-87427-92-3.

178. Johan Stedt 2014. *Wild Birds as Carriers of Antibiotic Resistant E. coli and Extended-Spectrum Beta-Lactamases* (ekologi/ecology) ISBN: 978-91-87427-93-0.

179. Kerstin Golker 2014. *Fundamental Studies on Molecularly Imprinted Materials* (kemi/chemistry) ISBN: 978-91-87427-94-7.

180. Helena Ackesjö 2014. *Barns övergångar till och från förskoleklass. Gränser, identiteter och (dis-)kontinuiteter* (pedagogik/education) ISBN: 978-91-87427-95-4.

181. Astrid Skoglund 2014. *Kommunikativa strategier i texter om tobaks-
avvänjning. Innehåll, argumentation och modellläsare* (svenska språket/Swedish)
ISBN: 978-91-87925-06-1.
182. Stefan Gustafsson 2014. *Electromagnetic Dispersion Modeling and Analysis
for Power Cables* (fysik/physics) ISBN: 978-91-87925-07-8.
183. Magdalena Elmersjö 2014. *Kompetensfrågan inom äldreomsorgen. Hur upp-
fattningar om kompetens formar omsorgarbetet, omsorgsbehoven och omsorgsrelation-
en* (socialt arbete/social work) ISBN: 978-91-87925-08-5.
184. Elisabeth Lindberg 2014. *Tid för vårdande möten. Att vidmakthålla och ut-
veckla vårdandet med patientperspektivet i fokus* (vårdvetenskap/caring science)
ISBN: 978-91-87925-09-2.
185. Andrew Wekesa Wamukota 2014. *Fishermen and fishtraders in Kenya: Re-
sources, livelihoods and the influence of global markets* (miljöveten-
skap/environmental science) ISBN: 978-91-87925-10-8.
186. Henrik Hansson 2014. *Treatment of wastewater generated by wood-based
dry industries: advanced oxidation processes and electrocoagulation* (miljöveten-
skap/environmental science) ISBN: 978-91-87925-11-5.
187. Thorbjörn Ahlgren 2014. *Institutionalisering på hemmaplan. En idé resa i
den sociala barnavården* (socialt arbete/social work) ISBN: 978-91-87925-14-6.
188. Tora Hammar 2014. *eMedication – improving medication management us-
ing information technology.* (biomedicinsk vetenskap/ biomedical sciences)
ISBN: 978-91-87925-15-3.
189. Lars Fonseca 2014. *Det godkända fusket – normförhandlingar i gymnasiesko-
lans bedömningspraktiker.* (pedagogik/education) ISBN: 978-91-87925-16-0.
190. Ulf Larsson-Olaison 2014. *Convergence of National Corporate Governance
Systems: Localizing and Fitting the Transplants.* (ekonomistyrning/accounting
and management control) ISBN: 978-91-87925-17-7.
191. Richard Shukia 2014. *Learning and Teaching to Read in Kiswahili in Pre-
Primary Classes in Tanzania: Teachers' Beliefs and Instructional Practices.* (peda-
gogik/education) ISBN: 978-91-87925-18-4.
192. Louise Elmlund 2014. *QCM-based sensing using biological and biomimetic
interfaces.* (kemi/chemistry) ISBN: 978-91-87925-20-7.
193. Anna-Kari Trondman 2014. *Pollen-based quantitative reconstruction of
land-cover change in Europe from 11,500 years ago until present – a dataset suitable
for climate modelling* (miljövetenskap/environmental science) ISBN: 978-91-
87925-21-4.
194. Marianne Dahl 2014. *Fritidspedagogers handlingsrepertoar. Pedagogiskt ar-
bete med barns olika relationer* (pedagogik/education) ISBN: 978-91-87925-22-
1.

195. Kirsi Jarnerö 2014. *Vibrations in timber floors – Dynamic properties and human perception* (byggt teknik/civil engineering) ISBN: 978-91-87925-23-8.
196. Bahtijar Vogel 2014. *An Open Architecture Approach for the Design and Development of Web and Mobile Software* (datavetenskap/computer science) ISBN: 978-91-87925-24-5.
197. Olga Maskenskaya 2014. *Abundance and fractionation of rare earth elements in calcite and other secondary minerals in fractures in the upper kilometre of crystalline bedrock, SE Sweden* (miljövetenskap/environmental science) ISBN: 978-91-87925-25-2.
198. Ann-Kristin Fjällström 2014. *Signaling factors related to atrophy and hypertrophy in denervated skeletal muscle* (biomedicinsk vetenskap/biomedical science) ISBN: 978-91-87925-26-9.
199. Marie Gunnarsson 2014. *Lärande i arbetslivets övergångar* (pedagogik/education) ISBN: 978-91-87925-27-6.
200. Pontus Wallin 2014, *Authoritarian collaboration. Unexpected effects of open government initiatives in China* (statsvetenskap/political science) ISBN: 978-91-87925-29-0.
201. Simone Scarpa 2014, *The spatial manifestation of inequality: residential segregation in Sweden and its causes* (socialt arbete/social work) ISBN: 978-91-87925-32-0.
202. Kim Evertsson 2014, *Protein expression and modifications in denervated atrophic and hypertrophic skeletal muscle* (biomedicinsk vetenskap/biomedical science) ISBN: 978-91-87925-33-7.
203. Changxun Yu 2014, *Metal abundance and behavior in geo-materials affected by metal-sulfide oxidation* (miljövetenskap/environmental science) ISBN: 978-91-87925-36-8.
204. Markus Lindh 2014, *Bacterioplankton population dynamics in a changing ocean* (miljövetenskap/environmental science) ISBN: 978-91-87925-38-2.
205. Gustaf Olsson 2014, *Modelling of Molecular Interactions in Mixtures: Fundamental Studies and Applications* (kemi/chemistry) ISBN: 978-91-87925-41-2.
206. Ke Yang 2015, *Regulation of artemisinin biosynthesis in Artemisia annua L* (biomedicin/biomedical science) ISBN: 978-91-87925-43-6.
207. Henrik Nilsson 2015, *Kultur och utbildning – en tolkning av två grundskolors mångkulturella kontexter* (pedagogik/education) ISBN: 978-91-87925-44-3.
208. Ida Ohlsson Al Fakir 2015, *Nya rum för socialt medborgarskap. Om vetenskap och politik i "Zigenarundersökningen" – en socialmedicinsk studie av svenska romer 1962–1965* (historia/history) ISBN: 978-91-87925-46-7.

209. Antonina Danylenko 2015, *Decision Algebra: A General Approach to Learning and Using Classifiers* (datavetenskap/computer science) ISBN: 978-91-87925-47-4.
210. René Nordstierna 2015, *Osäkra möten. Perspektiv på metodbildning för det konstnärliga högskoleutbildningsfältet* (pedagogik/education) ISBN: 978-91-87925-48-1.
211. Reza Mahani 2015, *Magnetic solotronics near the surface of a semiconductor and a topological insulator* (fysik/physics) ISBN: 978-91-87925-49-8.
212. Niklas Andersson 2015, *När vårdande och lärande sammanfaller .Patienters, studenters och handledares erfarenheter av möten på en utbildningsvårdavdelning inom psykiatrisk vård* (vårdvetenskap/nursing) ISBN: 978-91-87925-50-4.
213. Luca Fumarco 2015, *Essays on discrimination in the marketplace* (nationalekonomi/economics) ISBN: 978-91-87925-51-1.