Gender and Ethnic Discrimination in Hiring
- Evidence from Field Experiments in the German Labor Market -

Der Fakultät für Wirtschaftswissenschaften der
Universität Paderborn
zur Erlangung des akademischen Grades
Doktor der Wirtschaftswissenschaften
- Doctor rerum politicarum -
vorgelegte Dissertation
von

Andre Kolle

geboren am 3. September 1984 in Seesen

2014

Vorwort

Diese Arbeit wäre sicherlich nicht oder zumindest nicht in dieser Form entstanden, wenn mir nicht die Unterstützung zahlreicher Menschen zuteil gekommen wäre, für die ich jedem einzelnen unendlich dankbar bin. Die Aufzählung aller würde sicherlich den Rahmen sprengen, eine Liste ist aber vom Autor auf Anfrage erhältlich. Trotzdem möchte ich an dieser Stelle einige Personen besonders hervorheben.

Allen voran möchte ich meinem Doktorvater Prof. Dr. Bernd Frick für seine unermüdliche Hilfe, seinen Zuspruch für und seine Aufgeschlossenheit gegenüber einer nicht immer unumstrittenen Vorgehensweise sowie die immer wieder erkenntnisreichen und motivierenden gemeinsamen Diskussionen danken. Mein besonderer Dank gilt auch Prof. Dr. René Fahr, der mir ebenso offen und geduldig Anregungen und Verbesserungsvorschläge gegeben und mir jederzeit mit Rat und Tat zur Seite gestanden hat. Auch den weiteren Kommissionsmitgliedern, Prof. Dr. Burkhard Hehenkamp und Prof. Dr. Martin Schneider, möchte ich für ihre Bereitschaft, sich meiner Arbeit anzunehmen, recht herzlich danken. Letzterem gebührt dabei mein besonderer Dank, da mein Weg durch die Arbeit als studentische Hilfskraft an seinem Lehrstuhl erheblich mitgeprägt wurde.

Mindestens genauso entscheidend haben aber auch meine Kolleginnen und Kollegen am Lehrstuhl, Prof. Dr. Christian Deutscher, Dr. Marcel Battré, Dr. Linda Kurze, Friedrich Scheel, Anica Rose, Tobias Neuhann und Filiz Gülal, von denen viele zu sehr guten Freunden geworden sind, zur erfolgreichen Erstellung dieser Arbeit beigetragen. Ich kann mich wirklich glücklich schätzen, in einem solchen Team und in einer solch positiven Atmosphäre gearbeitet haben zu dürfen. Im selben Atemzug möchte ich hier auch unsere studentischen Hilfskräfte erwähnen, die mich in vielerlei Dingen rund um meine Doktorarbeit und darüber hinaus unterstützt haben.

Abschließend möchte ich mich bei meinen Eltern, meiner Familie, meinen Freunden und meiner Freundin Marleen dafür bedanken, dass sie meinen Weg mitgeprägt, immer wieder für die nötige geistige Zerstreuung gesorgt und mir zu jeder Zeit mit Liebe und Kraft zur Seite gestanden haben. Ihr alle seid ein Teil dieses für mich so großartigen Erfolgs!

Andre Kolle — Paderborn im März 2014

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AGG	General Act on Equal Treatment (Allgemeines Gleichbehandlungsgesetz)
AFQT	Armed Forces Qualification Test
BA	German Federal Employment Agency (Bundesagentur für Arbeit)
BHPS	British Household Panel Survey
BIBB	Federal Institute of Vocational Education and Training (Bundesinstitut für Berufsbildung)
cdf	Cumulative distribution function
CDU	Christian Democratic Union
CPS	Current Population Survey
CSU	Christian Socialistic Union
DGB	The Confederation of German Trade Unions (Deutscher Gewerkschaftsbund)
DIW	German Institute for Economic Research (Deutsches Institut für Wirtschaftsforschung)
ESS	European Social Survey
FDP	Free Democratic Party
GoF	Goodness of fit
GSOEP	German Socio-Economic Panel (Sozio-ökonomisches Panel)
ILO	International Labour Organization
LIAB	Linked Employer-Employee Data
LPM	Linear Probability Model
LR	Likelihood ratio
ML	Maximum Likelihood
NLS	National Longitudinal Surveys
NLSY	National Longitudinal Survey of Youth
NPD	National Democratic Party of Germany
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary Least Squares
PSID	Panel Study of Income Dynamics
PUMS	Public Use Microdata Sample

SEO	Survey of Economic Opportunity
SES	Structural Earnings Survey
SPD	Social Democratic Party of Germany

1 INTRODUCTION

A major challenge in contemporary business environments is recruiting qualified staff that meets the increasing job requirements. Due to the fierce competition for talent and the demographic change characterizing labor markets, firms and the economy as a whole are required to activate unused potential and rely on demographic groups insufficiently considered in previous hiring campaigns (e.g. The Bundestag, 2002; European Commission, 2011). However, looking at the stylized facts for Germany and other industrialized countries reveals that, among others, women and ethnic minorities still have worse employment outcomes in comparison to men and native Germans, respectively. They have inferior human capital endowments when entering the labor market, have lower labor force participation and employment rates, are underrepresented in high-paying industries, occupations and firms and are eventually paid less (see chapter 2). A compelling explanation for these outcome differences is the prevalence of discrimination in the market place which has been a point of focus among equal rights activists, policy makers and researchers all over the world. According to the German General Act on Equal Treatment (AGG) from 2006, discrimination exists whenever individuals are subject to differential treatment on the grounds of race or ethnicity, gender, religion or ideology, disability, age or sexual orientation.

Discrimination has been found to prevail in various domains (e.g. Riach and Rich, 2002; Pager and Shepherd, 2008). Research areas include the housing, credit and product market. Studies on housing discrimination focus on residential segregation and rely on field experiments that investigate differences in access to purchase and rental units (Yinger, 1986; Ross and Turner, 2005; Ewens et al., 2012). Discriminatory behavior in the credit market is predominantly demonstrated in the context of mortgage lending. Here, administrative data including a wide range of financial and property background variables are used, just as data from audited inquiries by testers from varied racial backgrounds (Munnell et al., 1996; Ladd, 1998; Pope and Sydnor, 2011). With respect to service and product markets, the most prominent research papers compare price offers to otherwise equally endowed racial groups by conducting field experiments (Ayres, 1995; Ayres and Siegelman, 1995), analyzing the correlation between the share of blacks and the price level in the local area (Graddy, 1997) and investigating systematic group differences between court cases filed for consumer discrimination (Harris et al., 2005). Systematic disadvantages in these markets have not only been documented in cases of racial and ethnic minorities, but also prevail against women (Ayres and Siegelman, 1995; Goldberg, 1996; Harless and Hoffer, 2002) and disabled people (Gneezy and List, 2004).

The largest body of theoretical and empirical literature on discrimination, however, undoubtedly exists in the labor market. Altonji and Blank (1999: 3168) define discrimination here as "[…] a situation in which persons who provide labor market services and who are equally productive in a physical or material sense are treated unequally in a way that is related to an observable characteristic such as race, ethnicity, or gender." Engaging in this field of research matters for two reasons: first, because discrimination is prohibited by law (e.g. AGG, 2006) and, second, because differential treatment based on factors unrelated to productivity creates costs to employers and may lead to forgone income (e.g. Becker, 1971). The latter perspective is supported by empirical studies using firm-level and sports data. Gwartney and Haworth (1974), for example, provide evidence from professional baseball and find that clubs contracting an above-average share of black players are able to significantly increase both their wins per unit costs and home team attendance. Similar results are presented by Szymanski (2000). Using longitudinal data from English soccer over a period of 16 years (including 39 teams), he finds a positive relationship between team performance and the share of black players on a team. More precisely, the costs per unit of success are 5 percent higher for discriminators, i.e., those teams whose proportion of blacks is below the league average. Put differently, clubs that disfavor black players have to pay a 5 percent premium on top of their wage bill to be as successful as non-discriminators.

Hellerstein et al. (2002) extend the empirical literature on discrimination to the business environment. They match U.S. census and survey data including information on workforce characteristics and profitability measures, which is then particularly used to assess the correlation between the share of females and company success. The analysis supports the hypothesis that the proportion of women has a positive impact on profitability and that companies with an above average share of women outperform discriminators. Long-term effects with respect to gender discrimination and firm closure, however, cannot be identified. This, on the other hand, is suggested in a study by Weber and Zulehner (2009). Analyzing the survival rates of around 30,000 startups in Austria, they find that firms with a share of women below the average survive 18 months less as compared to those with an average or above-average percentage. Moreover, the surviving startups systematically increase the proportion of female employees as a rational reaction to the prevailing market mechanisms. The bottom line of all these studies is essentially the same: firms benefit from effectively avoiding labor market discrimination.

Despite its legal and economic importance, researchers find it hard to undoubtedly identify the prevalence of discrimination and its driving factors (e.g. Pager and Shepherd, 2008; Lang and Lehmann, 2012; Charles and Guryan, 2013). The methods used particularly depend on which stage of the employer-employee interaction is considered. Wage discrimination, for example, is predominantly looked at by conducting regression analyses using administrative data (e.g. Hellerstein and Neumark, 2006). Differential treatment across groups is then investigated controlling for differences in e.g. worker and job characteristics. Decomposition techniques further allow disentangling the effects from differences in characteristics and returns to these characteristics (Blinder, 1973; Oaxaca, 1973). The use of administrative data generally carries the risk of omitted variable biases and unobserved heterogeneity in individual characteristics both because detailed productivity measures are rarely available (Altonji and Blank, 1999). Moreover, these data may well serve for assessing wage gaps across groups, but are either unavailable or inappropriate for uncovering discrimination in access to certain jobs and hierarchical positions. Conducting surveys on attitudes and discriminatory practices against minority groups, on the other hand, would very likely elicit dishonest responses and thus biased results. Pager and Quillian (2005), for instance, reveal significant differences between what employers state and how they actually (re)act. In other words, stated and revealed preferences are likely to diverge.

A way to overcome the methodological challenges touched above is the use of field experiments (Harrison and List, 2004). Unfortunately, only a few studies are able to explore the effect of institutional changes on firms' recruiting behavior. One prominent exception is Goldin and Rouse (2000). They make use of a natural experiment, i.e., a procedural change in the hiring process of U.S. orchestras from open to blind auditions, and find a significant increase in the share of women after the sex of the candidates has been anonymized during the initial stage of the screening process. Alternatively, a strand of literature has used the audit and correspondence method in order to detect discrimination in access to employment (e.g. Charles and Guryan, 2013). These studies try to separate any joint effects that go back to differences in worker and workplace characteristics by matching job candidates with respect to socio-economic characteristics and human capital endowments. The experimental design further allows effectively reducing the biasing effects from i.) self-selection into industries and occupations, ii.) unobserved heterogeneity (of applicant characteristics), iii.) social desirability (which is especially an issue when using survey data) and iv.) applicants' unrevealed preferences. The matched pairs apply for the same job providing the same amount and quality of productivity information. Yet, the applications differ with respect to one major characteristic which distinguishes the majority from the minority group, i.e., for instance, applicants' gender or ethnic origin. Any statistically significant differences in firms' aggregate responses to each group can then be regarded as evidence for discrimination (Riach and Rich, 2002).

The prevalence of systematic differences in employment outcomes, however, raises the question as for its underlying sources. In fact, researchers find different explanations for unequal treatment depending on their field of study. Pager and Shepherd (2008), for example, identify sociological and psychological causes for discrimination which they classify into individual, organizational and structural factors. These factors in turn are found to shape people's tastes and group perceptions and thus form the grounds for two fundamental economic theories of discrimination, namely taste-based (Becker, 1971) and statistical discrimination (Arrow, 1971; Phelps, 1972; Aigner and Cain, 1977), which constitute the theoretical framework of the present thesis.

1.1 Research Gap and Research Questions

Reviewing empirical studies on unequal treatment, research on wage discrimination has clearly drawn the most attention inside and outside the German labor market (e.g. Darity and Mason, 1998; Altonji and Blank, 1999). Yet, wage discrimination may only be the 'tip of the iceberg' as group differences in pay are influenced by factors that, on their own, may be subject to discrimination. Previous findings particularly highlight the role of group segregation across industries and occupations on remuneration (e.g. Groshen, 1991; Fields and Wolff, 1995; Huffman and Cohen, 2004). Whenever a demographic group is systematically disadvantaged entering certain jobs while another group has unrestricted access, inequalities of the gender distribution across sectors are produced. The effect of these inequalities may be twofold. On the one hand, they may enhance the wage gap even though this may not provoke outright pay discrimination and, on the other hand, they may induce self-selection since disadvantaged groups adapt their career plans as a response to anticipated labor market drawbacks (Pager and Shepherd, 2008). Thus, assessing discrimination during initiation of work relationships, i.e., in the recruitment process, seems to be of particular interest and can be considered a precursor of discriminatory practices at later stages.

Empirical research on hiring discrimination has been conducted in multiple countries considering various demographic groups and using a wide array of methodological approaches (e.g. Riach and Rich, 2002). At first glance, the findings from most of these studies seem to be very homogenous. Regarding gender discrimination, for example, differential treatment is found to vary by job type where women are discriminated in male-dominated jobs while men are disfavored in female-dominated professions. Racial and ethnic minorities, on the other hand, are found to be disadvantaged independent of job types, but dependent on skin color and nationality. However, there are some exceptions that particularly demonstrate that the prevalence and magnitude of discrimination may be sensitive to certain conditions. These conditions in turn may reflect employers' motives to treat one group worse than another, all other things being equal. Indeed, there is spurious evidence that employers discriminate based on their distastes and productivity perceptions linked to group membership. Empirically, though, the emphasis thus far has predominantly been put on whether and to what extent discrimination exists. Disentangling the effects from taste-based and statistical discrimination is therefore one major challenge that will be addressed during the course of this thesis (Charles and Guryan, 2013).

Bearing in mind the enormous theoretical and empirical work on hiring discrimination, quite surprisingly, research in the German labor market is quite limited. Even demographic characteristics most commonly investigated in the existing literature, i.e., gender and ethnic origin, lack thorough evidence in particular concerning access to employment. The stylized facts and previous empirical research suggest that differential treatment in disfavor of either group prevails. Preliminary evidence supports this notion. Goldberg et al. (1996), for instance, investigate discrimination against native (first generation) Turks in eleven occupations in the mid-1990s and find evidence of significant disadvantages against the minority group. Furthermore, in a more recent study, Kaas and Manger (2012) find an average probability of receiving a positive response from employers that is 5 percentage points lower for candidates with a Turkish-sounding name as compared to their German-named counterparts. They also demonstrate that discrimination disappears if the applications include an additional reference which they interpret as evidence for statistical discrimination. However, whether their results also hold in another institutional context remains to be tested. Moreover, unlike for ethnic minorities, even less research has been undertaken on gender differences in access to employment and the conditions under which differential treatment evolves.

The purpose of this thesis therefore aims to extend prior research by investigating gender and ethnic discrimination in the recruitment process of German employers. Using correspondence testing, further insights should be provided into the prevalence as well as the factors influencing discrimination. In particular, the study compares response probabilities of men and women as well as native Germans and second generation Turks when applying for apprenticeship positions in predominantly technical occupations. The experimental design allows separating whether employers' decisions are in line with the predictions of the taste-based and/or statistical discrimination approach. Specifically, the thesis investigates the following questions:

- Do females and/or second generation Turks suffer from hiring discrimination in the German labor market for apprenticeships?
- If so, what are the factors that enforce or mitigate discriminatory behavior?
- Do taste-based and statistical discrimination affect the prevalence and/or magnitude of differential treatment?

The results may not only be of interest to the scientific community, but may be of significant practical importance. First of all, the study identifies whether discrimination is an issue that is relevant – statistically and economically. If so, it sheds more light on its underlying sources. In fact, policy implications might differ depending on the type of discrimination. In Germany, for example, policy makers have recently tested the introduction of anonymous applications in order to increase the chances of minorities of being invited to a job interview (Krause et al., 2010; Krause et al., 2012b). Now, in order to assess the rationality of such measures, empirical studies should, on the one hand, ex ante identify the prevalence and causes of discrimination and, on the other hand, evaluate their success ex post (Åslund and Nordström Skans, 2012). The former aspect clearly motivates this thesis.

1.2 STRUCTURE OF THE THESIS

The remainder of the thesis is organized as follows: chapter 2 presents stylized facts that highlight the situation of women (2.1) and ethnic minorities (2.2) in the German labor market and descriptively compares their situation with the respective majority group (males and native Germans).

Chapter 3 gives a literature overview that, on the one hand, discusses the advantages and drawbacks of different methodological approaches used to identify discrimination (3.1) and, on the other hand, reviews previous empirical findings investigating different labor market outcomes by gender (3.2.1) and ethnic origin (3.2.2). The empirical methods are further classified into regression-based approaches (3.1.1) and experiments (3.1.2) where laboratory (3.1.2.1) and field experiments (3.1.2.2) are distinguished. Insights on gender (3.2.1) and ethnic differences (3.2.2) are provided separately for wages and employment rates and for research inside and outside the German labor market. Concerning wage inequalities, only a brief overview of existing work is given whereas, with regard to employment differences, particularly the results from correspondence studies are focused upon. Moreover, in section 3.2.3, empirical research that reveals various sources of discrimination is presented. Here, the emphasis is especially placed on the separation of economically motivated factors.

Chapter 4 starts with the theoretical framework. Recruiting is analyzed within a principal-agent setting (4.1.1) and theories explaining labor market inequalities are developed (4.1.2). More specifically, different employment outcomes are explained by pre-market inequalities (4.1.2.1), human capital theory (4.1.2.2), segmented labor market theory (4.1.2.3) as well as theories of labor market discrimination (4.1.2.4). The latter are further divided into economic, i.e., taste-based (4.1.2.4.1) and statistical discrimination (4.1.2.4.2), and non-economic theories (4.1.2.5). After that, section 4.2 presents the conceptual framework that formally describes the hiring decision with special reference to the prevalence of different sources of discrimination. Based on the theoretical and empirical considerations, section 4.3 then develops testable hypotheses for both the study on gender and ethnic discrimination.

Chapter 5 comprises the empirical part. In section 5.1, the importance and suitability of the labor market for apprenticeships is highlighted (5.1.1.1 and 5.1.1.2) and the experimental design is described in detail (5.1.2–5.1.5). Section 5.2 presents the data (5.2.1), descriptive results (5.2.2) and empirical analyses (5.2.3) of the gender study. It further tests the hypotheses, discusses the findings and relates them to theory as well as to prior empirical research inside and outside the German labor market (5.2.4). Section 5.3 has a similar structure reporting the results on ethnic discrimination. Additionally, section 5.4 provides a brief methodological note that compares the outcomes of pairwise and single application tests and demonstrates the reliability of the correspondence approach.

Finally, chapter 6 draws conclusions, highlights the contributions to both the scientific community and practice and provides directions for future research.

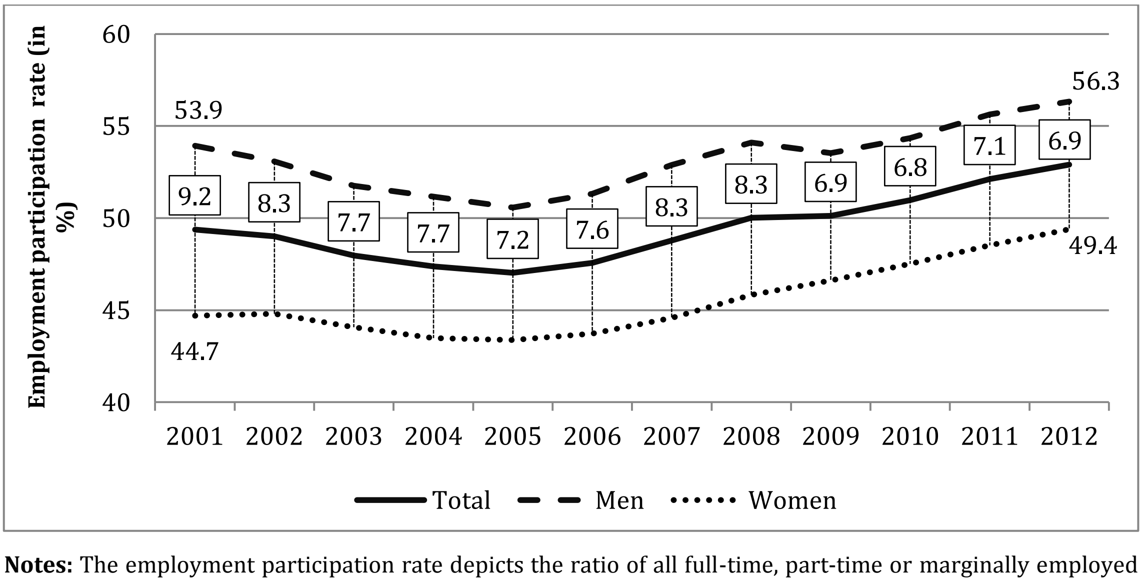
2 STYLIZED FACTS

This section reports stylized facts about the labor market situation of men and women as well as native Germans and people with migration background. It highlights the existing discrepancies in labor market outcomes between majority and minority workers and provides tentative evidence on where these observable employment differences might stem from.

2.1 THE SITUATION OF WOMEN IN THE GERMAN LABOR MARKET

Annual data from the German Federal Employment Agency (BA) shows that after a decline from 2001 until 2005, the employment ratio for both men and women has been rising except for a slight drop in 2009. The difference between men and women, however, is quite substantial but has also been declining over the last decade. While in 2012, 56.3 percent of the male population aged between 15 and 65 were gainfully employed, the respective figure for females was 6.9 percentage points lower. Coming from a 9.2 percentage points gap in 2001, the gender difference in employment has been oscillating around 7 percentage points within the last four years (see figure 2-1). The European Commission (2010) shows similar trends across the EU-27 countries and reports an average employment gap of 13.7 percentage points in 2008 and thus a significant reduction compared to 1998 (18.7 percentage points difference).

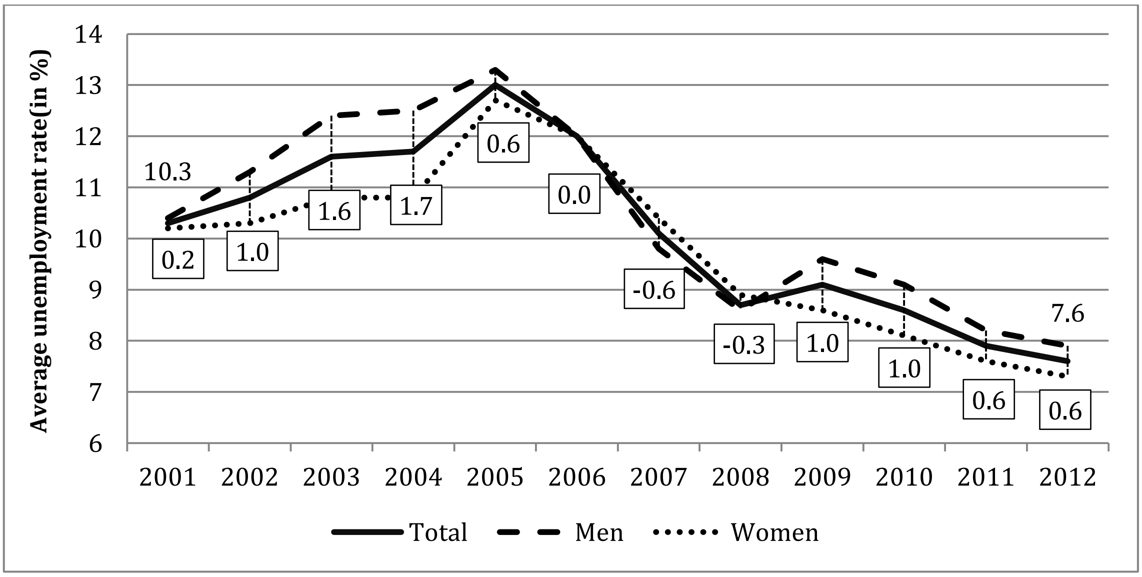
Figure 2-1: Average Employment Participation Rate of Men and Women Aged 15 and 65 Years in Germany



Notes: The employment participation rate depicts the ratio of all full-time, part-time or marginally employed among the entire population aged between 15 and 65 years.
Source: Own illustration based on BA (2013c).

Analogous to employment participation rates, figure 2-2 depicts unemployment among all employees separated by gender. After a peak in 2005 with around 13 percent, average unemployment rates decreased to 7.6 percent in 2012. Quite noticeably, the unemployment ratio of men has been exceeding the respective figure for women over the last decade except for the years 2006 until 2008. This is quite the opposite compared to the EU-27 average where women perform relatively worse compared to men (European Commission, 2010). Analyses from the BA (2012a) further reveal that transition rates in the labor market for men are higher relative to the labor market for women. The latter have a lower risk of becoming unemployed (0.8 versus 1.0 percent), but once being out of work also suffer from lower chances of finding a new job (6.0 versus 8.2 percent). Accordingly, the average unemployment duration of men (34.3 weeks) fell below the average duration of women (39.9 weeks) in 2011. Besides, the share of people who have been unemployed for 12 months or more was slightly lower for men (34 percent) than for women (37 percent).

Figure 2-2: Average Unemployment Rate of Men and Women in Germany



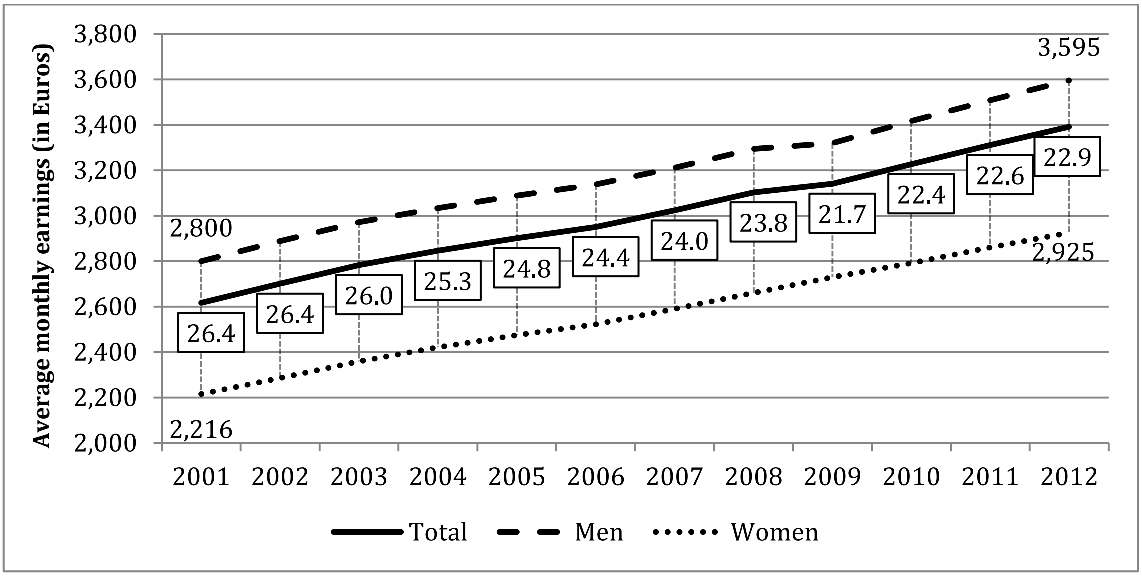
Source: Own illustration based on BA (2013a).

Comparing horizontal and vertical distributions across occupations and sectors as well as the number of working hours reveals further gender differences. While men generally work in sectors that are more prone to seasonal and economic variations, female professions are less volatile with respect to employment. For instance, in 2012, men made up more than 70 percent of all full- and part-time employees in sectors like manufacturing, transportation, mining and construction. In contrast, women were overrepresented in jobs belonging to the social sector, education, hospitality and public administration. These sectors do not only offer more stable working environments, but also permit more flexible working contracts which is underlined by a relatively high fraction of part-timers (BA, 2012c). As a consequence, the share of part-timers among women is significantly higher than among men. While every fifteenth man has reduced working times, roughly one out of three women does (BA, 2013d). These statistics go along with the EU-27 average that reveals an overrepresentation of women in part-time employment (European Commission, 2010). Gender differences also turn out to be quite substantial if firms' hierarchical levels are taken into account. Male employees are more likely to be found in high-skilled positions whereas women make up a larger fraction in skilled and unskilled jobs. However, the difference is most substantial in management positions that are almost twice as often filled by male than by female employees (Destatis, 2012b).

With respect to human capital endowments by gender, a first look at the latest figures from 2011 indicates that the share of high school graduates among the entire German population is higher for men than for women. However, the picture might be misleading. If scholastic achievements are observed separately by age cohorts, the fraction of female high school graduates turns out to be above the male share for people aged between 25-35 years and younger (Destatis, 2012b). A similar development can be observed with respect to professional qualifications. In the population, the difference between male and female unskilled workers is quite substantial (10.4 percentage points in 2011). Restricting the sample to all 25-35 year olds, though, makes this gap disappear. In the same vein, men having a degree from a professional school or university are overrepresented in the entire population, but are significantly outperformed by women among those aged between 25 and 35. Quite noticeably, all figures on human capital endowments and labor market segregation fit well into the EU-27 averages where women outperform men concerning the acquisition of university degrees but, given these superior human capital endowments, are channeled into lower-paying sectors (e.g. overrepresented in jobs such as health care and education) and hierarchical levels (e.g. underrepresented in management positions).

While the position of women in the labor market concerning educational endowments and professional qualifications has improved relatively to men, these developments thus far do not seem to have an impact on the gender pay gap. Figure 2-3 depicts average gross monthly earnings of all full-time employees working in the manufacturing and service sector. The 'raw' wage differences between men and women have been persistent over more than a decade and have only marginally declined from 26.4 percentage points in 2001 down to 22.9 percentage points in 2012. This, in fact, is clearly above EU-27 average which was reported to be 17.6 percentage points in 2007 (European Commission, 2010).

Figure 2-3: Average Monthly Earnings of Men and Women Working Full-Time in the Manufacturing and Service Sector in Germany



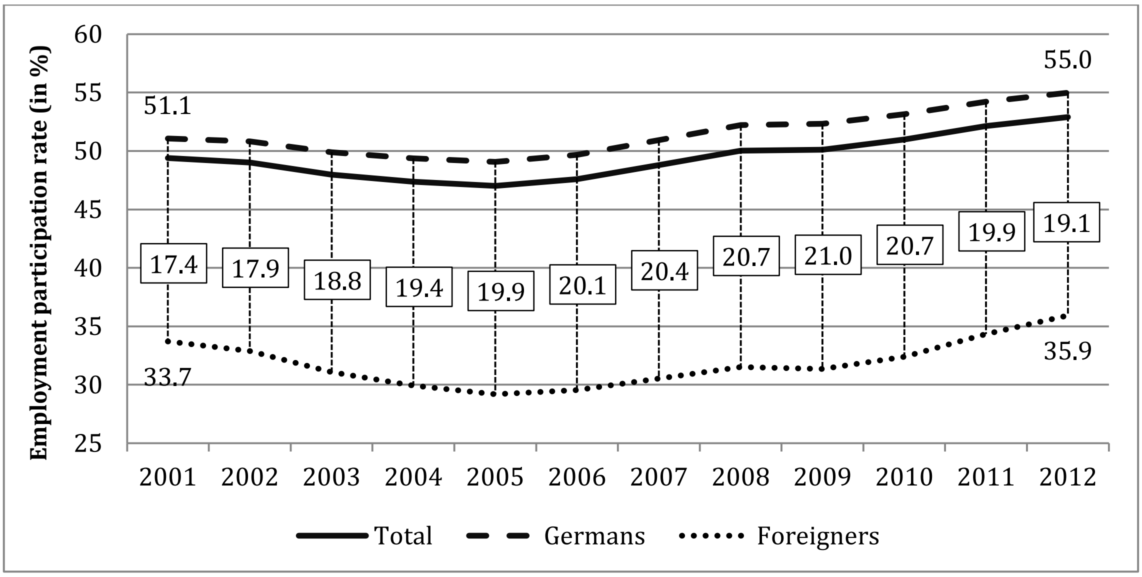
Notes: Reported earnings exclude fringe benefits.
Source: Own illustration based on Destatis (2013).

Summarizing, the German labor market shows substantial gender differences. Most importantly, women are less likely to be employed and also earn less than men. However, these stylized facts offer unconditional figures and do not take into account gender differences in e.g. horizontal and vertical distributions, working hours and human capital endowments. Therefore, they do not help to explain whether these differences are affected by supply- or demand-side factors or a mixture of both and whether hiring discrimination among others might be involved and serves as a possible explanation. Previous empirical research analyzing gender differences conditional on a variety of factors such as those mentioned above will thus be presented in chapter 3.

2.2 THE SITUATION OF ETHNIC MINORITIES IN THE GERMAN LABOR MARKET

Comparing the labor market situation of different ethnicities turns out to be a cumbersome task since it affords a proper differentiation between natives and people with a migration background. According to the BA (2012m), people possess a migration background if they either i.) do not have the German nationality, ii.) were born abroad and immigrated to Germany after 1949, or iii.) have at least one parent who was born abroad and moved to Germany after 1949. Unfortunately, administrative data in Germany primarily distinguish between nationalities rather than migration experience, i.e., only report separate figures for Germans and foreigners. In recent years, however, the requirements imposed on official statistics concerning information on migration status have been raised. Particularly the latest Microcensus offers detailed information separated by, inter alia, foreigners with own migration experience, Germans with own migration experience, foreigners without own migration experience and Germans without own migration experience (Destatis, 2012b). Accordingly, the first two are referred to as people with direct migration background while the latter constitute people with indirect migration background in the German Socio-economic Panel (GSOEP, 2012). Both statistics are also used to describe the labor market situation of migrants in this section. Nevertheless, where data are not available in detail, the figures on foreigners are used as a proxy. Aldashev et al. (2007), for example, find that the earnings prospects of people with migration background are similar to those of foreigners justifying the use of citizenship to approximate labor market outcomes.

Figure 2-4: Average Employment Participation Rates of Germans and Foreigners Aged 15 and 65 Years in Germany

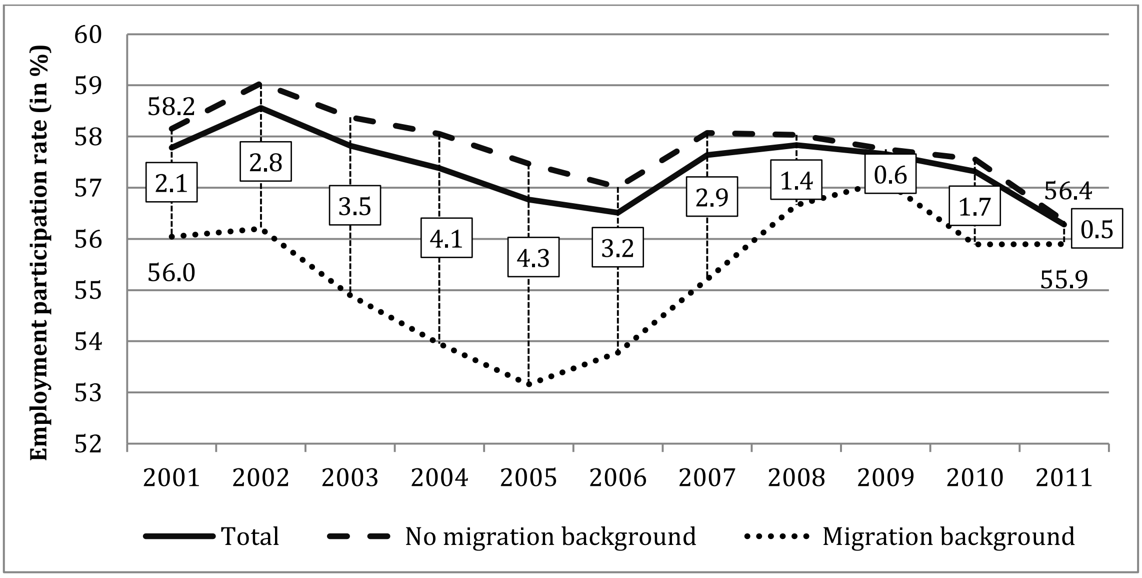


Source: Own illustration based on BA (2013c).

According to the Microcensus 2011, the population of Germany was 81.7 million of which roughly 16 million, that is almost 20 percent, either had a direct or indirect migration background. Thus, the way how such a substantial share of the society performs in the labor market is obviously of increasing importance. First turning to the participation rates, figure 2-4 shows a substantial gap between native Germans and foreigners that has been persistent from 2001 until 2012 and varied between 17.4 and 21.0 percentage points. While, except for a downturn in 2005, the participation rate of 15-65 year old Germans has constantly remained at a level above 50 percent, only 29 to 36 percent of all foreigners have been gainfully employed.

A closer look at GSOEP data for the same period of time, but with a special focus on migration status, indicates that the participation rates are quite heterogeneous across groups. Figure 2-5 suggests that ethnic differences in the share of people employed seem to be considerably smaller and have diminished over time. However, it has to be noted that immigrants most likely constitute a positively selected population in the panel so that participation rates may be overestimated.[1] In all cases, the ratios correlate and still show differences between native Germans and people with a migration history.

Figure 2-5: Average Participation Rates of People with and without Migration Background in Germany

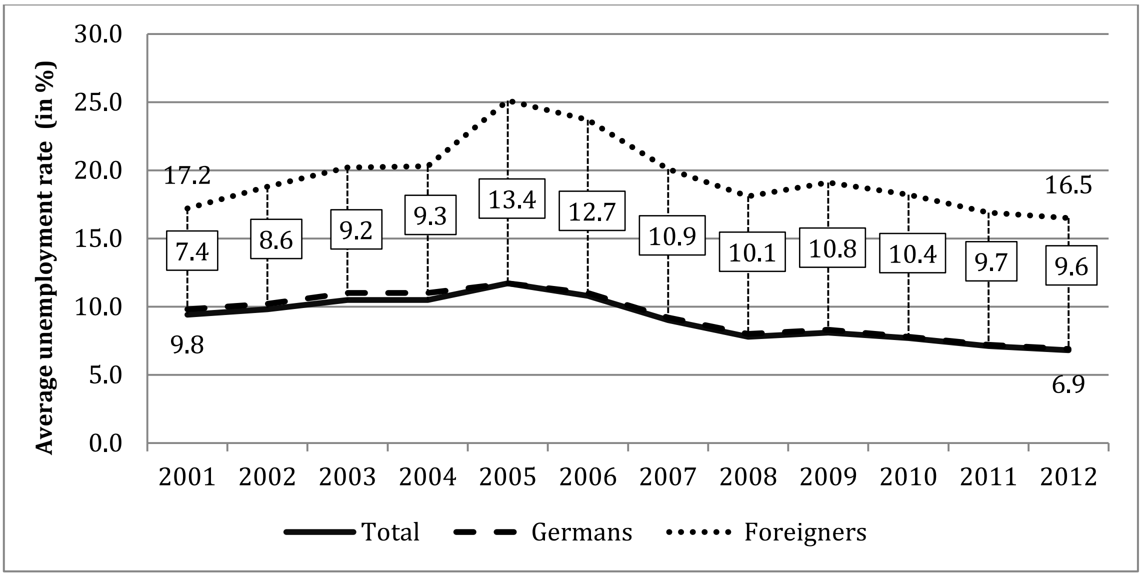


Source: Own illustration based on GSOEP data (GSOEP, 2012).

Compared to participation rates, unemployment rates separated by citizenship point in an opposite direction (see figure 2-6). Data from the BA for the last decade outline substantial and persistent differences between Germans and foreigners that reached their maximum (13.4 percentage points) during the economic downturn in 2005 and have, since, slightly decreased to 9.6 percentage points. Whereas in 2012 only 6.9 percent of all native German employees were registered as unemployed, more than twice the share of non-Germans was out of work (16.5 percent).

1 Note that apart from the participation rates of immigrants both the average and German employment ratio turn out to be higher in GSOEP data than in the statistics of the BA. I assume that especially sample selection issues drive these effects (see also Kroh, 2012).

Figure 2-6: Average Unemployment Rate of German and Foreign Employees in Germany



Source: Own illustration based on BA (2013b).

Referring to the distribution across sectors and branches, the stylized facts show that foreigners are overrepresented (relative to their share in the population) in hospitality, agriculture, transportation, construction and manufacturing and are less likely to be found in healthcare, finance and governmental occupations (BA, 2012c; GSOEP, 2012). Apart from that, the latest figures indicate that apart from an overall increase in the number of employees with reduced working hours during the last decade, among native Germans every fourth person was employed part-time in 2011, whereas among foreigners every fifth person had reduced working hours (Destatis, 2012a; GSOEP 2012; BA, 2013d).[2]

Labor market differences become most obvious if ethnic distributions at different hierarchical levels are considered. GSOEP data reveal that roughly 25 percent of native Germans work in management or high-skilled positions. In contrast, only around 17 percent of people with a migration background can be found in such positions. Apart from that, the ratio of unskilled employees is almost twice as large for people with migration background than for native Germans (GSOEP, 2012). Since hierarchical levels are closely related to educational and professional endowments, the job level differences are not at all surprising.

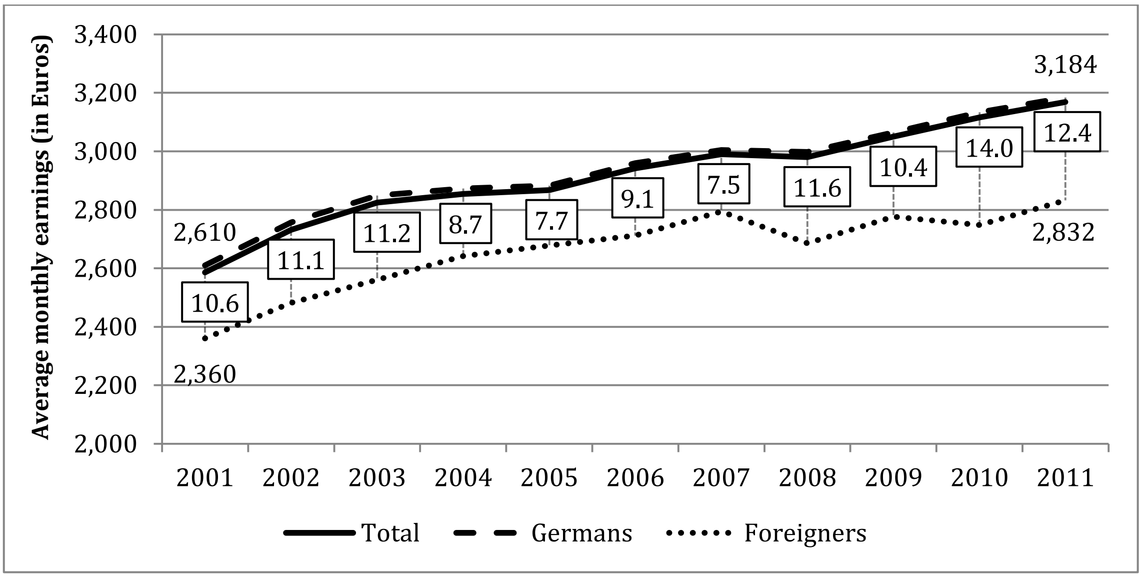
Looking at the recent figures on educational endowments conditional on citizenship and

[2] Note that the higher fraction of native German part-timers primarily goes back to a higher participation rate of native German (as compared to non-German) women who, as has been shown in section 2.1, constitute a higher share of part-time employees.

migration experience shows that native Germans have the lowest share of people with less than eight years of schooling. In contrast, a comparison of school dropout rates by different immigrant groups indicates that foreigners with own migration experience perform the worst. Simultaneously, however, they have the highest fraction of high school graduates (together with native Germans). What seems to be very odd in the first place, becomes quite reasonable if immigrant groups are considered separately. For example, it turns out that immigrants from EU countries outperform Turkish immigrants with respect to dropout and high school rates (Destatis, 2012a). This finding highlights significant variations in (pre) labor market performance among different immigrant groups. Furthermore, Microcensus data indicate that the socialization process in German society may affect performance at school as second generation immigrants perform better than first generation immigrants (Destatis, 2012a).

Similar to the distribution of educational endowments is the distribution of professional qualifications. The share of unqualified people is lowest among native Germans (15.4 percent) and highest among the foreign population that immigrated to Germany (48.5 percent). Again, a separation by selected ethnic origins shows substantial differences in the distribution of professional qualifications. Compared to the average of all people with a migration background, EU-27 immigrants have the highest fraction of university graduates and the lowest fraction of unqualified people. The latter, though, are most prominent among Turkish immigrants and German-born Turks (Destatis, 2012a).

Figure 2-7: Average Monthly Earnings of Germans and Foreigners in Germany



Source: Own illustration based on GSOEP data (2012).

As workers' (expected) productivity is closely related to their human capital endowments, the differences demonstrated above should map into a wage gap between native Germans and people with migration background. Although not accounting for additional control variables other than working hours, figure 2-7 emphasises this relationship. The average monthly earnings of Germans have exceeded foreigners' wages over the last decade. Pay differences have varied quite notably ranging from 7.7 up to 14.0 percentage points and have apparently increased during the financial crises from 2008 until 2011. However, without including a proper selection of potential covariates (such as human capital variables), the existing wage gap might be a result of both, differences in supply- and demand-side factors. Thus, more detailed evidence is required that analyzes ethnic employment and wage differences conditional on these factors. Such evidence will be provided in the next chapter.

3 LITERATURE REVIEW

This section first discusses different methods researchers apply in order to assess the presence and extent of labor market discrimination. In particular, the advantages and drawbacks of regression-based and experimental approaches are evaluated with regard to pay and hiring discrimination. Secondly, empirical research conducted in and beyond the German labor market is reviewed. Finally, empirical studies that successfully distinguish between different types of discrimination are presented in order to highlight the contrasting findings with respect to taste-based and statistical discrimination.

3.1 EMPIRICAL METHODS FOR UNVEILING DISCRIMINATION

A major challenge empiricists face when detecting actual labor market discrimination is to overcome the discrepancies between what economists call 'stated' and 'revealed' preferences. As will be shown, neither do employers truthfully state their preferences for certain demographic groups (e.g. Pager and Quillian, 2005), nor are minority workers able (or willing) to objectively evaluate the extent of discrimination they have suffered from during their working careers (e.g. Pager and Shepherd, 2008). Thus, the main objective of the following sections is to discuss whether and how different methods for unveiling discrimination tackle this challenge and present unbiased results of discriminatory treatment.

3.1.1 REGRESSION-BASED METHODS

Researchers broadly apply econometric tools such as regression techniques to microeconomic data. These data are either generated by surveys, collected by the government, provided by firms or emerge from what economists call 'natural experiments'. A prominent example that matches data from individual workers with establishment information is the Linked-Employer-Employee dataset (LIAB) which is administered and processed by the BA. Furthermore, the German Socio-Economic Panel, a longitudinal household survey conducted since 1984, and the Microcensus, a representative cross-sectional dataset covering 1% of all German households, provide rich sets of data that allow thorough analyses at the household and the individual level. Equivalents from the U.S. are, among others, the National Longitudinal Survey of Youth (NLSY), the Panel Study of Income Dynamics (PSID) (both longitudinal) and the Current Population Survey (CPS) (cross-sectional).

The surveys mentioned above obviously do not enquire employers' preferences towards certain demographic groups, nor do they ask employees whether they have been subject to any form of discrimination in the past. Both such designs would produce substantial bias as perceived disadvantages may be highly subjective and involve interviewer effects while employers, on the other hand, would certainly not admit discriminatory behavior since they would then have to face legal consequences harming their reputation (Pager and Shepherd, 2008).[3] Pager and Quillian (2005) convincingly demonstrate that personal distastes might not be truthfully stated or, to put it in their words, employers are not necessarily "walking the talk". They compare the results of a telephone survey with hiring probabilities from an audit study where black and white ex-offenders apply for a real job. Their findings suggest that firms which stated a higher likelihood of employing ex-offenders in a telephone interview actually revealed the same hiring probability than the average employer in the sample. Additionally, survey results do not show any racial differences in hiring while, in practice, blacks were significantly disadvantaged compared to white applicants (for similar findings on discrepancies between actions and stated views, see also Firth (1982)). Thus, empirical results based on self-reported behavior of employers or perceived discrimination of employees might be highly misleading and produce statistical artifacts (Pager and Shepherd, 2008).

However, even more 'objective' data do not permit the researcher to quantify the extent of direct labor market discrimination. Rather, the unexplained differentials from regression outputs can be considered a plausible proxy for discrimination, all other factors kept constant (Altonji and Blank, 1999). Blinder (1973) and Oaxaca (1973) introduced a framework that decomposes wage differentials into a fraction affected by endogenous variables such as productivity differences and differences in human capital endowments and a fraction explained by exogenous variables such as socio-economic differences. As their decomposition framework is widely considered as fundamental to research on wage discrimination and has seen a lot of derivatives and extensions (e.g. Brown et al., 1980; Reimers, 1983; Cotton, 1988; Neumark, 1988; see Oaxaca and Ransom (1994) and Silber and Weber (1999) for comparisons based on empirical data), it should briefly be discussed.

[3] In some studies, for example, subjects are asked for their past experiences with discrimination (e.g. Forstenlechner and Al-Waqfi, 2010). Obviously, these kinds of surveys are very prone to biases due to, inter alia, interviewer effects and a different understanding of what constitutes discrimination.

The basic idea is that the raw wage differential between demographic groups (e.g. blacks and whites or men and women) is attributable to differences in mean endowments, on the one hand, and differences in regression coefficients, i.e., in the returns to these endowments, on the other hand. Different rates of return imply that the market evaluates an identical set of endowments differently by demographic groups. It is this difference that can be interpreted as evidence of discrimination. In addition, any difference in the unexplained portion of the regression functions, i.e., in the shift coefficients (intercepts), also points at discriminatory behavior in either pre- or current labor market situations. Hence, using the last two measures, the fraction of discrimination among the entire wage differential can be calculated.

In order to decompose the effects of individual characteristics and the effects of discrimination, two regression models (denoted as the structural and the reduced form) for each demographic group are developed where the (log of hourly or annually) wage serves as the dependent variable. The structural model includes what is considered the full set of variables to estimate the wage regressions. This set consists of endogenous variables that provide information on e.g. education, industry, occupational position, vocational training, union membership and tenure and exogenous variables such as family background information, age, health conditions and the area of residence.[4] Some variables such as parents' education do not have a direct impact on the wage level but affect other endogenous variables such as education or career choice. For this reason a reduced form of the wage regression is estimated. Accordingly, the structural form estimates the wage conditional on the current socioeconomic situation while the reduced form estimates the wage based on the circumstances determined by birth.

In order to interpret the regression results, the between-group difference attributable to different endowments and the difference attributable to differences in the coefficients are compared. The latter provides information on how much the low-wage group (e.g. female employees) would earn if it had the same coefficients, i.e., for example, the same returns to schooling, as the high wage group (male employees). As explained above, differences in the estimated and the shift coefficients between the two groups indicate discrimination which can be expressed as a ratio of the raw wage differential in both models. Deducting the ratio of the reduced model from the ratio of the structural model yields the fraction of

[4] Note that the number and the nature of the independent variables are highly dependent on the data available. The variables listed here are taken from Blinder (1973).

discriminatory treatment that is based on unequal opportunities in access to, for instance, educational or occupational traits. Consequently, the decomposition technique enables researchers to break down the raw wage differential into a fraction that can be attributed to inferior endowments in the variables determined by birth, into a fraction that reflects direct discrimination in the wage setting process and into a fraction that accounts for discriminatory treatment in achieving the endogenous variables, i.e., pre-market discrimination.

The wage decomposition can well be explained by the studies of Blinder (1973) and Oaxaca (1973). The former uses data from the PSID survey in order to investigate the reasons for racial and gender pay differentials in the U.S. Besides actual hourly wage rates, the dataset includes detailed family background information which permits the dichotomization between endogenous and exogenous variables and thus a decomposition of the regression estimates. With respect to the 50.8 percent wage premium of white compared to black workers, Blinder finds that 30 percent are attributable to the latters' inferior endowment in socio-demographic characteristics such as parents' education or residential area of birth, 40 percent point at direct discrimination in the wage rates and 30 percent account for blacks' poorer opportunities in access to e.g. education. In contrast, he shows that the wage differential between white male and female employees (which adds up to 45.8 percent in favor of the former) is not based on differences in family background characteristics, but on differences in the regression and shift coefficients of the structural regression, i.e., direct wage discrimination (about two thirds of the raw differential) and inferior access of females to education and certain occupations (about one third of the raw differential).

The study by Oaxaca (1973) analyzes gender differences in hourly wages of white and black workers using a subsample of the Survey of Economic Opportunity (SEO) from 1967. He finds a gender pay gap of 54 percent in case of whites and of 49 percent in case of blacks, respectively. Decomposing these results reveals that discrimination accounts for 58.4 and 55.6 percent of the entire wage gap. More precisely, 19.3 (38.0) percent of the white (black) pay differential can be attributed to discriminatory treatment of females in access to certain occupations while 39.1 (17.6) percent account for differential evaluations of mean individual characteristics and (unexplained) differences in the shift coefficients. Hence, discrimination is the major source of the gender pay gap. Nevertheless, much of the wage differential does not stem from unequal pay for equal work, i.e., direct pay discrimination, but occupational segregation with women concentrating in lower-paying (service) jobs.

Independent of the econometric strategy, Altonji and Blank (1999) claim that the unexplained wage gap serves as an adequate proxy for labor market discrimination, but does not present a very direct form to measure systematic group differences.[5] Two main factors may bias the unexplained wage differential. Firstly, if occupational sorting and human capital investments in education and training were endogenous, i.e., influenced by (pre-) labor market discrimination, the unexplained gap would understate the extent of discrimination since it was partly captured by other independent variables included in the regression model. Whether the independent variables are affected by discrimination or whether differences in endowments simply represent different preferences is crucial, though very hard to disentangle by means of regression techniques (and also not fully accounted for by Oaxaca and Blinder's structural and reduced model). For example, women may dispose of inferior human capital endowments because they did not have equal opportunities in acquiring such endowments. On the other hand, they may invest less in their own human capital, may not apply for jobs in male-dominated occupations or may not aspire for senior positions because they anticipate unequal opportunities and adapt their career choices accordingly. Also, this could be a rational reaction when expecting a shorter career length (due to e.g. child-bearing activities).

Secondly, the extent of discrimination would be overstated if productivity relevant characteristics were omitted from the wage regression, i.e., included in the error term. Oaxaca (1973) admits that the estimated effect of discrimination crucially depends on the choice of the independent variables and that the unexplained gap may eventually disappear if a sufficient number of wage determinants is included. Farkas and Vicknair (1996), Neal and Johnson (1996) and Heckman et al. (2006), for instance, find a significant decrease or even complete disappearance of the gender pay gap if cognitive and non-cognitive abilities and skills other than schooling are incorporated in the wage regression. Yet, their results are refuted by Carneiro et al. (2005) and Lang and Manove (2011) who show that the inclusion of education causes the unexplained differentials to reemerge.[6]

5 That is why recent studies sometimes use terms like "residual gap" (Fransen et al., 2012) or "unobservable" component of earnings (Lee and Lee, 2012) instead of "discrimination" as a more neutral way to describe the unexplained wage gap.

6 Charles and Guryan (2011) also criticize the linear relationship assumed in models of the decomposition framework and point out that the impact of skills and abilities on labor market outcomes are most likely nonlinear and of unknown functional form which may cause substantial bias when assessing the extent of discrimination.

This debate outlines that regression-based findings on wage discrimination are very sensitive to alternative model specifications. Unfortunately, administrative data generally fail to provide detailed information on the production process and workers' productivity. A way to overcome these problems may be the use of insider data including detailed productivity information at an individual level. Such data, however, are rare, are commonly subject to strict data protection requirements and, of course, do not allow generalization.

Turning back to the findings by Blinder (1973) and Oaxaca (1973), a substantial fraction of the gender and racial pay gap can be attributed to occupational sorting, i.e., a systematic variation of demographic groups across jobs and industries. Even though the decomposition framework permits a thorough analysis of wage differentials and provides consistent (though potentially biased) evidence on pay discrimination, it may not be a suitable tool for assessing discrimination at an even earlier stage of the employer-employee interaction, that is, during the hiring process, or, later, during promotions to higher hierarchical levels (e.g. Petersen and Saporta, 2004; Charles and Guryan, 2011).[7]

The stylized facts from the German labor market demonstrate that demographic groups systematically differ regarding their distribution across occupations and hierarchical levels. In other words, labor markets are often horizontally and vertically segregated. Reasons for that not only go back to employers' discriminatory behavior. In fact, supply-side determinants that differ at the entry stage into employment as well as at later career stages may also have an impact on different employment outcomes across demographic groups (e.g. Lang and Manove, 2011). Analogously to the discussion on wage differentials, endogeneity issues play an important role as regression-based analyses lack evidence on the counterfactual situation, i.e., a market without discrimination (Harrison and List, 2004). Demographic groups may self-select into different occupations and hierarchical levels as a response to pre-labor market or anticipated discrimination, or simply because they have different preferences that, in turn, may be induced by societal role models (Eberharter, 2012). In addition, other factors such as the use of referral networks (e.g. Petersen et al., 2000; Ioannides and Loury, 2004; Caliendo et al., 2011), performance in

[7] Unequal opportunities in access to higher hierarchical levels, i.e., a glass-ceiling effect, have been documented in the seminal work by Lazear and Rosen (1990) and reproduced in various institutional settings (e.g. Weinberger, 2011; Johnston and Lee, 2012; Gobillon et al., 2012). Petersen and Saporta (2004) use the term "allocative discrimination" to account for the fact that discriminatory treatment may simultaneously be observed at various stages of the employer-employee interaction.

competition (Gneezy et al., 2003; Jurajda and Münich, 2011) and different dropout rates in the course of the hiring process (Arvey et al., 1975) may affect employment outcomes across demographic groups.

If not appropriately considered in the analyses, these factors would significantly bias findings on differential treatment and thus over- or underestimate the extent of discrimination. Consequently, Lang and Lehmann (2012: 8) point out that separating the effects of discrimination in the recruitment process from any other effects embedded in applicants' characteristics and their job search behavior may be even more challenging (compared to wage regressions). One major issue is data availability. Unlike wages, administrative data on unemployment rates and duration, entry and exit from unemployment as well as labor market participation contain only few, if any, individual-level information. Furthermore, company data from application processing are hardly available (exceptions are Arvey et al. (1975) and Petersen and Saporta (2004)) and, if so, only report who is hired, but lack information about who gets rejected. By generating (own) experimental data, however, researchers control for most of the above-mentioned supply-side differences and are thus able to directly identify discrimination in the recruitment process. Yet, experiments also face methodological challenges which will be discussed in the following.

3.1.2 EXPERIMENTS

Experiments allow controlling for any joint effects in the independent variables and try to minimize any bias originating from unobserved heterogeneity in workers' characteristics. The goal is to create a counterfactual situation in order to separate a treatment effect, i.e., observe the outcome of an untreated subject had it been treated. Thus, compared to administrative data, experiments provide a rather direct way to investigate discrimination in the labor market and allow generating data for empirical questions that would most likely have remained unanswered if only administrative data were available. In contrast, they enable the researcher to adequately match candidates and implement truly exogenous differences (e.g. of applicants' gender) that are unaffected by any endogenous variables determined in the field (Falk and Fehr, 2003). For example, male and female applicants may anticipate discrimination in jobs predominately occupied by the opposite sex which would discourage them from applying. Alternatively, only a highly-selected population, e.g. only high quality candidates, applies for non-stereotyped jobs. Such selection effects would significantly affect gender differences. Besides, pre-market disadvantages in the attainment of educational endowments may encourage occupational herding. If, for instance, women were systematically discriminated in Math which would negatively affect their grades, lower employment rates in technical occupations where Math grades are more important than, say, grades in Politics, would be a rational consequence rather than hiring discrimination. Being able to directly control these mechanisms is a major advantage of experiments. A direct test of discrimination would match two otherwise equally equipped candidates that apply for the same job and only differ with respect to one demographic characteristic. This procedure not only creates a treatment and a control group, but the experimental setting also permits replicability of the findings.

Harrison and List (2004) list various methods to create the counterfactual. These methods are either econometric tools used together with administrative data such as propensity score matching and instrumental variable regressions or rely on natural or controlled experiments. In line with their name, natural experiments compare the outcomes between a treatment and a control group in a naturally occurring environment. Thus, subjects can be observed in a real context that involves real stakes. Unfortunately, researchers do not come across such data very often (for an exception of this, see e.g. Goldin and Rouse (2000) and Wozniak (2012)). This in turn calls for the implementation of experiments that construct a control group via randomization.

3.1.2.1 LABORATORY EXPERIMENTS

Where alternative data are not available or do not contain sufficient information to draw causal inferences, data may be generated in the laboratory.[8] Here, researchers have the possibility to observe exogenous ceteris paribus changes as subjects' preferences are induced by controlled effort cost and production functions. Thus, endogeneity problems can be dealt with to a certain extent which allows the experimenter to clearly identify e.g. factors influencing the decision. Additionally, biases due to information asymmetries and unobserved activities such as sabotage can be excluded or separated (by observing e.g. outcome differences between anonymous and face-to-face interactions) as the experimental framework and subjects' communication is under the researcher's control.

8 Researchers distinguish various forms of laboratory experiments including scenario and neuroeconomic experiments. Here, only the general advantages and disadvantages of laboratory experiments are elaborated. A thorough discussion would be beyond the scope of this thesis and can be found in e.g. Harrison and List (2004).

In the same vein, the underlying circumstances are known and can be influenced by the researcher. Such circumstances include the number of subjects involved in an interaction and whether this interaction is repeated or just one-shot. Lastly, the thorough control also permits replicability of the experiment and its results, which facilitates the verification or falsification of the hypotheses developed (Falk and Fehr, 2003).

However, laboratory experiments face several objections that need to be carefully addressed. Firstly, the majority of laboratory experiments use students as the subject pool because they are generally easy to get access to, do understand the underlying rules and have rather low opportunity costs. Critics argue that students may not be representative, may lack experience with certain tasks and provide little socio-demographic variability. Conversely, the incomplete control over recruiting of subjects from outside university carries the risk of further sample selection and attrition bias. Research comparing the results from different subject populations varies with respect to the quantitative findings, but shows strong similarities in the qualitative patterns (Falk and Fehr, 2003).

Secondly, commodities chosen in an experiment might not appropriately represent those in the field and, as a consequence, might cause subjects to behave differently. In other words, relatively low incentives may induce different behavior as opposed to rather high incentives (such as monetary payouts or legal consequences). However, Camerer and Hogarth (1999) outline that subjects' behavior is very little if at all dependent upon changes in expected earnings. Besides, any reservations about the size of the stacks may be tackled by conducting experiments in poor countries where the stakes are more meaningful to the subjects (Falk and Fehr, 2003).

Thirdly, a small number of observations may limit the applicability of parametric data analysis techniques and may fail to produce statistically robust results. These limitations, however, are rather weak since observations can be increased at any time. Moreover, researchers have engaged in large scale experiments that allow a comparison to the results from small sample studies (see Falk and Fehr (2003) for prominent examples).

Fourthly, tight control may carry the risk that subjects behave differently when they are observed, i.e., either feel social pressure to behave in a certain manner (known as the 'Hawthorne effect') or act how they believe the experimenter wants them to act (the so-called 'experimenter effect') (Harrison and List, 2004).

Fifthly (and probably most commonly mentioned in the literature), criticisms have been raised concerning the internal and external validity of laboratory experiments. While internal validity may be implemented by a proper experimental design, external validity includes more general objections on whether the inferences drawn prevail outside the laboratory. Realism can for example be added by conducting real effort experiments and providing a real context (Falk and Fehr, 2003). More convincingly (or at least complementary) and beneficial to the generalizability of the findings from laboratory experiments, however, may be the implementation of field experiments.

3.1.2.2 FIELD EXPERIMENTS

The nature and design of field experiments is quite similar to laboratory experiments. Harrison and List (2004) classify field experiments as artefactual, framed or natural. While the first two have an informed nonstandard subject pool, natural field experiments observe uninformed subjects following their every-day business. So, ideally, external validity is maximized by the field environment and internal validity is maintained by a sufficient set of controls. Furthermore, natural experiments guarantee that subjects do not only make simple statements, but actually (re)act according to their preferences (recall the initial discussion about 'stated' and 'revealed' preferences).

Field experiments investigating hiring discrimination can be designed in various ways. A strand of literature has used matched-pair experiments denoted as audit or correspondence testing in order to find differences in access to employment conditional on a treatment variable such as gender or ethnic origin. These methods try to control for any effects that stem from differences in workers and workplace characteristics by matching equally qualified pairs of job candidates who apply for the same position. The applications only differ with respect to one major characteristic which distinguishes the majority from the minority group where the former (latter) generally represents a higher (lower) share of employees in the respective labor market segment. Based on firms' aggregate callbacks to each group, the prevalence of differential treatment can be tested. A callback is generally referred to as a situation where the employer promotes the candidate to the next stage of the recruitment process which could be, for example, a job interview. Since individual characteristics are controlled for, differences in market expectations, preferences and social ties (networks) can be ignored and the effects of group-specific selection into certain occupations and hierarchical levels can be excluded, any aggregate callback differences that turn out to be statistically significant can be attributed to discriminatory practices on behalf of employers (Riach and Rich, 2002; Pager, 2007).

Prior matched-pair studies use different measures to report the extent of discrimination. The main differences stem from the treatment of firms that do not call back any of the applicants. Riach and Rich (2002) discuss how the results from correspondence and audit studies should be reported and interpreted. They argue that employers rejecting both applicants should be treated as non-observations as it is not clear to the researcher whether an actual evaluation of the candidates has taken place or whether the vacancy was already filled which would have made such an assessment obsolete. Thus, they recommend calculating the net discrimination rate by subtracting the number of occasions where only the majority candidate received a callback from the number of occasions where only the minority candidate received a callback conditional on employer's callback to at least one candidate. Consider (1) as the total number of matched pairs, (2) as the number of cases where neither of the candidates received a callback, (3) as those occasions where at least one candidate received a callback, (4) as situations where both received a callback, (5) as 'majority-only' callbacks and (6) as observations of 'minority-only' callbacks, this formally yields:

$$\text{net discrimination rate} = \frac{(5) - (6)}{(3)}.$$

The gross discrimination rate, on the other hand, considers all employers addressed which makes it a less conservative measure of differential treatment. Hence,

$$\text{gross discrimination rate} = \frac{(5) - (6)}{(1)}.$$

Analogous to the net and gross discrimination rates, dividing the ratio of majority callbacks by the ratio of minority callbacks among those employers that gave at least one candidate a positive response yields the odds ratio:

$$\text{odds ratio} = \frac{\frac{(4) + (5)}{(3)}}{\frac{(4) + (6)}{(3)}}.$$

Including the observations of those firms ignoring both applications (2), yields the following success ratio:

$$\text{success ratio} = \frac{\frac{(4) + (5)}{(1)}}{\frac{(4) + (6)}{(1)}}.$$

The measures presented are the same independent of whether audit or correspondence testing is applied. However, both methods differ with respect to their experimental design. While the former train real-life applicants such that similar behavior during telephone and job interviews is evoked, the latter only send out résumés of fictitious applicants. Thus, audit studies allow the researcher to evaluate discriminatory practices at every stage of the hiring process. Heckman and Siegelman (1993) and Heckman (1998), however, point out the problems that occur due to demand effects and a lack of control, especially during a personal job interview. The correspondence method, in comparison, gives unaltered evidence of unequal treatment since it focuses on written applications that minimize unobserved heterogeneity. The major shortcoming of this method is that observations are confined to the first step in the recruitment process. Nevertheless, this problem seems to be less severe. In fact, reviewing the results of previous audit studies, Riach and Rich (2002) show that discrimination is most evident before personal contact takes place, i.e., when written applications are assessed.

Further criticisms highlight the problem of effective matching (Heckman and Siegelman, 1993; Heckman, 1998). In this respect, Harrison and List (2004) note that partial matching may sometimes be worse than no matching. For example, if men and women are expected to have the same average productivity, but different in-group productivity variances, it depends on the employer's threshold level which group he prefers. If the threshold level is high, it is rational to choose a member of the higher variance group since a higher fraction will meet the high standard. Conversely, if the threshold level is rather low, the lower variance group should be favored as they are less likely not to meet firms' requirements. In other words, candidates that look homogenous on any other characteristics except for the one treated (e.g. gender or ethnicity) are not necessarily perceived as being equal which might cause bias in the regression estimates produced. Consequently, study designs should include variations in other individual characteristics to allow for an investigation of the treatment effect conditional on other independent variables.[9]

Another objection to be addressed has to do with hidden connotations of individual characteristics such as names and profile pictures. Correspondence studies usually use the former as an indicator of candidates' gender and/or ethnic affiliation. Typically, name registers are consulted to choose a set of (gendered) native and ethnic-sounding names. However, Fryer and Levitt (2004), for instance, show that names may not only convey information on group membership, but might be associated with socioeconomic status. Further studies reveal that names are used to infer people's age, attractiveness and

[9] See Neumark (2012) for a thorough discussion of implicit assumptions (embedded in correspondence studies) on group differences in the unobservables and their effect on employment outcomes.

intelligence (e.g. Rudolph and Spörrle, 1999; Rudolph et al., 2007; Cotton et al., 2008; Arai and Skogman Thoursie, 2009; Watson et al., 2011). These findings indicate that employers may form productivity beliefs based on applicants' names rather than their gender or ethnic origin which would dilute experimental control and make the separation of an unbiased treatment effect impossible. Therefore, a proper correspondence design requires the implementation and control of name effects (see e.g. Bertrand and Mullainathan, 2004) for within-group name-based outcome differences in a correspondence setting). Similarly, the attachment of a profile picture which is common in the German labor market needs to take into account beauty effects as, inter alia, investigated by Hamermesh and Biddle (1994), Mobius and Rosenblat (2006) and Rooth (2009). Especially if differential treatment based on age or gender is evaluated, beauty controls need to be considered. This could be done by implementing a variety of profile pictures that are then included as dummy variables in the econometric analyses.

Additional challenges have their origin in the nature of the correspondence method and the recruitment practices in general. Since the hiring process within a firm is like a 'black box' to the researcher, employers' responses do not reveal whether callbacks are based on individual or group decision making. While the latter permits social learning, the former does not. This, however, may lead to systematically different employment outcomes across groups and may thus affect the extent of discrimination. Apart from that, the type of jobs suitable for audit and correspondence studies are limited. Senior positions, for instance, require prior professional experience which is hard to signal due to a lack of credible references. Besides, the longer the employment history and the more credentials are provided, the higher is the treatment effect bias unintentionally created by unobservable productivity information. Also, both audit and correspondence methods are suitable for revealing discrimination in recruitment, but are rather inappropriate procedures for uncovering discrimination in other domains of the employee-employer interaction such as access to training, promotions and lay-offs.

Lastly, researchers criticize the deceptive nature of audit and correspondence studies. Riach and Rich (2004) deal with the question of whether these methods are ethical and represent a legitimate research practice. Referring to benefits and drawbacks of alternative methods presented above, they trade off the disadvantages some demographic groups have from discriminatory practices against the economic costs employers face when processing fictitious applications. They conclude that an application of the matched-pair experiments using fictitious applications is well justified if certain quality standards implicitly agreed upon throughout the history of these methods are met (see e.g. Riach and Rich, 2002). Above all, this includes promptly and politely withdrawing from the applications in case of employers' callbacks.

In the labor market, Fidell (1970), Levinson (1975) and Firth (1982) were the first to use audit and correspondence methods to study gender differences in hiring, while Jowell and Prescott-Clarke (1970), Newman (1978) and Firth (1981) conducted matched-pair testing to assess ethnic discrimination in the recruitment process. Later, the use was extended to other socio-demographic characteristics such as age (Bendick et al., 1999; Lahey, 2008; Riach and Rich, 2006a, 2007b, 2007a), religious affiliation (Banerjee et al., 2009; King and Ahmad, 2010; Siddique, 2011), obesity and attractiveness (Agerström and Rooth, 2011; Rooth, 2009; López Bóo et al., 2013; Ruffle and Shtudiner, 2013), sexual orientation (Ahmed and Hammarstedt, 2009; Weichselbaumer, 2013), leisure time activities and physical fitness (Rooth, 2011), maths skills (Koedel and Tyhurst, 2012), criminal records (Baert and Verhofstadt, 2013) and unemployment experiences (Falk et al., 2005; Oberholzer-Gee, 2008). In addition, domains other than the labor market were addressed (see e.g. Ross and Turner (2005) for the housing and Gneezy and List (2004) for the product market). Detailed results of more recent studies on gender and ethnic discrimination will be presented in the subsequent section.

Summaryzing, a review of the methods applied in the empirical literature on labor market discrimination shows that regression-based studies prevail with respect to the analysis of wage differentials while experimental approaches are most commonly used when assessing differences in hiring. In the context of the latter, field experiments have proven advantageous compared to laboratory experiments as well as administrative and/or survey data. They provide a real context, minimize selection and firm specific effects and do not depend on different perspectives, expectations and information available to the respondents. In addition, they most strongly promise to reveal employers' true rather than their stated preferences. Due to these advantages, the correspondence method will be applied for data collection in this thesis.

3.2 Empirical Evidence on Different Labor Market Outcomes by Gender and Ethnic Background

This section discusses empirical findings on wage and hiring differences by gender and ethnic origin. Unlike the stylized facts from chapter 2 that display largely unconditional employment and wage differences, the studies reviewed below control for confounding effects, decompose the existing gaps and try to identify the prevalence and extent of labor market discrimination against women as well as racial and ethnic minorities. Of course, the literature presents only a snapshot of the available work and focuses on seminal papers as well as the most recent publications. Findings from outside the German labor market are largely restricted to research in the U.S. while empirical studies on the German labor market are presented separately.

3.2.1 DIFFERENT LABOR MARKET OUTCOMES BY GENDER

As direct evidence on gender hiring discrimination in the German labor market is very limited, the following subsections focus on related literature that provides supply- and demand-side explanations for the prevailing gender differences inside and outside the German labor market. First, the findings on wage and, afterwards, the findings on employment disparities are presented where labor market discrimination is identified by a variety of methods as discussed in section 3.1.

3.2.1.1 FINDINGS ON GENDER WAGE DIFFERENCES OUTSIDE THE GERMAN LABOR MARKET

Economists studying the causes and consequences of the gender wage gap can look back on a long history of empirical research of which some widely cited papers are presented below. Decomposing the factors impacting on median hourly and weekly earnings of male and female full-time employees, Blau and Beller (1988) find a narrowing gender pay gap in the U.S. In particular, cross-sectional estimates from 1971 and 1981 CPS data show an increase in the female-male wage ratio. Their results suggest that, firstly, a decline in direct wage discrimination and, secondly, changing gender roles may account for this trend. As a result, women have increased labor force participation which has, in turn, increasingly fostered their own and employers' decision to invest in general and specific human capital. On the other hand, occupational segregation and women's lower returns to schooling are found to mitigate the reduction in wage differentials. In the same vein, Blau and Kahn (1997) find with PSID data that both relative improvements of women's human capital endowments as compared to men's and a decline in discrimination against female employees have led to a decrease in the U.S. gender wage gap between 1979 and 1988. In particular, women's average labor market experience increased relative to men's, they benefitted from changes in occupational patterns that strengthened the role of jobs in the service sector where women were overrepresented and they were less affected from real wage losses due to deunionization.[10] These effects outweighed changes in the wage structure that particularly disfavored low-skilled workers among whom women were overrepresented. As both labor supply and demand of females increased, the overall progress in particular for high-skilled women was slowed down (O'Neill and Polachek, 1993).[11] Consecutive analyses point toward a slowdown in the convergence of wage differentials between men and women in the 1990s as compared to the 1980s. Comparing hourly earnings from three waves of PSID data (1979, 1989 and 1998) reveals that while women's human capital endowments continued to increase and returns to skills remained constant, developments towards a rather equal gender distribution across occupations stagnated. Where women entering the labor market were a positively selected population in the 1980s, changes in women's labor force structure might have provoked systematic variations in unmeasured characteristics slowing down the decline in the gender pay gap (e.g. Darity and Mason, 1998; Blau and Kahn, 2006; Mulligan and Rubinstein, 2008).[12]

As the studies presented above indicate, gender differences in human capital endowments are generally held responsible for a substantial part of the gender pay gap. One reason why these differences occur can be found in women engaging in childbearing and -rearing activities. Anticipating parental leave may deter both employers and employees from undertaking human capital investments thus leading to systematic gender differences. Furthermore, employment interruptions associated with motherhood create a relative gender gap in accumulated labor market experience. As a result, women earn less than men which is why the literature often refers to the so-called 'family' or 'motherhood' gap (e.g. Mincer and Polachek, 1974; Miller, 1987; Korenman and Neumark, 1992; Waldfogel,

10 In general, empirical analyses indicate that the degree of unionization is negatively related to the gender pay gap (e.g. Even and Macpherson, 1993; Doiron and Riddell, 1994).

11 Blau and Kahn (1992, 1999, 2000, 2003) show that changes in the wage structure not only explain within-country variations in the gender pay gap over time, but also help to explain cross-country wage differences. Their findings consistently indicate that women tend to be "swimming upstream". While human capital endowments have narrowed, the returns to high skills (which women were, on average, inferiorly endowed with) increased relative to the returns to low skills.

12 For similar results from meta-analyses using studies with data from inside and outside the U.S., see e.g. Jarrell and Stanley (1998, 2004) and Weichselbaumer and Winter-Ebmer (2005). Reassessing the results by Blau and Kahn (2006), Lee and Lee (2012), however, offer quite surprising insights. They find that the reported decrease in the gender pay gap may be prone to measurement error and, in fact, be smaller than suggested. The reason is that the earnings variable systematically differs depending on whether the survey is self-reported or proxied by another household member. As more women have become household leaders over time and have thus self-reported their earnings in the survey, gender differentials may have been systematically biased. These findings underline the sensitivity of survey data and the necessity of being aware of any potential sample selection effects.

1997, 1998; Erosa et al., 2002; Brown et al., 2011; Theunissen et al., 2011; Belley et al., 2012; Glauber, 2012).[13]

Even though former studies interpret the unexplained gender wage differentials as appropriate evidence for discriminatory treatment, unobserved heterogeneity of productivity-related characteristics as well as problems from omitted variable bias remain. Madden (1987) carefully addresses these issues and reveals that gender differences do not occur as a result of (unobservable) investment decisions, but due to gender discrimination in access to training. Contrasting this, Kim and Polachek (1994) show that addressing unobserved heterogeneity significantly decreases the unexplained gender wage gap. They built a balanced panel from PSID data with more than 2,600 individuals over a course of 12 years (1976-1987) and estimate fixed and random effects models. Their main finding demonstrates that adjusting for worker heterogeneity results in a decrease of the unexplained wage differential from 40 to 20 percent. Addressing endogeneity that stems from e.g. the decision (not) to take up employment (because the wage offers are below workers' reservation wages) decreases the unexplained gender pay gap even further to less than 10 percent.

Apart from gender differences in labor force participation rates, horizontal and vertical segregation remain persistent factors influencing the gender wage differential. Even though women's earnings have grown faster than men's due to a shift to higher-level occupations and steeper wage growth within job levels (Gittleman and Howell, 1995), women still tend to be overrepresented in low-paying industries and low-skilled occupations (e.g. Darity and Mason, 1998; Blau and Kahn, 2000). Put differently, it is the gender composition across industries and jobs that significantly contributes towards explaining the gender wage gap (e.g. Sorensen, 1990; Groshen, 1991; Fields and Wolff, 1995). However, empirical estimates on the extent of this crowding effect yield varying results depending on the data and aggregation of occupational controls (e.g. Dolton and Kidd, 1994; Bayard et al., 2003). While some researchers find that remuneration within job-cells, i.e., the same occupations within the same establishments, only marginally differs across the sexes (e.g. Groshen, 1991), others reveal that women still earn significantly less than men even within narrowly defined jobs at the same employer (e.g. Gupta and

[13] Furthermore, another strand of research finds women to trade in more flexible and family-friendly working conditions for lower wages and promotion probabilities which circulates under the term 'compensating differentials' in the literature (e.g. Filer, 1985; Glass, 1990; Glass and Camarigg, 1992).

Rothstein, 2005; Bayard et al., 2003). Using longitudinal data, Macpherson and Hirsch (1995) further show that as much as two thirds of the gender composition effects on wages are endogenous and can be explained by occupational characteristics and unmeasured skill and taste differences.[14] Gender wage differences, though, have not only been found to arise from occupational crowding, i.e., the so-called 'glass door' effect, but may also be caused by segregation across hierarchical levels, commonly referred to as the 'glass ceiling' effect. Quantile regression results from Europe, the U.S. and Canada indicate that the gender pay gap is most prominent in the upper tail of the earnings distribution (Arulampalam et al., 2007; Chzhen and Mumford, 2011; Weinberger, 2011; Javdani, 2013).

In line with the findings by Madden (1987), Lips (2013) argues that the pre-market choice to invest in human capital cannot be considered as gender neutral, but may be affected by a gender-specific component that itself might entail discrimination. In contrast, women may voluntarily invest less in pre-market human capital than their male counterparts as they have different preferences for such investments. In order to address these opposing approaches (assigning labor market differences to either the demand or supply side), in the last decade, researchers have been trying to incorporate variables that reflect gender differences in (wage and career) expectations (e.g. Filippin and Ichino, 2005; Chevalier, 2007; Grove et al., 2011; Schweitzer et al., 2011; Frick and Maihaus, 2013), (educational, job choice, risk and competitive) preferences (e.g. Bowles et al., 2001; Croson and Gneezy, 2009) and non-cognitive skills (e.g. Heckman et al., 2006; Müller and Plug, 2006). The explanatory power of these variables, however, seems to vary quite substantially. As a consequence, the effects from changing social attitudes (about the role of women in society) on the gender wage gap also remain rather suggestive.

One prominent exception is the study by Backes-Gellner et al. (2013) which assesses the relationship between regional differences in the attitude towards women in the labor market and wages. Therefore, the authors use the Swiss Earnings Structure Survey (ESS), an employer-employee linked dataset, and approval rates to two amendments in the Swiss constitution (1981 and 2000) promoting gender equality in the labor market (and thus make use of variations in people's revealed rather than stated preferences). Most notably,

[14] Some authors also report a systematic shortfall of wages in female- as compared to male-dominated jobs although skill requirements and other wage-relevant factors are comparable (e.g. England et al., 1988). This phenomenon is also referred to as 'valuative discrimination' in the literature (e.g. Petersen and Saporta, 2004). However, the documentation is difficult and empirical papers produce rather mixed results (see the discussion by Tam (1997), England et al. (2000) and Tam (2000)).

they find that within-firm remuneration varies across cantons and gender. The gender pay gap is larger in cantons with a lower approval rate and explains about 50 percent of the within-firm variation of the gender pay gap, ceteris paribus. Similarly, Fortin (2005), conducting cross-country comparisons between 25 OECD countries with data from the World Values Surveys (which, in turn, only includes information on people's stated preferences), establishes a relationship between egalitarian views on gender issues in the labor market and actual employment differences. While recent age cohorts have a rather liberal attitude and support labor market equality, perceptions of women as homemakers are found to cause a slowdown of the narrowing gender wage gap. Both of the aforementioned studies can thus be regarded as strong evidence for a linkage between societal role models and the gender pay gap where the former may substantially impact on the latter.

3.2.1.2 FINDINGS ON GENDER WAGE DIFFERENCES IN THE GERMAN LABOR MARKET

Given the extensive research on the reasons for the gender pay gap, empirical studies focusing on the German labor market are relatively scarce. Finke (2011) uses the Structural Earnings Survey (SES) 2006, a dataset including rich information on gross hourly wages as well as socio-economic and job characteristics, to investigate the gender pay gap in Germany. Comparing more than 1.5 million male and female employees, she finds a raw wage differential of 22.2 percent of which roughly two thirds (62.7 percent) are explained by differences in endowments while 8 percent of the wage gap remain unexplained. Looking at the variation explained by the regression model, differences in jobs and hierarchical positions have the largest impact (44.1 of 62.7 percent). Concerning the unexplained part, the major effect is captured by the constant which, on the one hand, may stem from direct pay discrimination but, on the other hand, may also reflect unobserved heterogeneity.

Further analyses that investigate the distribution of men and women across industries and hierarchical levels and its impact on wages have been conducted by Fitzenberger and Wunderlich (2002), Busch and Holst (2011, 2012) and Bechara (2012). The latter reveals that at the time of labor market entry, the gender wage gap can be almost fully explained by women selecting into lower-paying occupations and firms. Fitzenberger and Wunderlich (2002) assess gender wage differences across the skill distribution in Germany over a period of more than 20 years (1975-1995) controlling for cohort effects. In the observation period, they find a narrowing wage gap. However, earnings growth differs across skill levels with low- and medium-skilled women benefitting most while the reduction of pay differences is particularly small for high-skilled females as opposed to their equally-qualified male counterparts. Busch and Holst (2011) investigate the effect of horizontal and vertical segregation on gender wage differentials in management positions. Using GSOEP data from 2001-2008 and controlling for selection into managerial positions as well as differences in human capital endowments, they find support for a systematic wage lag in female-dominated as opposed to male-dominated jobs resulting in lower pay for women. A decomposition of the wage differential further reveals that 35 percent of the variation in wages cannot be explained by the regression model which they suggest might indicate discriminatory practices prevalent in the labor market. Further studies reveal that wage discrimination in female occupations is restricted to large employers (Busch and Holst, 2012), is significantly smaller in public as opposed to private companies (Melly, 2005) and turns out to be most prominent in firms without a works council (Jirjahn, 2011).

Contrary to the former studies that use large-scale publicly available datasets, Pfeifer and Sohr (2009) use firm-level data from one single German company covering a period of seven years (1999-2005). They find an unconditional gender pay gap of 15 percent for blue-collar and 26 percent for white-collar workers. This gap however decreases to 13 percent for both production and administration workers if individual characteristics reflecting human capital endowments and working hours are included in the estimation. The gender pay differential even further declines (3.5 percent for blue-collar and 8 percent for white collar workers) as soon as controls for hierarchical levels are included in the wage regressions. Examining the earnings profiles, the results indicate that the gender pay gap for white-collar workers decreases with tenure.[15]

3.2.1.3 Findings on Gender Employment Differences outside the German Labor Market

Quite a few challenges prevail when the reasons of gender differentials in access to employment should be assessed. These challenges particularly concern the availability of data with an adequate set of control variables. Therefore, the regression-based literature

[15] Pfeifer and Sohr (2009) interpret their findings as evidence for statistical discrimination (see also section 3.2.3.3). Inherently, employers pay women less than men since they have less accurate expectations about women's productivity. However, learning that women are as productive as men, employers adjust their wages which leads to a reduction of the gender pay gap over time.

is rather scarce. Indeed, research generally focuses on the effects of occupational segregation on wages rather than identifying the factors for occupational segregation and differential treatment in access to employment (e.g. Darity and Mason, 1998). Some exceptions, however, are available.

Investigating U.S. Census and survey data from 1940 to the late 1980s, Coleman and Pencavel (1993) show that women's labor market attachment differs across skill levels. In fact, high-skilled women have increased their working hours since World War II, but low-skilled women significantly reduced them as opportunity costs of taking up employment have risen or, put differently, reservation wages have increased. England (1982) uses NLS data from 1967 to show that among 30 to 44 year old women, the type of occupation does not have an impact on the effect of the time spent out of the labor force on wages. In other words, selecting into female- rather than male-dominated jobs does not seem to make a difference. Reviewing the U.S. literature, she also claims that segregation and child-rearing as the two main determinants of the gender pay gap are unrelated. More precisely, women do not trade in career interruptions and mother-friendly work environments for on-the-job training, higher earnings and better career prospects (which contrasts the findings presented in footnote 13) (England, 2005).

However, empirical research on the effect of gender-specific job choices seems to shed more light on differences in hiring outcomes. Eberharter (2012) assesses the impact of individual and family background characteristics on occupational choice and its relation to wages across different countries. Relying on longitudinal data from the U.S. (PSID), the U.K. (BHPS) and Germany (GSOEP) over a period of three decades (1980-2010), she demonstrates that even though the level of horizontal and vertical segregation has decreased from one generation to the next, occupational choice is still gender-specific and does not markedly differ across countries. The reason for that may be rooted in applicants' preferences as e.g. shown by Fernandez and Friedrich (2011). They use data from 5,315 telephone applications successfully directed to a call center over a 13-month period. At the application stage, the job candidates were asked about their preferences for typically male- (computer programmer) and female-dominated (receptionist) occupations. Not surprisingly, gender stereotyping already exist at the pre-hiring stage. Even though hiring probabilities were unknown to the candidates and (self-assessed) skills were held constant, female applicants gave the job as a receptionist a significantly higher rating than male applicants while men preferred the rather masculine occupation as a computer programmer.

Apart from supply-side evidence on why segregation in the labor market occurs, hiring differentials are also found to originate from demand-side factors. Various laboratory experiments provide clear evidence for gender-stereotyping in the evaluation of application forms where men are perceived as more suitable for tenured and high-level positions as well as in male-dominated domains (Fidell, 1970). Additional information on applicants' quality, though, eliminates or, at least, reduces this 'gender-job-bias' (Glick et al., 1988; Heilman et al., 1988).

Outside the laboratory gender discrimination in access to employment has been the subject of research in a number of countries and occupations. Goldin and Rouse (2000) use audition records and personnel rosters to study the effect of a procedural change in the hiring process of U.S. orchestras on the employment of female musicians. Observing 588 auditions with more than 7,000 individuals over a course of almost 40 years, they find that the change from open to blind auditions explains approximately one third of the increasing fraction of women among new hires while an increase of women in the applicant pool is responsible for another third. Overall, the introduction of blind auditions accounts for 25 percent of the increase in the share of women being employed which they suggest provides evidence for discrimination against female musicians.

Since natural experiments such as the one quoted are rare, researchers have started to carry out their own field experiments relying on matched-pair testing. Most of these studies investigate whether gender discrimination is influenced by the job type and may thus affect horizontal sex segregation. As one of the first, Levinson (1975) uses telephone inquiries in order to test for differential treatment in 'sex-inappropriate' jobs, that is, whenever the majority of people employed in a certain occupation is of the opposite sex. Overall, he finds evidence of what he denotes as "clear-cut" discrimination, i.e., cases where either of the candidates is rejected while the counterpart is either redirected or directly interviewed, in one third of the 246 inquiries. Yet, women in male-dominated jobs are discriminated somewhat less (28 percent) than men in female-dominated occupations (44 percent). One explanation he suggests is that employers fear being regarded as discriminatory against women. Apart from that, he concludes that the degree of sex-stereotyping measured as the proportion of opposite sex employees in a specific occupation affects the extent of differential treatment. Hence, not surprisingly, Nunes and Seligman (2000) testing in-person applications of male and female candidates in auto-shops located in San Francisco, find strong evidence for discrimination against the female applicant.

Apart from the findings from audit studies, researchers conducting correspondence tests have come to quite similar results of which a selection is summarized in table A-1 in the appendix. Reasons for contradictory results across countries may, on the one hand, have their root in differences in occupational gender distributions (Booth and Leigh, 2010). On the other hand, cross-country differences in labor market regulations (especially with respect to prevailing affirmative action policies) and gender roles in society may help to explain the heterogeneous results. For instance, in the Swedish labor market where gender differences have historically been smaller than in other countries, Carlsson (2010) does not find significantly lower callback rates for women in male-dominated jobs.

Looking more closely at discrimination towards women, Hitt and Zikmund (1985) reveal that the gender effect per se is not statistically significant. However, if applications of women signal a commitment to equal employment opportunity issues, hiring differences occur. A similar idea is pursued by Weichselbaumer (2004) who investigates discrimination of male applicants in female-dominated jobs and of female candidates in male-dominated jobs in the Austrian labor market. In particular, she studies how different sex stereotypes and personalities affect gender discrimination. Therefore, she distinguishes between résumés of women that convey feminine traits and appearance and those that convey rather masculine characteristics. Across the entire sample, discrimination towards men and women prevails in female-dominated and male-dominated jobs, respectively. Perhaps surprisingly, results do not change when personality is controlled for. Neither do 'masculine' women have an advantage in male-dominated jobs compared to women with rather 'feminine' characteristics (both perform significantly worse than the male candidate), nor do 'masculine' women have a disadvantage when applying for female-dominated occupations.

Apart from the importance of job types, correspondence tests are also implemented to study the role of (expected) maternity and parenthood on hiring probabilities. While Albert et al. (2011) fail to find relative discrimination against 37-year-old married women with children in the Spanish labor market, the results by Correll et al. (2007), using field data from the U.S., indicate that mothers suffer from significantly lower callback rates as opposed to childless women. Furthermore, evidence from France and the U.K. highlights that expected maternity particularly disadvantages women in getting access to high-skilled and career-oriented jobs (Firth, 1982; Duguet et al., 2005; Petit, 2007).

3.2.1.4 FINDINGS ON GENDER EMPLOYMENT DIFFERENCES IN THE GERMAN LABOR MARKET

To the best of the author's knowledge, empirical findings on direct gender discrimination using the audit and correspondence method do not exist in the German labor market. In fact, even regression-based studies that focus on hiring differences and the reasons for occupational gender segregation are rather scarce.

Fitzenberger et al. (2004) compare labor force participation and employment rates of men and women from West Germany over a period of 20 years (1976-1995). They use Microcensus data in order to compute employment and participation profiles by gender that account for time, age and birth cohort effects. Their findings indicate that employment and participation rates of men and women have narrowed over time. While men's labor market attachment has declined, women's participation rates have increased due to changes in labor demand and increasing opportunities of part-time employment. In particular, low- and medium-skilled women are responsible for this trend as their opportunity costs of not entering the labor force have increased. However, while age-employment profiles of males remained unaffected, those of females are still characterized by an M-shape due to the family phase. Employment patterns further indicate that full-time employment decreases while part-time employment strongly increases with age. This development is primarily influenced by female cohort effects suggesting that medium- and high-skilled women increasingly engage in part-time employment.

Given these general employment patterns, Kunze and Troske (2009) investigate gender differences in job mobility and job search behavior of displaced men and women contingent on the life-cycle. They use a two percent random sample drawn from the social security records covering almost three decades (1975-2001). The dataset only includes 20 to 60 year old workers who have been displaced due to establishment closures and contains information on employment spells and wages. Estimating different survival models and controlling for unobserved heterogeneity, the authors find that gender differences in displacement spells are primarily influenced by female workers in their prime age (between 20 and 35 years) who have significantly longer unemployment spells than their male counterparts. In fact, in the age cohort 56 to 60 years, women even have shorter spells of displacement than men. Thus, the results suggest that fertility decisions and (expected) maternity help to explain gender differences in labor market participation. Further estimates indicate that wage drops after displacements are slightly higher for women than for men (Crossley et al., 1994). Even though only prevailing in some age cohorts (20-25 and 46-50 years), these findings once again indicate that access opportunities to (new) employment may impact wages differently by gender.

3.2.1.5 Conclusion

Empirical research has demonstrated that while women have increasingly entered the labor market and have benefitted from narrowing human capital endowments, they are still paid lower wages due to, inter alia, the anticipated costs of maternity leave, decreasing returns to skills in low-skilled jobs, direct wage discrimination as well as occupational segregation. Labor market segregation, in turn, has been shown to result in both, women being overrepresented in lower-status and lower-paid jobs as well as women dominating in lower hierarchical positions within occupational categories.

Overall, research in the German labor market yields quite similar findings than studies from abroad: differences in individual characteristics and segregation across industries and hierarchical levels explain the major fraction of the pay gap. Besides, there is still a substantial share of unexplained differences that may be a result of wage discrimination. However, while human capital endowments have converged over time, labor market segregation still seems to be a major determinant of the gender pay gap, especially as female occupations are found to face a wage penalty compared to male-dominated jobs.

The question thus remains whether gender differences in access to certain jobs and occupations influence the wage effect and whether these differences originate from the labor-supply or -demand side. Here, regression-based evidence provides rather mixed results indicating that self-selection as well as discrimination by employers explain the variations in participation rates and occupational distributions. Direct evidence from previous correspondence and audit studies, however, supports that gender discrimination is present in 'sex-inappropriate' jobs for both male and female applicants. Hence, Riach and Rich (2002) conclude that prior findings are consistent with the hypothesis that gender roles in society have an impact on horizontal sex segregation as they evoke gender discrimination in certain occupations.

3.2.2 Different Labor Market Outcomes by Ethnic Background

Analogously to the literature review on the development and sources of gender differences in the labor market, the subsequent section provides an overview of some frequently cited papers investigating ethnic wage and employment inequalities. In order to account for country-specific peculiarities, empirical results from the German labor market are again presented separately.

3.2.2.1 FINDINGS ON ETHNIC WAGE DIFFERENCES OUTSIDE THE GERMAN LABOR MARKET

When analyzing relative black-white earnings in the U.S. over time, a lot of similarities to the development of the gender pay gap and to wage inequalities of immigrants in other industrialized countries can be observed.[16] During the 1950s to 1970s, the racial wage gap has narrowed with two reasons accounting for this development. On the one hand, blacks have benefitted from more resources in education which improved schooling quality relative to whites (Smith and Welch, 1989). And, on the other hand, legislative enforcements, particularly the Civil Rights Act, have contributed to labor market equality as blacks increasingly invested in human capital and had better access to certain occupations and industries (Card and Krueger, 1993). As a result, the racial skill gap has continuously decreased until the late 1990s (Altonji et al., 2012).

However, the narrowing of the wage gap due to skill convergences has slowed down and even reversed during the 1980s (see Juhn et al. (1991) for an extended discussion). Firstly, the wage structure started to change. The change particularly disadvantaged low-skilled workers among which blacks (and other ethnic minorities such as Hispanics) were overrepresented (e.g. O'Neill, 1990; Gottschalk, 1997). In response to this price reduction, labor force participation in the low-skilled sector fell as the wages offered deceeded reservation wages. The population of blacks who remained in the workforce was thus positively selected. Empirically, such selection needs to be properly accounted for and, indeed, has reduced the black-white wage convergence of males even further (e.g. Brown, 1984; Chandra, 2000; Juhn, 2003; Western and Pettit, 2005; Fearon and Wald, 2011; Hunt, 2012).[17]

Secondly, the extent of labor market discrimination was found to have increased during the late 1970s to 1980s. While in 1976 about 19 percent of the wage gap between black and white men could be attributed to different intercepts and lower return rates for blacks, this share increased to 26 percent in 1985 (Cancio et al., 1996). In line with that,

---

[16] Note that most of the research presented below investigates wage differentials between blacks and whites in the U.S. Yet, inferences from these findings on the prevalence and extent of discrimination against other ethnic groups and in other labor markets need to be drawn carefully. To illustrate this, previous research has used skin-shades to proxy different ethnic affiliations (e.g. Telles and Murguia, 1990; Darity et al., 1996; Goldsmith et al., 2007). Indeed, these studies have established a relationship between skin-shades and wage differences. The results suggest that a 'darker' skin color, ceteris paribus, leads to a larger wage gap. Thus, the reported black-white wage differentials may rather constitute the upper bound compared to other immigrant-native wage disparities.

[17] With respect to females, the situation is somewhat similar. Unlike whites, the population of black females in the labor market is positively selected. Consequently, wage gap estimates are likely to underestimate the actual extent of wage differentials (Anderson and Shapiro, 1996).

Altonji and Blank (1999) find that the fraction of the black-white wage gap explained by differences in return rates and the intercepts has increased when CPS data from 1979 and 1995 are compared. Their results indicate that earnings differences have increased from 16.5 to 21.1 percentage points. Even though both the amount attributable to endowments and parameters increased, the impact of the latter reflecting discrimination rose relative to the former. In other words, groups' (skill) endowments narrowed, but were more unequally rewarded.

Using longitudinal data from the NLS (1966-1981), Kilbourne et al. (1994) find that labor market experience, education and cognitive skill requirements as a proxy for hierarchical positions make up the largest proportion of the racial earnings gap for both men and women. In contrast, other independent variables such as marital status, the share of female employees and industrial segmentation contribute only marginally, if at all. Though not explicitly discussed by the authors, a rather substantial fraction of the pay gap still remains unexplained which may, inter alia, indicate the prevalence of labor market discrimination (and thus supports the findings presented above).

If, however, the main covariates such as schooling or labor market experience systematically differ as a consequence of e.g. racial group differences in family and school environments, the actual wage gap may be over- or underestimated and spurious evidence of discrimination may be provided. In order to control for these potential differences, an unbiased measure of skills and abilities is required. Fortunately, the NLS include information on the Armed Forces Qualification Test (AFQT), a measure of verbal and mathematical skills originally designed to determine an individual's qualifications for military service. Arguing that these test scores are racially unbiased and reflect differences in schooling quality and family background, O'Neill (1990) shows that controlling for AFQT scores, schooling and potential labor market experience reduces the white wage premium quite substantially. About three quarters of the remaining black-white earnings gap among 22-29 year old men can be explained by her regression model. In fact, adding actual labor market experience makes the wage differential almost fully disappear. Later, Neal and Johnson (1996) have somewhat reproduced these findings. They included AFQT scores as the only productivity-related measure revealing that pre-market skill differences explain the entire racial pay gap for females and a substantial fraction for males. Therefore, they conclude that policy actions should focus on the alignment of schooling quality rather than quantity when tackling racial differences in labor market outcomes (see also Maxwell, 1993). However, there is no consensus about the O'Neill and Neal and

Johnson results. Rodgers and Spriggs (1996) and Carneiro et al. (2005), for example, show that wage differences reemerge if alternative model specifications are considered. This discussion illustrates that the racial pay gap may already originate from pre-market differences even though they are likely not to be responsible for the entire disparity.[18]

Apart from pre-market factors, experience, seniority, training and job mobility are documented to affect racial wage differences. Though, it is again not clear whether this is due to an endowment or a return effect. D'Amico and Maxwell (1994) show that disparities in experience endowments rather than different return rates are the main force behind the black-white earnings disparities in early career years. Yet, following young high school graduates from the NLSY sample over 13 years (1979-1991), Bratsberg and Terrell (1998) refute these results and report that blacks are less rewarded for accumulated experience than whites.

Further evidence on wage differentials between natives and ethnic minorities can be traced back to differences in occupational and hierarchical distributions (e.g. Carrington and Troske, 1998; Huffman and Cohen, 2004; Aydemir and Skuterud, 2008; Pendakur and Woodcock, 2010). Barth et al. (2012) demonstrate with employer-employee linked data from Norway that differences in unemployment spells and career prospects explain 40 percent of the wage gap between natives and immigrants. In particular, immigrants fail to advance to higher-paying firms and thus experience flatter wage growth than their native counterparts. In the same vein, Eliasson (2013) reports that inequalities with regard to job mobility among the highly educated in the Swedish labor market account for a large fraction of the ethnic wage gap. These two examples indicate that, similar to gender wage differences, horizontal and vertical segregation need to be considered as additional factors influencing the ethnic and racial wage gap.

3.2.2.2 FINDINGS ON ETHNIC WAGE DIFFERENCES IN THE GERMAN LABOR MARKET

In order to put the existing evidence into perspective and to find similarities in the qualitative results, it may be worthwhile explicitly focusing on empirical findings on ethnic

[18] For a brief overview on the debate of AFQT scores, see also Darity and Mason (1998), Lang and Manove (2011) and Lang and Lehmann (2012). The impact of pre-market factors in explaining the wage gap is also found to differ depending on ethnic origin as e.g. shown by Black et al. (2006).

wage differences from the German labor market.[19] Velling (1995) analyzes a one percent sample of the 1989 employment register data including historical labor market information of 11,657 foreigners (from 14 different countries) and 105,204 Germans. He finds that differences in endowments make up the largest share (roughly 80 percent) of the overall wage gap which varies between 12.6 and 13.1 percent. The remainder can be attributed to discrimination where the magnitude is slightly higher (and thus endowment effects lower) if occupation dummies are excluded from the wage regressions. Using 14 waves of the GSOEP (1984-1997), Constant and Massey (2005) yield similar results. Despite assimilation in educational attainments, foreigners earn significantly less as they are overrepresented in lower status jobs and suffer from discrimination in the process of climbing up the job ladder (see also Riphahn, 2003). Yet, if occupational status is controlled for, average weekly earnings differentials decrease over time and completely disappear after 23 years.[20] Direct wage discrimination therefore only plays a minor role.

Whether the assimilation of wages differs between immigrant cohorts and skill groups, is, inter alia, investigated by Fertig and Schurer (2007). They analyze GSOEP data from 1984-2004 and show that earnings growth of ethnic Germans and persons who immigrated between 1988 and 2002 converges after 10 years. These results are robust to controls for unobserved heterogeneity across groups and sample attrition bias in the GSOEP.[21] Older immigrant cohorts (1955-1968 and 1974-1987), though, are found to suffer from flatter earnings profiles over their careers so that the wage gap widens rather than narrows over time. Detailed analyses by skill levels further reveal that differences in the earnings-experience profile are largest if high-skilled Germans and first generation immigrants are compared. Furthermore, with respect to industry differences, it is noticeable that the largest differences in the returns to experience occur in industries where the share of immigrants is lowest (Zibrowius, 2012).

Aldashev et al. (2007) use a more detailed distinction of people's migration history and compare the earnings prospects of native Germans, ethnic Germans, persons with

---

19 Note that the studies presented below compare the wages of employees in the German labor market. For empirical evidence on ethnic earnings differentials of the self-employed, see e.g. Constant and Shachmurove (2006), Constant et al. (2007), and Constant (2009).

20 Quite surprisingly and in contrast to prior empirical studies, Schmidt (1997) does not find significant monthly earnings differences between natives, ethnic German migrants, and foreign guest-workers if educational endowments, occupational status and industries are accounted for.

21 Constant and Massey (2003) evaluate the impact of selective out-migration on earnings assimilation using GSOEP data (1984-1997). They fail to find evidence for a selectivity bias driving the cross-sectional estimates of the immigrant-native wage gap during the observation period.

migration background and foreigners. Using GSOEP data over an 11 year period (1995-2005), they particularly look at the returns to educational achievements where achievements from abroad and from Germany are distinguished. In line with Fertig and Schurer (2007), they find that earnings of foreigners and people with migration background are significantly below those of natives regardless of gender and skill level (except for medium-skilled women). Moreover, these differences are found to widen with age and are highest among high-skilled employees. However, earnings histories of foreigners compared to people with migration background differ just as little as earnings of native and ethnic Germans (except for the high-skilled). With regard to differences in return rates, their results confirm the prevailing consensus that educational endowments received in Germany are rewarded significantly higher than those received abroad. This is particularly true for school and university degrees and is less pronounced in case of professional training.

Even though it does not matter empirically whether a somewhat narrow (people with foreign citizenship) or broad (people with migration background) definition of ethnic minorities is used, decomposing factors of differential treatment and analyzing wage assimilation processes by different ethnic affiliations produces quite heterogeneous results. For example, Lehmer and Ludsteck (2011) evaluate wage differences between native Germans and groups of immigrants (EU, East EU, Other East and Turkey) focusing on immigrants entering the German labor market between 1995 and 2000. As expected, decomposition analyses yield quite different results across groups (see also Velling, 1995). Netting out the effects due to differences in characteristics leaves an unexplained gap of more than 50 percent for most nationalities considered. However, if occupations are controlled for, the impact of characteristics increases. Nevertheless, the unexplained wage gap still accounts for 20 to 30 percent which, according to the authors, points toward direct wage discrimination and occupational segregation. Lastly, wage differentials are found to vary over the earnings distribution for citizens of some countries (including those from the EU) where the results of quantile regressions indicate sticky floor effects as discrimination seems to be larger in case of low-income earners (see also Panagiotis/Schluter, 2012). All these findings indicate that the factors influencing differential treatment and thus the magnitude of discrimination need to be separately addressed for each immigrant group (see also Lehmer and Ludsteck, 2012). Moreover, concerning intergenerational wage assimilations, a narrowing of the ethnic-native wage gap from first to second generation immigrants can be found only for some, but not all ethnic minorities (Algan et al., 2010).

3.2.2.3 Findings on Ethnic Employment Differences outside the German Labor Market

Few empirical papers have been published thus far to investigate whether ethnic differences in unemployment and labor force participation rates are accounted for by differences in observable and unobservable characteristics (Charles and Guryan, 2011). Bound and Freeman (1992) decompose the racial employment gap of young men contingent on educational levels (college, high school, school dropouts) and regions (Midwest, Northeast, South). They investigate CPS data from the mid-1970s to late 1980s and provide evidence that changes in industry and occupational composition, (de)unionization, decreasing minimum wages, relative educational improvements of whites and decreasing demand for blue-collar jobs have all contributed to a substantial drop in the employment of blacks.

The labor market situation of women, in contrast, does not seem to be characterized by diverging employment rates of blacks and whites (e.g. King, 1992; Anderson and Shapiro, 1996). Looking at census data twenty years prior (1940), during (1960) and after (1980) anti-discrimination legislation, Cunningham and Zalokar (1992) find occupational status convergence of black women leading to a narrowing black-white wage gap. For example, between 1940 and 1980 the share of women in private household jobs decreased dramatically (from 58.4 to 6.2 percent) while during the same period the share of professional and technical workers (from 4.6 to 16.1 percent) as well as clerical staff (from 1.3 to 29.0 percent) increased substantially and even exceeded the overall trend towards more skilled labor.

Apart from disparities in employment and occupational distributions, racial and ethnic differences in unemployment risks have widely been investigated. Fairlie and Sundstrom (1997), for example, use the Public Use Microdata Sample (PUMS) to study the changes of the racial unemployment gap in the U.S. for more than a century (1880-1990). They demonstrate that the unemployment rates did not differ until the late 1930s. After 1940, however, unemployment rates of blacks decreased less than those of whites and ended up at a ratio of two to one. This ratio remained almost constant until the 1990s and even increased thereafter. Still, part of the unemployment gap and its increase remain unexplained which the authors admit may be partly related to omitted variables such as changes in legislation, crime and family structures, but may also leave room for racial discrimination. Chiswick et al. (1997) investigate unemployment and employment patterns of U.S. immigrants with CPS data from 1979, 1983, 1986 and 1988. Unlike racial differences, both rates converge and gaps disappear after 3 and 10 years after arrival in the U.S., respectively. However, with respect to employment outcomes, differences across immigrant groups are observed with Asians doing best and Mexicans worst. Likewise, Arai and Vilhelmsson (2004) find higher unemployment risks for non-Europeans than for Europeans in the Swedish labor market even after controlling for the impact of worker characteristics, wage rates and unemployment risks across establishments. Both findings seem to suggest group differences in hiring discrimination. The latter explanation finds further support in Rooth (2002). He compares the employment outcomes of native Swedes with those of ethnic minority men who were adopted by Swedish families. All other things being equal, employment probabilities of these two groups differ by almost 10 percentage points. However, the differences vary by ethnic origin. Oaxaca-Blinder decomposition further reveals that more than two thirds of the variation in employment cannot be explained by schooling, age, marital status and the local unemployment rate. Acknowledging the peculiarities of adoptees' ethnic backgrounds leads the author to suggest that the unexplained gap originates from skin-color discrimination. Not surprisingly, these results are also in line with the findings on skin shades and wages presented in section 3.2.2.1.

Direct evidence on hiring discrimination has most convincingly been produced by field experiments such as correspondence and audit studies. Among these, without any doubt, racial and ethnic differences have attracted most researchers' attention (see table A-2 in the appendix for a selective list of correspondence studies and their results). Jowell and Prescott-Clarke (1970) were one of the first researchers who sent out fictitious résumés and reported the callbacks for British, Australian, West Indian, Pakistani and Cypriot applicants in the British labor market. All in all, they replied to 128 job offers in various occupations, e.g. sales and marketing, accountancy and office management, electrical engineering and secretarial jobs. As a result, they find that non-white (the latter three ethnic groups) as opposed to white (native Brits and Australians) candidates receive significantly fewer positive responses. Furthermore, altering the level of qualification shows that immigrants realize only minor returns to schooling and thus benefit less from higher quality résumés. A follow up study by McIntosh and Smith (1974) that doubled the number of observations supports the aforementioned findings. They trained and matched British, Greek and West Indian job candidates who then applied by phone. Callback rates between the first two groups do not turn out to be statistically different from each other. However, comparing firms' responses to the British and West Indian candidate yields a significantly lower callback rate for the latter.

The study by Riach and Rich (1991) was one of the first that used matched-pair testing outside the U.K. They created fictitious job pairs of male and female applicants and applied as sales representatives, clerks and secretaries in Australia showing that minority groups, i.e., Vietnamese and Greek immigrants, face discrimination in the recruitment process. Shortly thereafter, correspondence studies were also carried out in the U.S. (Bendick et al., 1991; Bendick et al., 1994; Kenney and Wissoker, 1994) and all across Europe. Bovenkerk et al. (1996), for instance, find differential treatment of male and female Moroccan and Surinamese immigrants in the Netherlands. Similar findings are reported by Angel de Prada et al. (1996), Arrijn et al. (1998) and Allasino et al. (2004) for male Moroccans in Spain, Belgium and Italy, respectively. A common trait of all these studies is that discrimination is most prominent in and often restricted to the first stage of the hiring process. In France, for example, Cediey and Foroni (2008) point out that 85 percent of all instances of discrimination against North and Sub-Saharan Africans are based on the evaluation of written applications, i.e., during the first step of the hiring process.

What can be considered as the most prominent work in this field is the paper by Bertrand and Mullainathan (2004). By sending out almost 5,000 applications in response to 1,323 job offers in the Chicago and Boston metropolitan areas, they show that African-Americans have a 50 percent lower callback rate compared to white Americans. Moreover, the results demonstrate that these differences neither vary across industries and occupations, nor are they contingent on the socio-economic characteristics of the applicants' neighborhood, on whether the firm is an equal opportunity employer or not and on whether the employer operates in the public or private sector. Bertrand and Mullainathan also altered the quality of résumés and sent out one pair of high and low quality applications to each vacancy. By doing so, they show that white applicants realize higher returns (in terms of callbacks) for high quality résumés than black candidates. Pager (2003), Pager and Quillian (2005) and Pager et al. (2009) support these results and, perhaps surprisingly, show that blacks statistically have the same callback rates than whites with a criminal record.

Regarding intergenerational differences, Carlsson and Rooth (2007) and Carlsson (2010) find differential treatment disadvantaging Middle-Eastern applicants in Sweden which according to the latter persists for first and second generation immigrants. Both studies also indicate that male recruiters discriminate significantly more. Remarkably, Oreopoulos

(2011) shows that discrimination exists in case of both, immigrants as well as native Canadians that have an ethnic-sounding name. Similarly, McGinnity and Lunn (2011) highlight that discrimination is not necessarily restricted to ethnic groups with other skin-colors and/or from low wage countries. They show that differential treatment in the Irish labor market is consistent for minority groups originating from Africa, Asia and Western Europe (Germany). Finally, research interacting ethnic origin with gender indicates that the effects may differ for men and women (e.g. Arai et al., 2011; Andriessen et al., 2012; Derous and Ryan, 2012). Arai et al. (2011), for example, show that high-quality résumés benefit minority women more than men and make discrimination disappear.[22]

Applying for small business transfers, i.e., taking over an existing business due to e.g. retirement of the previous owner, Ahmed et al. (2009) use the correspondence method to demonstrate that hiring discrimination not only exists in case of dependent employment, but may also affect the chances of becoming self-employed. Furthermore, Edin and Lagerström (2004), Eriksson and Lagerström (2012) and Blommaert et al. (2013) show that equally qualified job seekers from ethnic minorities are not only discriminated when actively applying for a job, but are also less likely to be contacted via an online hiring platform.

3.2.2.4 FINDINGS ON ETHNIC EMPLOYMENT DIFFERENCES IN THE GERMAN LABOR MARKET

The subsequently quoted studies represent a selection of empirical research conducted in Germany analyzing ethnic differences in unemployment risk and duration as well as employment participation rates and occupational distributions. Previous research has primarily relied on publicly available data with only a few exceptions having conducted field experiments. Kogan (2004), for example, investigates the transition into employment and unemployment using GSOEP data over a six year period (1995-2000). Her results indicate that native-immigrant differences are influenced by both human capital differences and segmentation across industries and occupational positions. In particular, first generation immigrants are channeled into unskilled labor and sectors where labor demand highly fluctuates which results in lower employment rates compared to native Germans (see also Constant, 1998). Second generation and EU immigrants, in contrast, do

[22] Additional reasons for minority-majority group differences in hiring are found to be based on systematic differences in application processing (e.g. Arvey et al., 1975), in recruiters' behavior (e.g. Giuliano and Levine, 2009; Giuliano and Ransom, 2011) and in applicants' job search methods (e.g. Holzer, 1987; Segendorf and Rooth, 2006).

not seem to be disadvantaged in finding new employment and bear the same risk of becoming unemployed as native Germans if tenure and job characteristics are accounted for. Unemployment duration also contributes substantially to differences between native Germans and immigrants' career paths which most obviously differ between natives and Turkish immigrants (Kogan, 2007). Kalter and Granato (2002) and Uhlendorff and Zimmermann (2006) support the finding that immigrant Turks in particular have significantly longer unemployment spells and are less likely to enter new employment. Most noticeably, their results extend to second generation Turks while, in line with Kogan (2004), guest-workers from other nationalities and their descendants are, ceteris paribus, hardly or not disadvantaged at all.

Other researchers have used employment probabilities as the outcome variable to measure ethnic differences in the German labor market. The main findings, however, remain the same. Despite controlling for socio-economic characteristics, employment gaps remain quite substantial. Algan et al. (2010), for example, find these gaps to vary across ethnic groups where both first and second generation Turks suffer most and have a 15.2 and 18.6 percent lower chance of being employed compared to native Germans. In other words, Turkish descendants are unable to realize superior employment outcomes than their parents. Further research by Kalter (2008), Heath et al. (2008) and Luthra (2013) report similar results and argues that some immigrant groups perform better over time and assimilate more quickly than others.

The importance of where educational endowments are attained is investigated by Brück-Klingberg et al. (2011). In particular, they study how different skill levels affect the hiring probability contingent on ethnic origin. Using survival estimates, they show that the return rates of education attained abroad and in Germany differ significantly. As a result, transition from unemployment to employment takes longer for both foreigners and ethnic minorities with German citizenship as opposed to native Germans.

Apart from differences in employment probabilities and distributions across sectors, immigrants are also found to be less likely to climb up the career ladder. Using GSOEP data (1984-1997), Constant and Massey (2005) find systematic differences in the allocation of occupational positions with workers of a migration background being less able to translate their human capital into higher job prestige. Similar results are produced by Luthra (2013) who analyzes employment outcomes and occupational attainments for different immigrant groups. Using Microcensus data from 2005, she shows that second generation immigrants of both sexes perform differently across immigrant groups but worse compared to native Germans.

The empirical findings from the German labor market presented thus far lack a direct measure of discrimination. Even though unemployment duration and employment gaps cannot completely be explained by human capital endowments and differences in the distributions across sectors, the unexplained fraction of the regression models may not necessarily reflect discriminatory treatment, but may also capture the effects from omitted variables such as family background information, language skills and social ties. Two studies try to circumvent these problems and assess the prevalence and extent of discrimination in access to employment by controlled field experiments. The results of both indicate that hiring differentials based on applicants' ethnic background may well be affected by the demand side and constitute discrimination on behalf of employers. Goldberg et al. (1996) conduct an audit and correspondence test where matched pairs of first generation Turkish immigrants and native Germans apply for semi- and higher-qualified jobs, respectively. In the audit study, the candidates made telephone inquiries to 333 job offers. In the end, members of the minority group were invited in 46 percent of all applications while the majority candidates received a callback in 53 percent of the cases yielding a 7 percentage points difference. Similarly, sending out more than 2,800 written applications in Berlin and the Rhine-Ruhr region, the authors find a 1 percentage point lower callback probability for the immigrant group. Unfortunately, no information on the statistical significance of these results is provided. Instead, the authors use the net discrimination rate which in both instances indicates unequal treatment at statistically conventional levels. A closer look also reveals that with regard to the correspondence study, discrimination of the minority candidates is restricted to commercial jobs only. Thus, the evidence is rather weak. More convincingly, Kaas and Manger (2012) find discrimination against equally qualified second generation Turks who apply for business internships. Here, the minority candidate is 5 percentage points less likely to receive a callback from employers. However, callback rate differences decline and become insignificant if the minority applicant attaches an additional reference letter providing favorable information on e.g. his qualifications, work effort and motivation.

3.2.2.5 Conclusion

Overall, differences in human capital endowments are shown to explain the largest fraction of the prevailing ethnic and racial wage gap inside and outside the German labor market. However, both average endowments and the size of earnings differentials vary quite substantially across immigrant groups. Consequently, the wage gap of some immigrant groups has narrowed over time while in case of others it has remained constant or has even increased. Similarly, while the unexplained fraction of the earnings estimates seems to have decreased after World War II, a substantial share still goes back to direct wage discrimination.

Another factor influencing ethnic wage disparities can be found in horizontal and vertical segregation with blacks and immigrants being channeled into lower-paying sectors and positions. Again, these phenomena can be traced back to discrimination in access to certain jobs. The findings quoted above point at substantial differences with respect to labor market participation, employment, unemployment and occupational distributions. In particular, the matched-pair studies presented provide direct evidence of discrimination in hiring towards certain minority groups, though they are (mostly) unable to identify its sources. With respect to discriminatory practices against blacks, Riach and Rich (2002: 503) conclude that prior field experiments "are more consistent with the majority white populations having a general 'distaste' (Becker, 1971), or 'social custom' (Akerlof, 1980), which motivates employers to discriminate against non-white applicants." However, it is yet not clear whether these conclusions hold true for immigrant groups in different countries and labor market segments.

3.2.3 Empirical Evidence on Different Sources of Discrimination

Charles and Guryan (2011) and Neumark (2012) argue that it is a fundamental challenge to disentangle the effects from taste-based and statistical discrimination. Firstly, because both approaches predict the same labor market outcome, i.e., discrimination towards a certain demographic group and, secondly, because findings supporting one approach can often be explained by some version of the other. In the following section, selected studies are presented that provide empirical evidence for either taste-based or statistical discrimination. However, not surprisingly, many of these studies find support for both theories.

3.2.3.1 Mixed Evidence

Gneezy et al. (2012) analyze a series of field experiments on age, gender, race, sexual orientation and disability discrimination and conclude that characteristics given by birth such as race or gender underlie statistical discrimination whereas other characteristics that may be subject to change while a person grows up such as sexual orientation are associated with taste-based discrimination. However, their results are based on studies outside the labor market. In fact, they conduct an audit study in the product market where ten white and black testers bargain for a car purchase at five different dealers in the Chicago area. In order to reduce unobserved heterogeneity, the testers are instructed to stick to a uniform pre-determined bargaining strategy. While no racial differences with respect to initial and final offers for low-end cars can be observed, interestingly, blacks on average receive a 1.5 percent ($630) higher initial and a 3 percent ($1,010) higher final offer for high-end cars. If car dealers had distastes for racial minorities, they would offer higher prices to minority buyers of both low- and high-end cars. As the price differences only exist in the high-end market, the authors expect statistical discrimination to be present. Unfortunately, however, they do not provide further empirical evidence on e.g. different search costs across groups depending on the cars' quality levels. Thus, their interpretation remains rather suggestive and leaves room for alternative explanations.[23]

Sometimes the empirical evidence neither convincingly supports taste nor statistical discrimination as shown in the study by Bertrand and Mullainathan (2004). On the one hand, customer discrimination is very unlikely to account for the racial hiring gap as the extent of discrimination does not vary conditional on whether or not the jobs require high communication skills and customer contact. On the other hand, statistical discrimination would suggest that the provision of additional productivity related information would decrease or perhaps even eliminate differential treatment. However, the opposite holds: callback rate differences are largest whenever high-quality applications including supplementary credentials are dispatched. As an alternative explanation, the authors argue that racial differences occur because recruiters start with sifting the pool of applicants and stop reading the applications if they are confronted with a distinctively black name. Ironically, they do not mention that this is what would be expected by either taste or statistical discrimination, i.e., group membership serves as a pre-selection device due to employers' distastes or group-based productivity beliefs.

In contrast to Bertrand and Mullainathan, Carlsson and Rooth (2008) find evidence for both economic explanations on why (ethnic) minorities are discriminated. In particular, they relate 23 percent of the hiring gap to the minority applicants' foreign qualifications

---

[23] Scott Morton et al. (2003) provide more convincing evidence for the prevalence of statistical discrimination in the market of new car purchases. They show that while minority customers pay a 2% price premium offline, the difference in buying prices disappears if online purchases are considered. They explain their findings by reduced information costs through on the Internet.

which they interpret as evidence for statistical discrimination and the remaining 77 percent to group membership per se. Decomposing the remaining difference indicates a mixture of both, on the one hand, employer and coworker discrimination as male recruiters and firms with a high share of male workers discriminate somewhat more and, on the other hand, statistical discrimination as recruiters presumably (need to) rely on sifting due to time constraints which results in a predominant rejection of minority applicants. Either way, all papers quoted so far outline the ambiguities that evolve if the different sources of discrimination should undoubtedly be identified.

3.2.3.2 EVIDENCE SUPPORTING TASTE-BASED DISCRIMINATION

Taste-based discrimination has been found to negatively affect the labor market outcomes of both women and ethnic minorities. Analyzing job offers from a Chinese internet job board, Kuhn and Shen (2013) show that preference related job targeting, i.e., discrimination against either men or women in opposite-sex stereotyped jobs, significantly decreases with the jobs' respective skill requirements. As with higher job requirements, search costs, foregone income for not filling the position and potential losses associated with adverse selection increase, their findings are in line with Becker's taste approach (Becker, 1971). In the same vein, Baert et al. (2013) conduct a correspondence test to uncover ethnic hiring discrimination in Belgium's youth labor market addressing occupations that differ with respect to the demand for labor. Indeed, the results reveal that employers respond to scarcity. While callbacks do not differ for vacancies that are difficult to fill, the minority candidates are clearly discriminated in occupations where demand for labor is rather low.[24]

The question, to what extent taste discrimination against ethnic minorities can be

[24] Somewhat related to taste-based discrimination is monopsonistic discrimination which is caused by group differences in labor-supply elasticities. The effects originating from these differences are illustrated by Hirsch et al. (2009). They exploit regional variations in demand-side competition for labor to assess the gender pay gap. Firms in metropolitan areas that face harsh competition for talents in the labor market are found to discriminate consistently less (over a 30 year period) than their counterparts from rural areas. The authors argue that unlike in Becker's model, employers in rural areas do not incur costs by discriminating the female minority because, otherwise, these employers would be driven out of the market in the long run which is not observed in the data. In contrast, women living in "hot-spots" simply have more outside options and therefore higher wage elasticities than in regions where alternatives are limited. As a consequence, employers' monopsy power and thus their ability to discriminate is somewhat constrained in big cities whereas in rural areas the opposite applies. Similar results are also published by Hirsch et al. (2010) and Ransom and Oaxaca (2010) who analyze differences in employment and quit rates conditional on gender-specific wage elasticities. Furthermore, Hirsch and Jahn (2012) demonstrate that, for the same reason, ethnic minorities are willing to accept lower wage offers than their native counterparts.

explained by societal attitudes towards these minorities has also been addressed in the recent literature (e.g. Charles and Guryan, 2008). Some of this research has linked the results from matched-pair studies with information on public opinions. Carlsson and Rooth (2011), for example, use survey data on attitudes towards ethnic minorities and the results of a previous correspondence test in the Swedish labor market. They assume that employers located in a certain region adapt the population's opinion on immigrants in that area. In fact, their findings reveal that discrimination is more likely in areas where the average employer has a more negative attitude against immigrants. However, this effect is only statistically significant if the sample is restricted to low-skilled occupations. Similarly, Rooth (2010) asks recruiters primarily involved in a fictitious field experiment to participate in an implicit association test that measures automatic attitudes and stereotypes towards ethnic minorities (for more details about the implicit association test, see section 4.1.2.5). The results show that implicit associations towards Arab-Muslim candidates are negatively correlated with callback rates and affect the outcome of the recruitment process to a statistically significant extent (for similar results from the Australian labor market, see Booth et al. (2012)).[25]

Further evidence of discrimination in line with Becker is provided by Szymanski (2000) who exploits data from professional soccer. He shows that some clubs are willing to accept poorer performance on the pitch than others by signing a below-average share of black players. Undoubtedly, these findings support preference-based discrimination. Moreover, as any (negative) effects on attendance as a potential signal for customer discrimination can be excluded, differential treatment likely goes back to either club owners' or other teammates' prejudices against black players, i.e., denotes employer or coworker discrimination.

Empirical studies that explicitly investigate whether taste-based discrimination originates from employers, coworkers or customers are mainly restricted to the latter (e.g. Holzer and Ihlanfeldt, 1998). Audit study results from restaurant hiring by Neumark (1996), for example, indicate that discrimination against women might be based on customers' preferences. While callback rates to male and female applicants do not differ in low- and

[25] Temporary events provoking increased media coverage and public perceptions, in contrast, are not found to affect the extent of discriminatory behavior. Neither do Åslund and Rooth (2005) find higher employment differentials of ethnic minorities after 9-11, nor do Carlsson and Rooth (2012) find lower hiring gaps after the use of correspondence testing was widely discussed in the media (see Pope et al. (2011) for opposite results in the sports environment).

medium-priced establishments, they do in high-priced restaurants where both male waitpersons and male customers dominate. Although these customers are not expected to have a general distaste towards women, hiring male staff signals tradition and prestige on behalf of the restaurants and thus may be thought to emphasize its superior positioning. Some latest field results from the Netherlands point into the same direction and uncover customer discrimination as a potential source of why people with a foreign sounding name have lower chances of being recruited compared to their native counterparts. In particular, majority-minority callback differences are twice as high in jobs that require (high) customer contact (8 percentage points) than in those without (4 percentage points) (Andriessen et al., 2012).[26]

While previous research reports some convincing evidence for customer discrimination, researchers have thus far struggled to unveil and disentangle the effects that originate from employers' and coworkers' preferences. One exception includes the studies by Haile (2009, 2012, 2013) who shows that disabled, female and minority coworkers decrease employees' well-being which, in turn, might induce employers to place these group at a disadvantage in the recruitment process.

3.2.3.3 EVIDENCE SUPPORTING STATISTICAL DISCRIMINATION

In addition to taste-based discrimination, many authors have related their findings on gender and ethnic labor market disparities to statistical discrimination. Gneezy et al. (2012) conduct experiments that test people's willingness to help others in everyday life situations. For these experiments, age-, gender- and race-matched testers were confronted with two distinct tasks. First, they should drop either a pen or a pair of keys and report whether they were picked up and returned by someone else. And, second, they should ask for a dollar for the parking meter or directions to a well-known location somewhere around. Overall, young black men did significantly worse in both tasks. The performance of older minority candidates, however, did not differ compared to the control group. Relating these findings to criminal records in Chicago during that time shows why: crime rates among young black men were by far the highest. Thus, the modest willingness to help young black men stems from people's fear of being robbed. People use group

[26] Another strand of research again uses sports data to show that customers' tastes foster racial (ethnic) employment and wage disparities (e.g. Kahn and Sherer, 1988; Kalter, 1999). For an overview of these studies, see also Kahn (1991). More recently, though, Kahn (2009) reports that racial hiring, wage and retention differences in U.S. basketball have been eliminated due to a decline in customer discrimination.

membership to draw inferences on the probability of being subject to robbery and therefore rationally prefer to help the white rather than the black testers. Theoretically, their behavior goes along with statistical discrimination. In the same vein, Knowles et al. (2001) provide interesting evidence that police officers search cars of black drivers more often for carrying drugs not because of racial distastes, but because they try to maximize their ratio of successful searches. They develop a model that relaxes assumptions according to which racial prejudices impact on policemen's decisions. In particular, they allow blacks to respond to increased searches by reducing illegal activities. In fact, this is exactly what the data suggest. Even though blacks have a higher probability of their vehicles being subject to search (as a result of inferences made by the police officers), guilt probabilities do not differ between blacks and whites.

In the labor market, regression-based studies by Neumark (1999) and Pinkston (2003) find that a large portion of females' wage setbacks can be explained by men's productivity signals having a stronger effect on starting wages because they are perceived as more reliable by employers. In line with what Pinkston denotes as screening discrimination, employer learning through tenure then has a greater impact on women's than on men's wage profiles. In other words, as employers' beliefs on women's future productivity become more accurate, gender wage differences decline. Further evidence for employer learning reducing labor market inequalities is also provided when the black-white wage gap is analyzed (e.g. Pinkston, 2006; Kim, 2012).

However, not only repeated interactions, but also the provision of credible signals may lead to decreasing labor market differences as denoted by Siniver (2011). He exploits a natural experiment to investigate the reasons for which immigrant physicians in Israel are discriminated on the basis of wages. In particular, physicians entering Israel prior and past the introduction of an obligatory licensing examination in 1989 are observed. The study provides two important insights. First, compared to physicians who immigrated prior to the obligatory licensing, the institutional regulation has affected the remuneration of post-licensing immigrants positively. And, second, the post 1989 immigrant-native wage gap has disappeared after 5.5 years while that of earlier immigrants remained. Both, the discontinuity in 1989 and the wage convergence of the treated group, i.e., those physicians that were required to take a test on their qualifications, point at statistical discrimination since the official approval of immigrant physicians' licenses has decreased employers' uncertainty about physicians' productivity and have thus led to a removal of labor market differences over time. In line with these findings, Kaas and Manger (2012) provide field evidence demonstrating that ethnic hiring differentials in the German labor market are motivated by statistical rather than taste-based discrimination. In particular, they show that the inclusion of additional productivity information leads to a convergence of hiring probabilities of native and immigrant applicants while in the absence of such credentials, the latter are significantly disadvantaged in terms of callback rates. Again, these findings support the idea that employers are inherently less able to correctly predict minorities' future productivity and therefore use the (usually lower) group average as a proxy. Due to the provision of credible signals, these group proxies become relatively unimportant so that especially minority applicants are evaluated on the basis of observable characteristics conveyed by their applications.

Finally, the importance of additional information available to the employer is also supported by findings from the laboratory. Heilman (1984) asks 77 university students to evaluate the résumés of fictitious applicants and judge on a nine point scale whether these candidates should be interviewed for a job or not. Moreover, the subjects rated the applicants' expected success in the job. The application forms were matched and only varied with respect to applicants' gender and whether a reference letter by a professor was attached or not. While in some cases this reference letter included information of either high or low job relevance, in the control group such credential was omitted. Not surprisingly in terms of statistical discrimination, the findings indicate that job suitability and potential success do not differ across gender if highly job relevant information is provided. Otherwise, though, men fare significantly better than their female counterparts. In a larger scaled study with 241 college students, Heilman et al. (1988) later reproduce the aforementioned results and show that additional information that proves women to be of high ability makes gender differences in subjects' evaluations disappear while a significant gap persists in the absence of such information.

4 THEORETICAL BACKGROUND, CONCEPTUAL MODEL AND HYPOTHESES

The following section develops the theoretical framework that helps explaining the labor market differences across gender and ethnic groups as presented in chapter 2 and 3. A special focus is laid upon the distinction between different types of discrimination as the empirical part explicitly tries to disentangle the effects from taste-based and statistical discrimination. Accordingly, a conceptual model is presented that formally describes how different preferences and information asymmetries affect the hiring outcome. Finally, based on the theoretical considerations and previous empirical findings, the hypotheses to be tested with the data from the field experiments are derived.

4.1 THEORETICAL BACKGROUND

At the beginning of the theory section, the employee-employer interaction particularly during the hiring phase is considered from a principal-agent perspective where the basic assumptions of New Institutional Economics hold. Afterwards, economic theories that explain differences in (pre-) labor market outcomes of individuals and demographic groups are elaborated. First, human capital theory and the dual labor market hypothesis are referred to in order to separate any effects on labor market outcomes that stem from differences in workers' and workplace characteristics from the effects that are based on discriminatory treatment. Second, the two seminal economic theories of labor market discrimination, i.e., taste-based and statistical discrimination, are presented in more detail. Finally, non-economic theories that may be regarded as a cause to prejudices and stereotypes are discussed.

4.1.1 RECRUITMENT AS DECISION UNDER UNCERTAINTY

Principal-agent theory provides a suitable framework that helps explaining agents' behavior when confronted with decisions under uncertainty such as hiring (e.g. Ross, 1973; Jensen and Meckling, 1976; Fama, 1980; Grossman and Hart, 1983). Based on the fundamental assumption that information in markets and, as a consequence, contracts signed in these markets are incomplete, the agent (in the context of this thesis: the applicant) has superior information on her quality which in turn is ex ante unknown to the principal (here: the employer/ recruiter). The latter is thus confronted with a decision under uncertainty that Akerlof (1970) in his seminal paper illustrates, inter alia, by referring to the automobile market. Assuming that such a market entails good and bad cars, but quality is unobservable to buyers, the average price sellers demand would overpay bad and underpay good quality. Since the costs from selling overpriced low-quality cars, so-called “lemons”, are borne by the market, every individual seller has an incentive to offer poor quality. The buyer, on the other hand, constantly faces the risk of selecting “lemons”. As these “lemons” are worth less than the average market price, the buyer would only be willing to pay a price below the market average. Anticipating this, sellers in turn lower the offered quality. In the end, under asymmetric information, average quality and market size shrink until the market eventually breaks down. To avoid a market breakdown, economic institutions such as guarantees or brands may serve as a signal to the buyer that she bargains for high quality cars.

In the labor market or, more precisely, in the hiring context, an employer (principal) faces the problem of adverse selection whenever he is unable to distinguish between high- and low-quality (i.e., more or less productive) applicants (agents). To be able to identify and sort out “lemons”, he may rely on certifications such as high school diplomas or university degrees. Likewise, an employer may prefer one demographic group over another not because he is prejudiced, but because group membership serves as a quality device for applicants that are otherwise hard to distinguish (Akerlof uses this example to show why minorities fare worse in entering employment). Furthermore, he can implement screening mechanisms in the recruitment process. Such mechanisms comprise e.g. résumé evaluations, (telephone and face-to-face) interviews, assessment centers, or probation periods and should help the employer to reduce uncertainty about applicants’ productivity.

Yet, as proposed by Spence (1973), even from an agent’s perspective, it might be worthwhile to offer ability signals that ex ante lower asymmetric information and improve employers’ productivity beliefs. The basic rationale is that the production of signals creates costs where costs are negatively correlated with productivity. Agents select the amount of signals that maximize expected profits, i.e., the differences between offered wages and signaling costs. In order to successfully distinguish high- from low-quality agents, signaling costs must differ across groups in such a way that the production of ability signals pays off for high-, but is unprofitable for low-quality agents. Moreover, a sufficient number of distinguishable signals is needed such as, for example, years of schooling or different university degrees. Signaling theory then shows that the market arrives at different equilibria in which the value of signals is reproduced, i.e., confirms employers’ beliefs.

However, indices, that Spence refers to as demographic characteristics determined by birth (e.g. race or gender), may affect productivity beliefs as well. Whenever demographic groups differ with respect to their opportunity structures, that is, have different signaling costs, and thus invest differently in the production of signals, two distinct equilibria arise. The lower level equilibrium of one group as opposed to the other is self-perpetuating. Spence denotes this situation as a "lower level equilibrium trap" (Spence 1973: 374). In essence, this trap forms the ground for group differences in the returns to e.g. education and statistical discrimination as will be discussed later in this chapter.

Besides screening and signaling, the principal might induce self-selection by offering a distinct set of contracts that induces the agent to reveal her true quality. Wage contracts, for instance, may vary with respect to the ratio of fixed and variable pay. A higher fraction of the latter may attract high ability workers assuming that workers have the same risk preferences and act as utility maximizers. Conversely, workers of inferior productivity would select themselves into contracts where pay is predominantly fixed. Again, self-selection requires a sufficient set of contracts agents can choose from.

To briefly conclude, information asymmetries between principals and agents carry the risk of adverse selection (be it in the employment context, on the product market or anywhere else) which may eventually cause a market breakdown. To overcome these market inefficiencies, on the one hand, agents may invest in the production of signals that credibly shows them to be of high quality. On the other hand, principals may engage in screening or induce self-selection on behalf of the agents. In any case, agency costs arise that lead to a deadweight loss if compared to a market of symmetric information. Since a theoretical background highlighting the core problem associated with recruitment decisions has now been developed, next, theories that explain differences in labor market outcomes (including hiring) are presented.

4.1.2 Theories Explaining Labor Market Inequalities

As the stylized facts and previous empirical research indicate, demographic groups may differ with respect to all kinds of (pre-) labor market outcomes including scholastic achievements, unemployment and employment ratios, distributions across sectors and hierarchical levels, wages, promotion probabilities and quit rates, just to name a few. The following section presents some basic economic theories that explain these differences. However, these theories may be closely linked. As a result, labor market outcomes may reinforce each other leading to difficulties when trying to disentangle causes and consequences. Horizontal and vertical segregation, for example, may push minority groups into low-paying jobs, thus fostering already existing wage disparities. In addition, group differences may already evolve based on endowments, preferences and expectations brought to the labor market. That is why especially more recent empirical works as shown in section 3.2 account for unobserved heterogeneity and include proxies for factors influenced by pre-market developments in their regression models.[27]

4.1.2.1 Pre-Market Inequalities

Previous research on group differences in pre-school and school attainments relies on both economic and non-economic theories (Altonji and Blank, 1999). The former is mainly about beliefs and expectations on how the labor market rewards scholastic achievements. According to anticipated payoffs, parents invest differently in the schooling of their children shaping their endowments and preferences. This, for example, may result in ethnic minorities leaving school earlier than their classmates or girls focusing on other subjects than boys. Also, not surprisingly, these investments are often a response to expected labor market discrimination that lowers the playing field of those who suffer from discriminatory treatment. Furthermore, groups may differ in what Altonji and Blank (1999) refer to as comparative advantages. These differences are mainly an issue of gender. For instance, women are expected to work more efficiently in household production whereas men are assumed to perform better in physically-demanding jobs, both because they historically have more experience in either field. In addition, parents' investments often reinforce the gender-specific experiences contributing to gender segregation prior to employment.[28]

However, the behavior of girls putting emphasis on other subjects than boys and parents encouraging them to do so, cannot necessarily be explained by an economic rationale. Family, neighborhood, fellow pupils or society in general may have established role models and legal constraints that shape children's preferences, thus leading to group differences in early human capital accumulation (recall the results by Fortin (2005) and Backes-Gellner et al. (2013)). In an environment where women are primarily in charge for

---

27 Note that in the literature either the word pre- or non-market inequality is used (see e.g. Arrow (1971) for the latter). Both can be considered as synonymous.

28 See, for example, Mincer and Polachek (1974) for the factors (such as the number of children) influencing (gender-specific) family spending in human capital and Polachek (1981) for how early human capital acquisition affects occupational self-selection.

child-bearing and -rearing, they might not even develop a desire to acquire human capital and participate in the labor market. Moreover, discrimination embedded in the structure of the educational system and/or enforced by (pre-school) teachers may provoke pre-market inequalities.

No matter whether differences in intergroup educational outcomes are economically or non-economically motivated, in line with Spence (1973), they carry the risk of reinforcement. Whenever at least some members of a demographic group, for example blacks, are denied or restricted access to schooling, are channeled into lower quality schools or grow up in an environment that does not encourage them to acquire skills, employers start using membership, e.g. race, to infer the individuals' productivity. As a consequence, these employers rationally prefer whites over blacks in the recruitment process or contract blacks at lower wages than their white counterparts. Anticipating employers' behavior, blacks in turn underinvest in schooling and therefore confirm employers' beliefs. Hence, past and current labor market experiences may reinforce themselves.

Still, it is difficult to disentangle the effects from discrimination and any other factors causing labor market inequalities. What becomes obvious, though, is that if discrimination prevails, it should be regarded as a process rather than a steady state (Altonji and Blank, 1999; Pager and Shepherd, 2008). In other words, discrimination may be experienced prior to initial access into the labor market, i.e., during early skill acquisition, while entering the labor market (focused upon in the present thesis) and thereafter (e.g. with respect to wages and career paths).

4.1.2.2 HUMAN CAPITAL THEORY

According to Becker (1962, 1993) who can be considered the founder of human capital theory, individuals' skill acquisition follows a similar rationale than any other investment decision such as the acquisition of tangible products. Unlike these products, however, human capital is intangible and hard to transfer. Examples encompass investments in schooling or on- and off-the-job training, expenditures to maintain or improve health, the collection of labor market information and migration in order to take advantage of enhanced job opportunities. Theory suggests that human capital investments are rewarded by the labor market and associated with superior outcomes such as higher job seniority and wages (which Becker (1993) also supports empirically). Naturally, the positive effects vary contingent on the amount invested and the rates of return, thus producing differences in characteristics workers supply to the labor market.

Theoretically, given a utility maximizing individual, investments in human capital are undertaken whenever the rate of return is expected to be positive. The profitability, however, depends on the calculated (monetary and non-monetary) benefits as well as direct (e.g. tuition fees) and indirect (e.g. forgone income due to school attendance or participation in on- and off-the-job training) costs. Both benefits and costs are in turn affected by i.) the investment period, ii.) the degree of uncertainty, iii.) the mode of financing and iv.) the individual's ability (Becker, 1993). The former reflects the expected time spent in the labor market. Postponing labor market entry reduces career duration or, in other words, the time investments can be amortized and future gains be realized, and simultaneously carries opportunity costs. As a result, the present value of the investments' net effect decreases which ultimately leads to a negative rate of return. For this reason, individuals shift from learning to earning at a certain point in their lives, that is, they leave school in order to take up employment. Analogously, young workers have a higher incentive than older ones to invest in training activities, simply because they have more time to gain from the associated benefits. In the same vein, women have historically invested less than men in their own human capital as their overall career length in the labor market is expected to be lower due to e.g. child-rearing and other family duties. Thus, if the investment is financed by the employer (which can particularly be observed in case of firm-specific human capital spending), it would be economically rational to prefer men over women eventually resulting in the motherhood gap as reported in section 3.2.1.

By definition, human capital investments also carry a high degree of uncertainty since they are based on beliefs and expectations about future gains and costs. People are uncertain about how long they will actually (be able to) participate in the labor force, what their true abilities are (this especially applies to younger persons), how the labor market rewards their acquired skills, whether rewards change with e.g. technological progress and whether labor market inefficiencies such as discrimination (unexpectedly) enter their investment rationale. Furthermore, the market for human capital follows regularities also found in other capital markets. In particular, individuals face financial constraints that affect their investment decision where large expenditures (e.g. visiting university) are more difficult to afford and internal financing results in wealthier families investing more than poorer ones. Lastly, ability highly correlates with the rate of return and thus affects the extent of human capital investments. Assuming that two individuals had the same earnings without any investment in human capital and faced the same costs, more capable people would invest more since they can realize higher returns from their investment (Becker, 1993).

Adopting Becker's theoretical framework, Mincer (1974) develops an empirical model that relies on schooling and post-schooling investments as the main explanatory variables for annual earnings, since then referred to as the Mincer earnings equation and often used as the basic empirical model in the literature. The basic assumption is that not only pre-labor market, but lifetime human capital acquisition affects the earnings profile. By using data for white, urban, non-student men from the 1960 U.S. census, he empirically demonstrates that in order to correctly specify the relationship between human capital investments and earnings, estimations need to be clustered by schooling group and age cohort. Unlike previous studies that use age as a proxy for on-the-job training, he derives a variable that better reflects people's experience and thus more accurately predicts earnings.[29]

Linking Becker's theoretical considerations with the empirical findings presented in chapter 3 shows that human capital theory provides an appropriate framework for individuals' human capital investment decisions and helps to explain different labor market outcomes across groups. What has only briefly been touched up to this point is that the investment rationale especially during an individual's working career is not necessarily subject to the individual's decision alone, but may be influenced by an employer or induced by law. Knowing that women (at least temporarily) exit the labor market for child bearing and have on average higher absence rates than men, firms would ceteris paribus prefer the latter when it comes to specific training decisions. Similarly, legal regulations may force the employer to pay maternity leave making it more expensive to hire women instead of men. Yet, as will be shown below, either example relies on expectations over group behavior affecting firms' investment decisions and may therefore well point at the prevalence of statistical discrimination. More generally, if, ceteris paribus, access to human capital is systematically restricted for reasons that are based on demographic characteristics, discrimination might be present. Alternatively, differences in human capital endowments might simply arise because skill requirements vary across labor market segments. In this case, group differences in outcome variables only appear if some groups are overrepresented in one segment while others have mainly selected themselves into another segment. This argument is further developed in the next section.

---

29 If no direct information is available, experience can be proxied by deducting the length of schooling plus six (the age at which children usually start going to school) from the individual's age.

4.1.2.3 SEGMENTED LABOR MARKET THEORY

Another reason for different labor market outcomes is posited by segmented labor market theory (also referred to as the dual labor market hypothesis) which argues that the observed differences originate from job- rather than worker-related characteristics (Piore, 1979). Its theoretical foundation is the division of capital and labor. Since, in the short run, capital (e.g. machineries) is fixed, firms adapt their labor demand and reduce working hours or release some of their staff if necessary. However, in order to keep their production running, employers have an incentive to recruit, train and retain a sufficient number of workers that are capable of doing so. Inevitably, these types of workers will have stable and secure employment opportunities, thus constituting a firm's core workforce. As a consequence, all remaining workers bear even greater employment risks and are more likely to be released as a response of a declining demand.[30] The proportion of the latter is greater whenever demand is predictable in a way that allows the standardization of processes. Conversely, wherever the level of standardization is rather low, i.e., where workers perform multiple tasks that constantly need to be readjusted, considerable skills are required.

In short, variations in the production process lead to distinctions among workers and channel them into either a capital-intensive (primary) or a labor-intensive (secondary) sector. The former requires specific human capital investments, thus offering career opportunities and underlining the importance of internal labor markets, while the latter produces workers that are easy to substitute. In the primary sector, workers realize increasing returns to schooling and are compensated for on-the-job training. In contrast, the secondary sector links workers' remuneration mainly to the number of working hours and puts less emphasis on human capital endowments.[30] Jobs in this segment can generally be characterized as unskilled, low paying, involving unpleasant working conditions and carrying considerable insecurity. For either reason, workers have an incentive to move from the secondary to the primary sector.

At this point, it is important to notice that the evolution of segments per se is unrelated to certain industries and occupations. Highlighting the situation of foreign doctors in the U.S., Piore (1979), for example, demonstrates that even in high-qualified jobs 'dualism' may

[30] The fact that decreasing returns to low-skilled labor mitigated the convergence in participation rates and wages of both women and ethnic minorities (see chapter 3) empirically supports the regularities postulated by the segmented labor market theory.

arise. However, differences in skill requirements across industries (e.g. overrepresentation of migrants in construction and automobile jobs in France and Germany) make the occurrence of 'dualism' in some industries more likely than in others.

From a neo-classical perspective, labor market segmentation only evolves from differences in labor supply, particularly the human capital endowments workers bring to the labor market. However, some (groups of) workers may not be able to proceed from the secondary to the primary labor market because labor demand impedes any endeavors of doing so. A theoretical foundation for that is provided by economic theories of labor market discrimination elaborated in the next section.

4.1.2.4 ECONOMIC THEORIES OF LABOR MARKET DISCRIMINATION

While in a market characterized by imperfect information on workers' true productivity, differential treatment unrelated to individuals' actual abilities is sometimes inevitable, systematic discrimination against certain demographic groups is certainly not and, undoubtedly, represents inefficiencies in decision making. According to Aigner and Cain (1977: 178), "[g]roup discrimination in labor markets is evident when the average wage of a group is not proportional to its average productivity". These differences may, on the one hand, directly originate from differential treatment or, on the other hand, result from rules and procedures that have a disparate impact on otherwise equally treated groups, i.e., are disadvantageous to the minority (Pager and Shepherd, 2008). Either way, the empirical findings from chapter 3 (and particularly from section 3.2.3) suggest that the prevalence of discrimination as a major source affecting labor market inequalities cannot be excluded.

Unlike sociological and psychological approaches which are briefly referred to in section 4.1.2.5, economic theories of discrimination use an economic rationale (rather than behavioral patterns) to explain systematic differences in the treatment of individuals and demographic groups. In the literature, two basic frameworks are discussed. According to Becker's (1971) taste for discrimination approach, prejudices against certain demographic groups create disutility that enters the employer's, coworker's and customer's economic rationale and result in inferior labor market outcomes for the disadvantaged group. In contrast, statistical discrimination, as described by Arrow (1971), Phelps (1972) and Aigner and Cain (1977), refers to perceived group differences in worker's productivity due to imperfect information which translates into employers' rationally favoring of one demographic group over another. In the following, both theories will be discussed in detail.

4.1.2.4.1 TASTE-BASED DISCRIMINATION

In his seminal work, Becker (1971) proposes a theoretical framework that relates different labor market outcomes to “tastes for discrimination”. The basic assumption is that individuals have prejudices towards certain gender, ethnic background, social class, religion or personality attributes so that interacting with people who possess one or more of these attributes creates non-pecuniary costs, i.e., causes disutility. These costs are represented by a discrimination coefficient which enters the utility function and thus affects the price determination through market mechanisms. Put differently, individuals are willing to incur costs or forfeit income because they have a taste for discrimination and try to avoid getting in touch with certain demographic groups (recall, for example, the results presented by Szymanski (2000)).

Becker (1971) differentiates three types of taste-based discrimination, i.e., employer, employee (also denoted as coworker), and customer discrimination. According to the first, employers not only include objective and solely productivity-related criteria in decision-making. Instead, based on their personal tastes, they reject working with people from one demographic group while favoring workers from another. As a result the demand for the input factor discriminated against declines and so does its wage. In contrast, demand for non-prejudiced workers increases so that employers have to pay higher wages to the group of workers they prefer. This wage premium can be depicted as follows: π_i $(1+dc_{ie})$, where π_i is the wage rate offered by an employer i and dc_{ie} is the extent to which this employer discriminates, i.e., the discrimination coefficient. Since the increase in wage rates induces an increase in the price of labor as an input factor, aggregate production costs rise. The new equilibrium then generates higher costs that exceed the minimum costs of the previous factor combination. If tastes for discrimination are homogenous, i.e., either non-existent at all or equal across employers, employers face the same production costs from discriminatory behavior. However, in a market with perfect competition, i.e., identical production functions across firms, heterogeneity in the discrimination coefficients benefits employers with weak or no discriminatory preferences. These employers are able to produce at lower costs and can thus outperform their competitors. As a result, prejudiced employers lose market share and, according to Becker (1971), are eventually driven out of the market (which, except the study by Weber and Zulehner (2009), empirical research thus far fails to demonstrate). This process continues until only the least discriminatory firms survive.

As mentioned above, discrimination due to prejudices might not only originate from employers. Even coworkers may have certain distastes towards other demographic groups that creates disutility and causes economic costs. These costs vary contingent on the discrimination coefficient and can be stated as follows: π_j (1-dc_{jw}), where π_j is the wage rate of a worker j and dc_{jw} her respective discrimination coefficient. Hence, coworkers might be willing to compensate their personal distastes by accepting lower wages.

A third type of taste-based discrimination stems from distinct customer preferences. In order to overcome any disutility of buying from a prejudiced group of sellers, customers are willing to pay higher prices at sellers they do not have a prejudice against. Similarly to the case of employers, prices rise with an increase in the discrimination coefficient: p_k (1+dc_{kc}), where p_k is the price customer k pays for the commodity produced and dc_{kc} is the discrimination coefficient against the production factor, i.e., the minority worker involved in the production process. As a result, a taste for discrimination increases the costs of consumption.

In the recruitment context, employer discrimination might be a reaction to either own prejudices or employee (e.g. Haile 2009, 2012, 2013) and customer prejudices (e.g. Neumark, 1996). Especially the latter might have interesting consequences for the hiring outcome. Being aware of coworkers' or customers' distastes, employers might reject individuals from minority groups not because of their own disutility, but because they anticipate conflicts among the workforce or a decrease in sales. Thus, it might be economically rational to disregard minorities during the hiring process or at least offer them lower wages that compensate for the costs incurred by resolving conflicts and foregone sales. In turn, this also demonstrates that the different sources of discrimination are often hard to disentangle, in particular, if only employment ratios or actual wage rates can be observed (see also the discussion in section 3.2.3.2).

Apart from the employment and wage effects of discrimination, Becker (1971) discusses market segregation as a consequence of employers', employees' and customers' distastes. If a sufficient proportion of either party is prejudiced while the rest is not, minorities interact with non-discriminators more frequently than expected by random distribution. Given, for example, a market where discrimination against black workers prevails, this may eventually create a situation in which prejudiced employers hire only white workers that only serve white customers.

Subsequent research relying on Becker (1971) has theoretically shown that the extent of taste-based discrimination varies dependent on different model assumptions on how workers seek employment. In particular, models of random and directed search are distinguished. These models also assume that different tastes either originate from employers (Lang and Lehmann, 2012), coworkers (Sasaki, 1999) or customers (Borjas and Bronara, 1989). In random search models (e.g. Black, 1995; Bowlus and Eckstein, 2002; Rosen, 1997), employers and applicants meet randomly and wages, once negotiated, can be understood as take-it-or-leave-it-offers. Contracts are fixed whenever a satisfying (utility maximizing) wage-match-quality on behalf of either party is reached. However, the wage-match-quality is dependent on employers' prejudice levels. In addition, search costs enter applicants' decision rationale. The idea is straightforward: in the presence of prejudice, equilibrium wages are lower for minority workers. At some point these workers are willing to accept a job offer since costs of further search activities are expected to exceed the benefits from superior future employment contracts. Yet, anticipating minority workers' willingness to accept lower wages more rapidly than majority workers creates an incentive also to non-prejudiced firms to underpay minorities. Hence, the more prejudiced firms operate in a market, the higher is the monopsonistic power of non-prejudiced firms and, consequently, the higher will be the majority-minority wage gap. The inferior treatment by non-discriminators, though, should not be considered as discriminatory in terms of Becker, but is simply an economically rational response to increased market power.

Unlike in random search models, in directed search models (Lang et al., 2005) firms only determine one single wage unconditional on e.g. race (which is more realistic as conditioning wages on demographic characteristics violates anti-discrimination laws in most developed countries) and then choose the most productive worker (adjusted for any disutility they have). Yet, whether an employer is prejudiced or not is ex ante not obvious because prejudice matters only after applications have been evaluated. As certain preferences produce disutility that is incorporated in the productivity assessment, prejudiced workers might face discriminatory conditions. Assuming that workers are homogenous in terms of productivity, in the presence of employer prejudice, candidates from the majority group are always favored over those from a minority. As a result, while random search models help to explain the emergence of wage differences, models of directed search help to explain hiring differences (and are thus crucial for investigating discrimination in access to employment).

From a neoclassical standpoint, Becker's (1971) theory of taste discrimination implies that ultimately discrimination will disappear as competition drives discriminators out of the market. Two scenarios appear plausible: firstly, given a market with perfect competition and a sufficient number of non-prejudiced employers, discriminators suffer from declining demand until bankruptcy as they produce and sell at higher prices than their non-prejudiced competitors. Secondly, in order to remain competitive, employers simply abstain from prejudiced behavior and are thus able to contract workers at the same wages than non-discriminators. The major critique at Becker's approach specifically addresses these long-term consequences. Arrow (1971) argues that discrimination may prevail even in the long run if information asymmetries affect productivity beliefs that differ by demographic groups. This is known as statistical discrimination, a concept that will now be discussed.

4.1.2.4.2 STATISTICAL DISCRIMINATION

The theory of statistical discrimination as advocated by Arrow (1971) and Phelps (1972) claims that in a market of ex ante imperfect information on workers' productivity, otherwise "liberal", i.e., non-prejudiced, employers maximize expected utility from employer-employee interaction based on a priori productivity beliefs where these beliefs are formed based on "surrogate" information. Therefore, three basic conditions need to be met: First, employers should be able to distinguish two groups of workers at reasonable costs, for example, by easily observable characteristics such as race or gender. Second, workers' exact productivity should be ex ante unknown (as it is per definition in a market with imperfect information). Third, employers need to have a priori beliefs on workers' productivity that differ conditional on workers' group membership. For instance, if native workers have proven to be of superior productivity as compared to minority employees, employers would believe that in case of otherwise homogenous job candidates, native applicants' productivity exceeds that of minority applicants (Arrow, 1971).

These beliefs, in turn, may evolve from i.) employers' previous statistical experience, ii.) group differences in predictability of future productivity and iii.) prevailing role models. In case of the former, employers infer an individual's unknown productivity from past experience with members of the same demographic group, where the average productivity of the majority group is generally assumed to exceed that of the minority group (see the example presented above). As a result, minority workers either suffer from inferior hiring outcomes or are paid lower wages. Accordingly, Altonji and Pierret (2001) show that with employer learning on the productivity of minority workers (in their study: blacks) over time, wages increase by the same growth rate as for majority employees (whites). Yet, using group membership as inference for productivity especially seems to be an issue at the hiring stage.

An alternative explanation for this may be what Cornell and Welch (1996) denote as screening discrimination. It assumes that the observability of human capital signals differs across groups which results in employers favoring the group about which they possess most information. Broad empirical evidence suggests that observability is initially better in case of majority workers (e.g. Lang, 1986). However, in order to evaluate whether screening discrimination persists during the course of the employment relationship, static and dynamic models are distinguished. Lundberg and Startz (1983) develop a model showing that groups being subject to more measurement error, i.e., noisier productivity signals, undertake less unobservable human capital investments but, in contrast, have an incentive to overinvest in observable measures such as schooling (see also Lang and Manove, 2011). Altonji (2005) and Bjerk (2008) later introduce a dynamic model of screening discrimination that further explains why hierarchical segregation as a response to different promotion probabilities arises. In particular, the model argues that unequal opportunities in access to higher occupational positions come from employers acquiring productivity information on majority workers more rapidly than on minority workers.

Lastly, socio-cultural role models may create self-enforcing and persisting stereotypes that, in the absence of other productivity-related measures, serve as a suitable productivity device. Coate and Loury (1993) refer to this as rational stereotyping on behalf of employers. In essence, this is what has already been mentioned in section 4.1.2.1 when discussing pre-market differences: negative stereotypes towards minority workers result in lower human capital investments of these workers and, as a consequence, self-enforcing stereotypes. Indeed, the idea is also very similar to the lower equilibrium trap presented in connection with Spence's signaling model. Again, employers' justification stems from the fact that investments by one member of a group produce positive externalities for all other group members and vice versa. Thus, whenever human capital investments and productivity are imperfectly observable and average group investments differ, employers rationally favor members of the superiorly endowed group over those of the inferiorly endowed one. In the end, no matter how beliefs are formed, Arrow (1971) shows that if employers' expectations of mean productivity differ across groups, in equilibrium, differential treatment based on demographic characteristics occurs.

While Arrow (1971) and Phelps (1972) started to relate prior experience with members of a group to employers' expected productivity of this group, the idea has been further developed by Aigner and Cain (1977). They refer to "second moment" statistical discrimination if group differences in the precision of productivity relevant information occur. Employers are assumed to maximize the expected productivity discounted for risk where risk simply reflects the variance in workers' actual abilities. The variance is supposed to be higher for employees from the minority group since, due to inferior knowledge, their productivity indicators (such as test scores) are considered to be less reliable. Higher risk, in turn, creates costs on behalf of employers which directly translates into lower hiring probabilities and wage offers. Workers from the disadvantaged group might overcome the unequal risk distribution by producing additional productivity signals. However, the attainment of further ability signals generates extra costs so that disadvantages remain.

Theoretically, a higher variance in productivity measures could also be of benefit to the minority group. In a situation where the average ability of job applicants is fairly low compared to the market's threshold level, employers are ceteris paribus more likely to hire minority workers because of a higher chance to attract someone who meets the job requirements (which would then be the top performers). In contrast, if employers' threshold level is below the average ability of all candidates, workers from the low variance group (i.e., majority workers) would have an advantage as firms rather prefer a 'safe shot' (Neumark, 2012).

As a consequence of employers' productivity inferences based on group membership, vacancies with high turnover and replacement costs (skilled jobs) are more likely to be filled with employees with higher productivity expectations and more reliable productivity signals. Hence, the employer is less exposed to employment risks. Alternatively, people from minority groups are offered lower wages that compensate for the risk premium the employer has to carry.[31]

4.1.2.5 NON-ECONOMIC THEORIES OF LABOR MARKET DISCRIMINATION

Even though the economic concepts of discrimination are based on employers' prejudices and stereotypes towards certain groups of workers, they do not offer a suitable framework that helps to explain on which grounds prejudices and stereotypes evolve, nor do they address how people's attitudes and beliefs can be measured. This section will

[31] The trade-off between employment and wages given the prevalence of statistical discrimination has recently been established empirically by Dickinson and Oaxaca (2012). With data from a laboratory experiment, they show that while workers with equal mean, but higher productivity variance are discriminated in terms of wages, they are less likely to be unemployed, ceteris paribus.

therefore very briefly provide complementary insights on the causes of discrimination using sociological and psychological approaches.

According to Pager and Shepherd (2008), the reasons why people develop different tastes and stereotypes can be categorized into individual, organizational and structural factors. While the former describes the factors influencing discrimination on an individual level, the latter two ask whether the organizational, societal and political environment reinforce negative attitudes and beliefs. Greenwald and Banaji (1995: 7) define attitudes as "favorable or unfavorable dispositions toward social objects, such as people, places, and policies." In case of unfavorable dispositions, these attitudes are also referred to as prejudices which, as has been demonstrated, provoke tastes for discrimination. A stereotype, on the other hand, "is a socially shared set of beliefs about traits that are characteristic of members of a social category" (Greenwald and Banaji, 1995: 14). Whereas prejudices arise whenever a group of people, e.g. ethnic minorities or women, are negatively evaluated by others, stereotypes encompass judgements that may vary widely depending on which traits people associate with group membership. These traits may, in turn, simultaneously convey both positive and negative attributes. Greenwald and Banaji (1995) illustrate this by using, as an example, cheerleaders who are stereotyped as being attractive, but at the same time unintelligent. Either way, stereotypes are considered to be the basis of statistical discrimination as shown in the previous section.

Prior research further distinguishes between explicit and implicit attitudes and stereotypes. The former are directly measured by self-reported surveys and do not require much explanation. The latter, however, use indirect measures that ask people on a seemingly unrelated issue to assess their unconscious mental associations they have between groups and their attributes. Alternatively, people are invited to take tests constructed to reveal their implicit attitudes and stereotypes. One such example is the implicit association test (IAT) developed by Greenwald et al. (1998). The basic idea is as follows: attributes such as 'hardworking' or 'lazy' are categorized into certain groups such as 'white' and 'black' by hitting a key on a computer. In a 'compatible' treatment, these attributes need to be allocated according to persisting stereotypes, i.e., hardworking to whites and lazy to blacks. In a consecutive treatment, attributes and groups need to be paired counterstereotypically. In the end, the response time differential between both treatments is calculated which then can be interpreted as the implicit association subjects have towards certain groups.

Indeed, previous results documenting people's explicit and implicit tastes and beliefs are sometimes found to contradict each other (denoted as "dissociation"). People may have implicit attitudes and stereotypes which they would explicitly disavow. In the employment context, systematic patterns of implicit behavior benefitting one group over another would thus cause employers to unintentionally discriminate (e.g. Rooth, 2010; Booth et al., 2012). Whereas economic theories of discrimination assign a more active role to the employer, i.e., assume that prejudices and stereotypes are something that is controllable and of which people are aware, Bertrand et al. (2005) argue that the existence of these cognitive factors gives rise to an alternative, non-economic explanation on why labor market discrimination persists. Real-world evidence on market discrimination from tipping New York cab drivers (Ayres et al., 2005), negotiations over sports cards (List, 2004) and decisions whom to shoot in a video-game (Correll et al., 2002) may also stem from people's implicit associations rather than explicit prejudices or beliefs.

Another factor that influences the extent of discrimination is embedded in a firm's organizational structure. Highly formalized processes in hiring, promotion and remuneration, for example, provide an environment where discrimination is expected to be rather rare. The use of objective performance measures such as sales figures when deciding whom to promote or on which basis to fix payment obviously narrow the playing field for discriminatory practices. In contrast, informal and subjective performance evaluations probably leave more room for a treatment unrelated to productivity. Somewhat related to this topic, companies where occupational attainments are closely related to the use of informal networks are more likely to disadvantage minority workers whose average network within a firm is expected to be smaller and less influential (see also the discussion in section 3.1.1). Furthermore, internal measures such as diversity initiatives and the organizational context seem to matter. The former, for example, may be used to actively promote equal opportunities for minority groups (Pager and Shepherd, 2008).

Lastly, structural factors may affect how certain groups are treated in the labor market. Similar to what has been discussed in section 4.1.2.1, Pager and Shepherd (2008) argue that historical legacy and contemporary state policies such as castes in India, the apartheid in South Africa and Jim Crow laws in the U.S., as well as socio-cultural gender roles for e.g. child-rearing responsibilities evoke different preferences across demographic groups when entering the labor market which in turn shape employers' attitudes and beliefs. As a consequence, disadvantages accumulate (prior to entry) in the labor market and discrimination might be reinforced.

To conclude, this chapter has developed a theoretical framework that considers hiring as a decision under uncertainty where employers have imperfect information on workers' productivity at the pre-hiring stage. Furthermore, the chapter has presented economic theories that help to explain different labor market outcomes. Human capital theory relates these differences to differences in workers' endowments while segmented labor market theory attributes them to different workplace characteristics. However, controlling for the implications of these theories, i.e., keeping endowments and jobs constant, might still leave an unexplained gap. Economic theories of discrimination offer a rationale that sheds light on these unexplained differences and that relates inefficiencies to either tastes or productivity beliefs. Next, a conceptual model is developed that formally describes employers' hiring decision accounting for the prevalence of taste-based and statistical discrimination.

4.2 CONCEPTUAL MODEL

From an employer's perspective, an additional employee i is hired whenever her marginal productivity $\partial \mathrm{U}(\mathrm{i})$ exceeds her marginal costs $\partial \mathrm{C}(\mathrm{i})$, where the marginal productivity is determined by the employee's expected future productivity and the marginal costs are determined by a monetary (wage) component as well as a discrimination coefficient that depends on employer's prejudices against the employee's socio-demographic characteristics. Hence, an employer's treatment T whether or not to hire an additional applicant i can be written as follows:

$$\mathrm{T_i} = \begin{cases} 1, \text{if } \partial \mathrm{U_i} \geq \partial \mathrm{C_i} \\ 0, \text{otherwise.} \end{cases}$$

The economic theories of discrimination claim that, all other things being equal, employers either evaluate the expected productivity differently across demographic groups (which is referred to as statistical discrimination) or encounter a disutility when hiring applicants with certain characteristics predetermined by birth (which is described by taste-based discrimination). If either taste-based or statistical discrimination prevail, differential treatment occurs since marginal utility determined by the employee's productivity and marginal costs differ, respectively, and might result in a situation where it is economically rational for employers to hire an additional candidate of one demographic group, but to reject an applicant from the other. The following model referring to Neumark (2012) formalizes this differential treatment.

Let treatment T depend on the applicant's productivity-relevant characteristics P and a dummy variable G that stands for a certain socio-demographic characteristic, e.g. gender.[32]

$$[1] \quad \mathrm{T(P, G) = P + \lambda * G},$$

where G takes the value of 1 if the applicant is female and 0 if he is male. In general, either candidate is hired if her marginal productivity exceeds her marginal costs or, put differently, her expected productivity exceeds a certain threshold level that is a function of work requirements and wage costs. Differential treatment occurs if the applicants either vary in P or if $\lambda \neq 0$. Recall that in a controlled field setting such as the correspondence testing different labor market outcomes due to differences in human capital endowments (according to human capital theory) or occupational positions (according to segmented labor market theory) can be excluded since the applicants are carefully matched and only differ with respect to one specific attribute (here: gender). Now, given that productivity P is the same across groups, any $\lambda \neq 0$ describes discrimination that is purely based on an employer's distaste for either group. If, for example, λ is smaller than zero, women suffer from taste-based discrimination while the same happens to men if λ is greater than zero.

However, any preliminary conclusion with regard to discrimination à la Becker (1971) does not take into account that even though productivity indicators are controlled for within the experimental design of a correspondence study, the perceived productivity might differ across groups and firms. For this reason, the productivity P is split up into three components, i.e., the productivity-influencing factors $\mathrm{X_O}$ which can directly be observed by the employer, the productivity-influencing factors $\mathrm{X_U}$ which cannot immediately (or only at prohibitively high costs) be observed by the employer and firm-specific factors F. Hence, [1] extends to:

$$[2] \quad \mathrm{T(P(X_O, X_U, F), G) = P + \lambda * G}$$

The focus should now shift to the analysis of P. The firm-specific effect F that reflects differences in firms' threshold levels and accounts for intra-firm differences in the evaluation of the applicants can be disregarded given that F is normally distributed and statistically independent of $\mathrm{X_U}$.[33] Assumptions on the candidates' observed and unobserved productivity indicators $\mathrm{X_O}$ and $\mathrm{X_U}$, though, are crucial for the presence of

[32] Note that for simplicity in the present context G is considered as gender, but it could also be replaced by any other demographic characteristic such as race or migration background.

[33] In the empirical section, the estimations are clustered on employer-level to allow for unobserved heterogeneity in employers' decision-making processes and further include firm characteristics to see whether discrimination, if any, is robust across different types of firms.

statistical discrimination. Assume that

[3a] $P_{\text{female}} = X_{\text{O,female}} + E(X_{\text{U,female}})$ and

[3b] $P_{\text{male}} = X_{\text{O,male}} + E(X_{\text{U,male}})$.

If $P_{\text{female}} = P_{\text{male}}$ holds, the coefficient λ displays discrimination, if any, which is based on employers' tastes. However, satisfying this equation requires

[4a] $X_{\text{O,female}} = X_{\text{O,male}}$ and

[4b] $E(X_{\text{U,female}}) = E(X_{\text{U,male}})$

to be fulfilled. Given [4a] is satisfied by the verifiable signals provided in applicants' résumés, e.g. by school grades, employers' expectations on the unobserved productivity-building characteristics [4b] may still vary across gender. If the employer had full information, he would be able to determine [4b] for both of the candidates and, in case of equal preferences, hire the most productive person. Put differently, the firm would be indifferent between either of the candidates if both had the same productivity. However, the unobserved productivity of the candidates is stochastic and might differ in its mean and variance between the two groups.[34]

Employers may use the expected average group productivity as a means of evaluating the unobservables (Arrow, 1971; Phelps, 1972). This might lead to a situation where

[5] $E(X_{\text{U,female}}) \neq E(X_{\text{U,male}})$.

For instance, in male-dominated occupations employers might expect that, even though both candidates offer the same productivity signals, male apprentices are on average more capable to fulfill the requirements (because employers' previous experience with either group indicates men's higher productivity) and are thus preferred over women. If [5] holds, it may bias the extent of λ. In case $\lambda < 0$ (which stands for discrimination against the female candidate in the current example), $E(X_{\text{U,female}}) < E(X_{\text{U,male}})$ would overstate discrimination since employers also incorporate a higher mean productivity of males with respect to X_U in their employment decision. Thus, discrimination is unbiased and relies on λ only if the mean unobserved productivity is expected to be equal across groups. However, even then the results of differential treatment against either group may be

34 Note that a key assumption in correspondence testing is that due to the matching process even the unobservable productivity factors have the same mean, i.e., $E(X_{\text{U,female}}) = E(X_{\text{U,male}})$, which is the essential point of critique issued by Heckman and Siegelman (1993) and Heckman (1998).

misleading and contingent on the probability assumptions of the unobservables.

As proposed by Aigner and Cain (1977), it may well be that both groups are considered to be equally productive, that is

$$[6] \quad E\left(X_{O,male} + X_{U,male}\right) = E\left(X_{O,female} + X_{U,female}\right),$$

but that the variance in the quality of unobserved productivity differs across gender. Assume that the employer has a certain threshold level c and only hires a candidate whose expected productivity exceeds these minimum requirements. Formally,

$$[7] \quad T(P(X_O, X_U, F), G) = P + \lambda * G > c.$$

Given that the threshold level for recruiting any of the candidates is high and that the expected productivity $E(X_O + X_U)$ is equal for the male and female applicant, the employer might still prefer one group over the other even though $\lambda = 0$ holds. For instance, if X_O is set at a moderate level, X_U has to be perceived to be high before an employer is willing to hire any of the candidates. Now, analogously to the example presented in section 4.1.2.4.2, consider that males are expected to have a higher variance in X_U, the employer would correctly conclude that this group is also more likely to produce high achievers that meet the hiring standards. The opposite was true if the threshold level determines a fairly low standard. Then, ceteris paribus, females would on average realize better hiring outcomes as their probability of not meeting the standard is lower.

Both of the aforementioned approaches may lead to differential treatment which is not based on a disutility index, but on information asymmetries that employers try to reduce by making probability assumptions on unobservable productivity factors. That is why these concepts are referred to as statistical discrimination. Hence,

$$[8] \quad T(P_{female}, 1) - T(P_{male}, 0) = \lambda + E(X_{U,female} - X_{U,male})$$

gives the combined effect of taste-based and statistical discrimination if anything else (including other socio-economic characteristics) remain constant. This in turn provides a challenge to the design of correspondence studies and the analysis of their results. Even though both forms of discrimination are illegal and inefficient, there is a need to disentangle the combined effect since both forms are to be tackled by different strategies (see section 6.3). Econometrically, the extent of discrimination can be estimated from the regression

$$[9] \quad T\left(Y_{ij}\right) = \beta_0 + \beta_1 G_i + \beta_2 F_j + \varepsilon_{ij},$$

where $T\left(Y_{ij}\right)$ denotes the hiring outcome for applicant i at firm j, G_i is the gender dummy for applicant i, F is a vector of firm characteristics and ε_{ij} is a normally distributed random variable. Consequently, if $\beta_1 > 0$, the female candidate is more likely to be hired while the opposite is true for $\beta_1 < 0$. Note that the estimation coefficient β_1 only shows whether either party is being discriminated, but does not indicate the source of discrimination, i.e., whether it is based on employers' distastes or differences in information asymmetries. In order to identify the confounding effects of differential treatment, [9] has to be extended to include a set of independent variables that interact with the gender dummy and either represent taste-based or statistical discrimination. Hence,

$$[10] \quad T(Y_{ij}) = \beta_0 + \beta_1 G_i + \beta_2 F_j + \beta_3 G_i * TD_i + \beta_4 G_i * SD_i + \varepsilon_{ij},$$

where $G_i * TD$ depicts the effect that gender and a variable (or a set of variables) indicating taste-based discrimination have on the hiring outcome and $G_i * SD$ is a term accounting for the effect of gender and a regressor considering statistical discrimination. The conceptual model in [10] forms the basis for the empirical model to be estimated using the data generated with correspondence studies on gender and ethnic discrimination in chapter 5.

4.3 Hypotheses

Before the empirical analyses are conducted, testable hypotheses are developed based on the theoretical considerations and existing empirical research. These hypotheses also distinguish between the aforementioned competing approaches of where discrimination might stem from.

To begin with gender discrimination in recruitment, previous findings outside the German labor market have shown that female applicants are disadvantaged in male-dominated jobs, i.e., professions where the share of males is rather high and vice versa.[35] This research is closely related to evidence from the German labor market suggesting that men, for example, are overrepresented in technical occupations no matter whether they require a formal degree or a completed apprenticeship. The latter mainly include jobs as blue-collar specialists in industry. Here, future labor market scarcity is expected to be substantial, though hard to quantify. Nevertheless, considering previous research and the current situation in Germany's labor market for jobs with a male majority, the nature of the job is identified as the main moderator of differential treatment. More precisely, a

[35] Note that a correlation between the gender ratios and the extent of discrimination has not been in the scope of economic research so far and probably varies widely across different labor market regimes.

higher share of men often goes along with either physically demanding (craftsman) or socially stereotyped (computer programmer) jobs. This might be either the result of gender differences in human capital endowments required for these kind of jobs (see section 4.1.2.2), the prevalence of segmented labor markets (see section 4.1.2.3), a selection process (that in turn might stem from pre-market discrimination or the anticipation of lower chances with respect to future hiring outcomes (see section 4.1.2.1)), or discrimination in access to these jobs (see sections 4.1.2.4 and 4.1.2.5). Since the ceteris paribus condition is supposed to be met in correspondence studies (including the equality of observable human capital endowments) and any effects stemming from segmentation, selection and (other) pre-market differences can be neglected due to the experimental character of the study, this leads to the following hypothesis:

> $H_{job\ type}$: *The female applicant realizes fewer callbacks than her male counterpart in male-dominated jobs.*

Previous literature argues on the sources of gender discrimination and uses two economic approaches that help to explain why females suffer from a lower hiring probability in male-dominated jobs, that is, statistical and taste-based discrimination. The former states that discrimination is a rational reaction of employers based on asymmetric information that differs across gender. In other words, an employer is able to form more precise expectations about the future productivity of an applicant who is a member of a group the employer has been contracted and, hence, gathered previous experience with. Having equal productivity indicators of two applicants with different sexes would thus induce the recruiter to rely on additional information inferred from group membership. As this piece of information is more accurate in case of male applicants, females are rejected more frequently and gender discrimination arises.[36]

In order to reduce the importance of group membership, information asymmetries between employers and applicants have to be reduced. Without any unobservable characteristics, the employer would be able to perfectly predict the candidate's future productivity based on the information provided. However, the real hiring process deviates from this ideal situation (see section 4.1.1). Still, the idea prevails that additional productivity related signals increase the reliability of employers' productivity beliefs and therefore decrease the necessity to rely on group experiences as a productivity indicator.

[36] Note that this only holds in the present situation where male-dominated jobs are considered and is supposed to vary contingent on the share of females employed in a specific job.

In the context of male-dominated jobs, this means that the extent of callback differences between male and female applicants is reduced which would lend support to statistical discrimination. Accordingly, the hypothesis states:

$H_{certificate}$: *The provision of additional job-specific information reduces the extent of discrimination against the female applicant in male-dominated jobs.*

Statistical discrimination further claims that applicants should ceteris paribus be treated equally whenever employers' previous experience with either gender is the same with respect to quality and quantity. As previous studies indicate, this rationale holds for gender-neutral jobs in career entry positions where males and females on average realize the same employment outcomes. If, however, males are overrepresented in a particular labor market segment, employers can better evaluate the productivity potential of future applicants. Thus, anything else being equal, employers react economically rational by favoring men over women. Of course, the opposite is true for women in female-dominated jobs. As a consequence, the extent of discrimination against female applicants in male-dominated jobs should decrease with an increasing fraction of women already working in this segment. Since this fraction varies in the German labor market by region, the respective hypothesis can be derived as follows:

$H_{share\ of\ females}$: *The higher the share of female applicants in male-dominated jobs in a specific labor market region, the lower the extent of discrimination against them.*

Alternative to the hypotheses presented above, gender discrimination may be affected by different preferences for either group. As presented in section 4.1.2.4.1, employers may be willing to pay higher wages or forfeit income in order to avoid any disutility arising from working with people that belong to the prejudiced gender. Employers may prefer one group over the other because of their own utility function or as a reaction to the distaste their employees and customers, respectively, might face. Even though these three forms are hard to disentangle, they all lead to worse employment outcomes for the minority group. However, taste-based discrimination comes at a certain price and should differ with the price level. In other words, if an employer is confronted with additional search costs or is likely not to fill a vacancy, he would rather recruit a member from the disliked group, say a woman, than incurring an even greater disutility by continuing the hiring process or leaving the position vacant. In line with this, scarcity in the regional labor market may serve as a proxy for this price mechanism. Whenever in the previous year more jobs were offered than suitable candidates were available, an employer should rather hire people from the minority group, e.g. women in male-dominated jobs, than facing an even greater utility loss. Hence, the following hypothesis is developed:

> $H_{scarcity}$: *The tighter the regional labor market in male-dominated jobs, the lower the extent of discrimination against the female applicant.*

In the same vein, the time interval until a position has to be filled represents a further constraint on behalf of the employer that signals a potential utility loss and may thus proxy potential costs. The more time until the job start elapses, the more search effort the employer has to expend and the higher is his probability of not filling the vacancy. Now, if two types of employers can be observed with one facing a rather long and the other one a rather short interval for staffing, the latter would be exposed to more economic pressure and, if the taste-based approach holds, is therefore expected to discriminate less, if at all. Along these lines, the respective hypothesis is derived:

> H_{timing}: *The shorter the time required for the vacancy to be filled, the lower the extent of discrimination against the female applicant in male-dominated jobs.*

As both, the study on gender as well as ethnic discrimination are conducted using the correspondence method and as both investigate discrimination in the same labor market segment, the development of the hypotheses referring to ethnic discrimination is very similar to that of the hypotheses presented above. The majority of matched-pair field experiments inside and outside Germany conclude that ethnic minorities (first and second generation Turkish immigrants in case of the German labor market) experience worse employment outcomes with respect to hiring probabilities (even though e.g. human capital endowments have been carefully controlled for). Based on these results that unequivocally point at ethnic discrimination in access to employment, the applicant with a Turkish migration background who represents the ethnic minority in the current study is expected to realize fewer callbacks compared to the German male candidate.

> $H_{minority}$: *The Turkish-named applicant realizes fewer callbacks than his German-named counterpart.*

Unlike the quite homogenous findings on the general prevalence of discrimination against ethnic minorities, the economic explanations for differential treatment are rather heterogeneous with a focus on the competing approaches of statistical and taste-based discrimination, respectively. In line with the conception of the study on gender discrimination, on the one hand, the provision of additional productivity signals and, on the other hand, the share of foreign applicants should serve as proxies that indicate the presence of statistical discrimination. The respective hypotheses can thus be formulated as follows:

> $H_{certificate}$: *The provision of additional job-specific information reduces the extent of discrimination against the Turkish-named applicant.*

> $H_{share\ of\ foreigners}$: *The higher the share of foreign applicants in a specific labor market region, the lower the extent of discrimination against the Turkish-named candidate.*

Again, employers, coworkers and customers, respectively, may also have different preferences for, e.g., native Germans and German-born Turks. Different preferences ceteris paribus map into different utility functions for working with or being served by either group and, as a result, produce hiring differentials. The economic pressure due to labor market scarcity, for instance, puts these tastes into a perspective and creates a tradeoff between two options, i.e., hiring a member of the prejudiced group or facing further staffing costs. Thus, taste-based discrimination persists whenever the extent of differential treatment between the majority and minority group decreases as a reaction to either an increase of labor market scarcity or a decrease of the time until the vacancy has to be filled. Referring to the case of ethnic discrimination then yields:

> $H_{scarcity}$: *The tighter the regional labor market, the lower the extent of discrimination against the Turkish-named candidate.*

> H_{timing}: *The shorter the time required for the vacancy to be filled, the lower the extent of discrimination against the Turkish-named applicant.*

Overall, the hypotheses developed above address the underlying research questions of this thesis. On the one hand, they focus on the prevalence of gender and ethnic discrimination in a certain segment of the German labor market ('$H_{job\ type}$' and '$H_{minority}$'). On the other hand, they postulate potential effects that allow identifying the factors influencing differential treatment ('$H_{certificate}$', '$H_{share\ of\ females/foreigners}$', '$H_{timing}$' and '$H_{scarcity}$').

5 EMPIRICAL ANALYSES

The empirical section presents the results from both the correspondence study on gender and the one on ethnic discrimination in the German labor market. Since the experimental design is the same for both investigations, it is described in detail first (5.1). After that, the results of the gender (5.2) and ethnicity (5.3) study are presented and discussed separately before the consequences of methodological variations on the results of such field experiments are addressed (5.4).

5.1 EXPERIMENTAL DESIGN AND PROCEDURE

As already mentioned in chapter 3, the experimental design of a correspondence study needs to account for local labor market characteristics and application standards and thus differs among countries, job types and seniority levels. Besides, the study should allow a reproduction of the results by implementing the same framework in future research. Therefore, in the following, a thorough presentation of the design and the procedure adapted in both field experiments is provided.

5.1.1 JOB MARKET FOR APPRENTICES

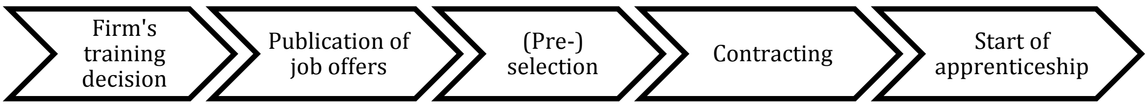
The correspondence studies conducted for the present thesis refer to the job market for apprentices. Its suitability for matched-pair testing, importance for the German labor market and latest developments will be outlined in the following sections.

5.1.1.1 SUITABILITY FOR CORRESPONDENCE TESTING

Investigating hiring discrimination in Germany requires a proper selection of the experimental framework. More precisely, the jobs focused upon using correspondence testing have to fulfill three main criteria. First, demand for labor must be sufficiently high so that an appropriate number of callbacks can be expected. Second, contract type and occupations have to be of particular importance for employers as well as for employees. Third, data on applicants' employment history must be kept to a minimum. The more information on e.g. prior labor market experience, unemployment spells and family breaks is provided, the higher is the risk of running into problems of an unobserved heterogeneity bias. In addition, supplemental information generally requires the attachment of additional credentials which in turn increases the likelihood that employers get suspicious of the deceptive nature of the correspondence method.

A labor market field that meets all these criteria and literally seems to be designed for correspondence testing is the labor market for apprenticeships. In the context of the dual training system in Germany, people learn a certified profession according to certain curricula during a period of 2.5 to 3.5 years. During this time, the apprentices partly visit vocational school and partly work for the training company they are employed in. Apprenticeships are also quite homogenous with respect to several other factors. The training programs start yearly, usually in August and September. However, job offers are published the entire year. While some employers recruit almost a year in advance (in the following referred to as 'early recruiters'), others offer their positions rather late (and are accordingly denoted as 'late recruiters'). Remuneration of the apprentices is typically settled by collective bargaining agreements and does not vary across apprentices applying for the same job.[37] The figure below illustrates the process that takes place before the apprenticeship contract is settled.

Figure 5-1: Application and Selection Process



Source: Own illustration.

Even though employers do not have a legal obligation to train apprentices, in 2011, 52.6 percent engaged in training activities (BIBB, 2012b).[38] Research investigating firms' decisions of whether or not to offer apprenticeship training usually distinguishes between investment and production strategies (e.g. Niederalt, 2005; Dionisius et al., 2009; Mohrenweiser and Zwick, 2009; Backes-Gellner and Mohrenweiser, 2010). The former considers apprenticeships as a means to circumvent asymmetric productivity information, to reduce hiring costs and to increase profits by paying the apprentices below their marginal product after the training period has ended. Consequently, these types of employers are more likely to extend their apprentices' contracts. On the other hand, firms following a production strategy use apprentices as cheap labor and do generally not offer

[37] Occupational variations in apprentices' pay, though, are common, but do not require further discussions as applicants are matched. Wages also differ slightly by region (e.g. East-West disparities) as they correspond to the local living standards. Yet, these differences are negligible. For information on the legal framework of apprenticeship contracts, see the Vocational Training Act (BBiG).

[38] The ratio of companies offering vocational training increases with firm size. Firms employing more than 200 people are found to have the highest training ratio (BIBB, 2012b).

permanent contracts after completion of the apprenticeships.[39]

According to the cost-benefit survey by the Federal Institute of Vocational Education and Training (BIBB) from 2007 where employers (N=2,986) self-reported the economic rationale behind their training decision, firms on average incurred net costs of around 3,600 Euros per apprentice and year (BIBB, 2009a).[40] However, these costs decrease over time and are eventually recovered by savings for not having to recruit qualified staff from the labor market and by the fact that former apprentices initially perform better than external recruits due to the specific human capital acquired. Moreover, employers mention the positive labor market signal that is sent out by the provision of vocational training as another reason for why they offer apprenticeships (see e.g. Backes-Gellner and Tuor Sartore (2010) for the signaling effect of apprenticeships). Employers' responses thus indicate that training is predominantly used to select qualified staff, decrease the probability of adverse selection, ensure future labor supply and build up reputation in the labor market which all go along with the aforementioned investment rather than a production strategy (BIBB, 2009a). Based on their productivity expectations gathered during the apprenticeship, employers have the possibility to offer a permanent contract at the end of the training period. Thus, from an applicant's perspective, being hired as an apprentice means having a foot in the door to future employment.[41]

From an individual level as well as a macroeconomic point of view, the labor market for apprenticeships matters: experts all over the world consider the dual system in Germany as a key ingredient for an ongoing supply of well qualified employees and specialized staff which in turn forms the ground for a fairly robust labor market in times of the international debt crisis. That is also why, in 2004, the German government together with employer representatives decided on an agreement (the so-called "Nationaler Pakt für Ausbildung und Fachkräftenachwuchs in Deutschland") which ensures that every

---

39 In line with employers' motives, Wenzelmann (2012) finds different allocations of productive and non-productive work tasks assigned to apprentices, which seem to depend on firms' training strategies and apprentices' educational endowments.

40 Analyses of employers' net costs indicate that medium-sized employers (50-499 employees) have significantly lower net costs per apprentice than small firms (10-49 employees) and that net costs are higher in the West compared to the East. Net costs, on the other hand, are not affected by job type (industry versus office jobs) and number of apprentices in a firm (BIBB, 2009a).

41 The Confederation of German Trade Unions (DGB) has been calling for inclusion of subsequent employment guarantees in apprenticeship contracts. Results from the 2007 survey further show that the ratio of firms extending the work contract (on average 57%) is highest in manufacturing (69%), in Eastern states (63%) and in large firms (89%) (BIBB, 2009a). For an empirical analysis investigating which employer characteristics affect the probability that an apprentice is offered a permanent contract after completion of the apprenticeship, see Bellmann and Hartung (2010).

applicant who is willing and capable to take up an apprenticeship receives an opportunity to do so (BA, 2005, 2007, 2010c).[42]

However, similar to the regular labor market, the market for apprenticeships is characterized by a certain degree of regional, occupational or educational mismatch causing apprenticeship positions to remain vacant. In the apprenticeship year 2010/2011, 34.8 percent of all training firms were not able to staff any or some of their vacancies offered. According to the BIBB (2012a), 67.8 percent of these firms note that applicants did not meet the company's educational requirements. This is the reason why they sometimes withdrew their job offers. Another 26.2 percent simply did not receive enough applications. Among the employers with unfilled vacancies, firms from Eastern Germany, rural areas and regions with a low degree of tertiarization as well as small-sized employers are overrepresented. Undoubtedly, these differences partly reflect difficulties in how to reach employers' locations (e.g. the availability and quality of public transportation is likely to be better in urban compared to rural areas so that apprentices find it more difficult to commute if employers are located outside metropolitan areas) and applicants' reservations against certain jobs and branches. Lastly, employers reported that 12.5 percent of the apprentices selected resigned before the apprenticeship started. In addition, about one fourth (23 percent) of all apprenticeship contracts were canceled during the training period (BIBB, 2009b, DIHK, 2011, BIBB, 2012b).[43] Both, unoccupied vacancies and early termination of employment relations create costs the employer tries to minimize. This, in turn, outlines the importance of proper apprentice recruitment and selection procedures.

In 2010, on average around 55 percent of an age cohort started an apprenticeship for the first time. However, this share significantly varied across gender (66.1 percent of all German males started an apprenticeship as opposed to 49.0 percent of German females) and nationality (57.8 of German graduates compared to only 29.5 percent of graduates with foreign nationality signed an apprenticeship contract) (BIBB, 2012b). Table 5-1 gives an overview of the characteristics and job choices of the applicants for an apprenticeship in the reporting periods 2009/2010 until 2011/2012. According to these figures, every year roughly 550,000 people applied for an apprenticeship. These numbers depend on the

[42] In 2010, this agreement was extended for the second time and to date lasts until 2014 (BA, 2010c).

[43] See BIBB (2009b) for differences between training firms with and without unfilled vacancies as well as reasons for the dissolution of contracts during the training period.

business cycle, the share of people going to university and the fact that a recent school reform doubled the share of school graduates in some states (BIBB, 2012b). Among these applicants, roughly 45 percent were females and between 11.0 and 11.6 percent were non-Germans. The largest proportion of foreigners was represented by Turks who accounted for almost 50 percent of the people from abroad. With respect to applicants' age and their educational endowment, table 5-1 shows that more than 40 percent finished middle school and around 65 percent were younger than 20 years at the time of their application. Around 60 percent of the apprenticeships addressed service apprenticeships while approximately 37 percent were dedicated to jobs demanding technical tasks.

Table 5-1: Characteristics and Job Choices of Applicants for Apprenticeships by Reporting Period

	Fraction in %	**Fraction in %**	**Fraction in %**
	2009/2010[1)] (N=552,168)	**2010/2011[1)]** (N=538,245)	**2011/2012[2)]** (N=559,877)
Females	45.4	44.9	44.9
Foreigners	11.0	11.2	11.6
(Turks)	(5.3)	(5.4)	(5.5)
Aged under 20	64.1	65.2	65.9
Middle school	41.5	42.4	42.5
Technical apprenticeships	37.4	37.0	36.9
Service apprenticeships	59.5	60.2	57.8

Notes: Technical and service apprenticeships are classified according to the job classification of the BA from 1988[1)] and 2010[2)], respectively. A reporting period lasts from October 1st of the previous until September 30th of the following year.
Source: BA (1988, 2010a, 2010b, 2011, 2012b).

Descriptive statistics of applicant characteristics across these two job types clearly highlight gender differences (see table 5-2). Male applicants dominate technical apprenticeships (approximately 85 percent) while service apprenticeships have a majority of female job candidates (63.5 percent). With respect to the share of foreigners and middle school graduates, however, only minor differences among the job types can be identified.

Table 5-2: Characteristics of Applicants for Apprenticeships by Job Type for the Reporting Period 2010/2011

	Fraction in %	**Fraction in %**	**Fraction in %**
	All apprenticeships (N=538,245)	**Technical apprenticeships** (N=199,063)	**Service apprenticeships** (N=323,756)
Females	44.9	14.8	63.5
Foreigners	11.2	10.0	12.4
Middle school	42.4	40.9	43.6

Notes: Difference to 100 due to omitting apprenticeships from the agricultural and mining sector.
Source: BA (2011).

Apart from the fact that apprenticeships are meaningful to both employers and apprentices, they are quite suitable for the correspondence testing since they address entry-level jobs. This implies that the majority of people who apply for an apprenticeship are career starters who have recently graduated from or are in their last year at school. As a consequence, only a limited employment history needs to be designed and the amount of credentials can be kept to a minimum. With respect to gender differences this also implies that the expected costs of maternity leave do not enter employers' decision rationale and can therefore be neglected.

5.1.1.2 SCOPE OF APPRENTICESHIPS IN PRESENT STUDIES

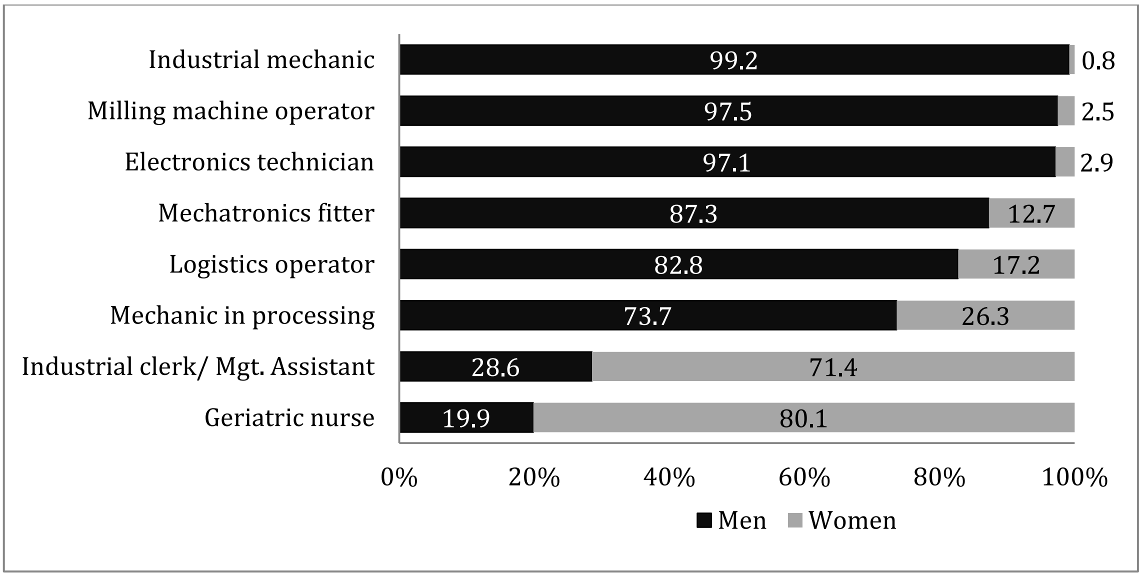
Both the gender and ethnicity study focus on technical apprenticeships. In particular, six rather technical training professions that belonged to the 50 most frequently chosen apprenticeships in 2010 are addressed, i.e., industrial mechanic (German: Industriemechaniker/-in), mechatronic fitter (Mechatroniker/-in), milling machine operator (Zerspanungsmechaniker/-in), mechanic in plastics and rubber processing (Verfahrensmechaniker/-in für Kunststoff- und Kautschuktechnik), electronic technician (Betriebselektroniker/-in) and warehouse logistics operator (Fachkraft für Lagerlogistik). In case of the investigation on gender discrimination, this range of jobs is extended by apprenticeships as geriatric nurse (Altenpfleger/-in), industrial clerk (Industriekaufmann/-frau) and management assistant for office communication (Kaufmann/-frau für Bürokommunikation) which, from the apprentices' perspective, belong to the 20 most favored jobs in the same year (BIBB, 2010b).[44] Comparing full-time employees working in the jobs considered for subsequent investigations reveals huge variations in the fraction of females. These variations justify a classification into male- and female-dominated jobs. The former include technical occupations where the share of females varies between 0.8 and 26.3 percent while the latter comprise service jobs with a share of women above 70 percent.[45] With respect to the distribution of foreigners across occupations, no obvious differences emerge. A closer look at the share of certified employees, though, reveals substantial differences across jobs with a range between 50

[44] Overall, 348 certified apprenticeship professions were listed in 2010. This number remained constant over the last decade (BIBB, 2012b).

[45] The data for full-time employees are supported by the figures for new apprenticeships. In the years 2009 until 2011, the fraction of women starting an apprenticeship in service professions was roughly between 60 and 80%. In male-dominated jobs, however, only between 4.4 and 11.5% of the new hires were female (BIBB 2010a, 2011a, 2012b).

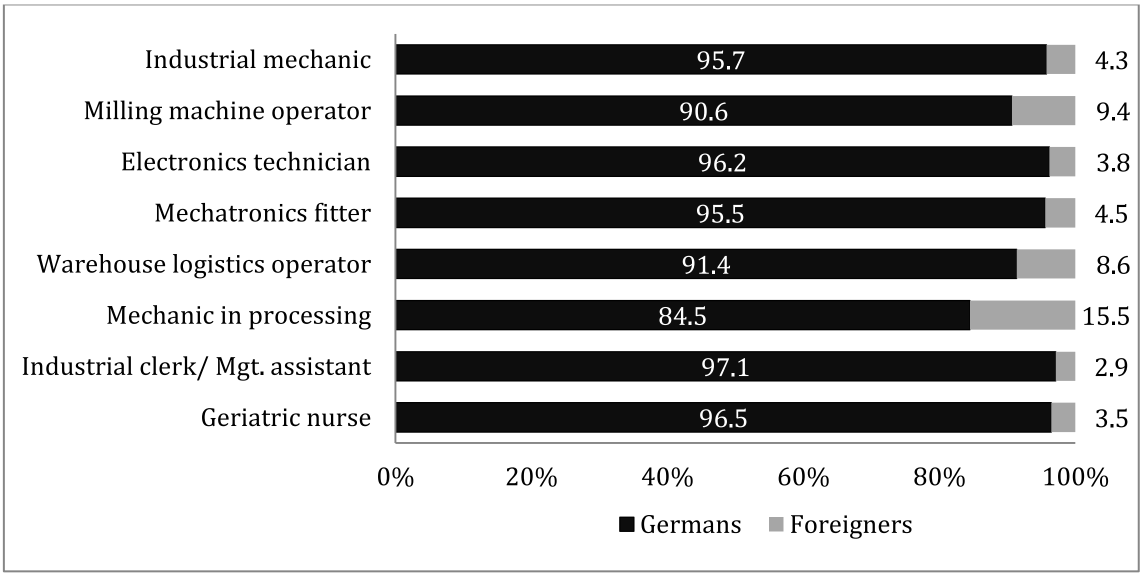
and 90 percent (see figures 5-2, 5-3 and 5-4).

Figure 5-2: Full-Time Employees in Selected Jobs by Gender



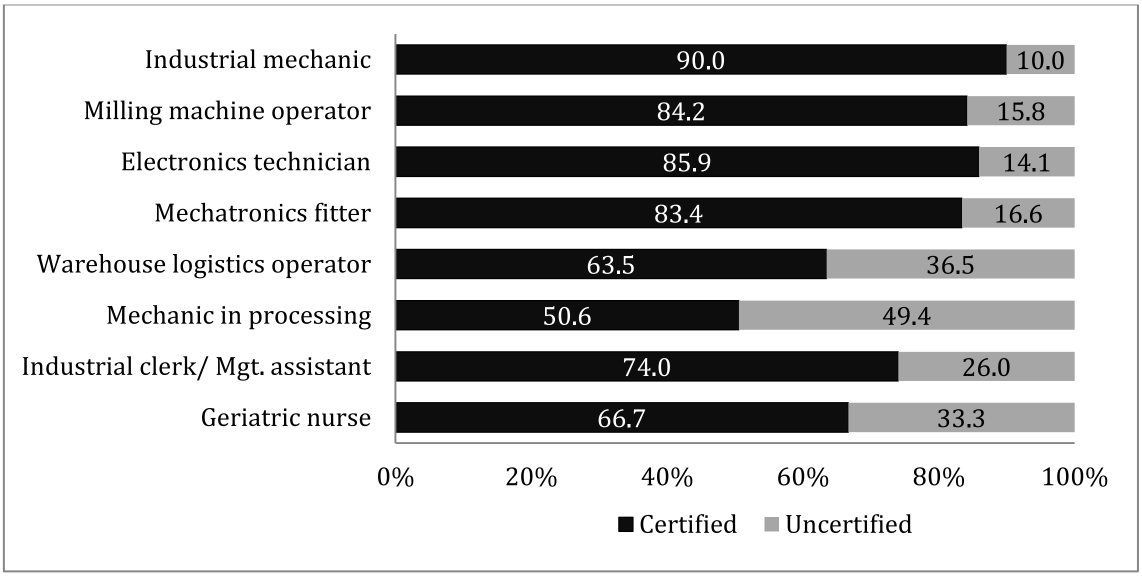
Notes: For industrial clerks and management assistants for office communication no disaggregated data are available. Proportions denote an unweighted average of the years 2005, 2007 and 2009.
Source: Own illustration based on BA (2012d, 2012e, 2012f, 2012g, 2012h, 2012i, 2012j, 2012k).

Figure 5-3: Full-Time Employees in Selected Jobs by Citizenship



Notes: For industrial clerks and management assistants for office communication no disaggregated data are available. Foreigners denote all non-Germans. Proportions denote an unweighted average of the years 2005, 2007 and 2009.
Source: Own illustration based on BA (2012d, 2012e, 2012f, 2012g, 2012h, 2012i, 2012j, 2012k).

Figure 5-4: Full-Time Employees in Selected Jobs by Certification



Notes: For industrial clerks and management assistants for office communication no disaggregated data are available. Certification refers to all people who have successfully finished an apprenticeship. Proportions denote an unweighted average of the years 2005, 2007 and 2009.
Source: Own illustration based on BA (2012d, 2012e, 2012f, 2012g, 2012h, 2012i, 2012j, 2012k).

5.1.2 VACANCIES

In this section, the access to and the requirements of job offers addressed by the applicants within the correspondence studies are presented. The vacancies for the apprenticeships were taken from the job platform of the German Federal Employment Agency. Weitzel et al. (2011a, 2011b) show that approximately 77 percent of all employers place their employment advertisements online.[46] Applications referred to apprenticeships starting in 2011 and 2012, respectively, and were sent out at three different points in time, i.e.,

- May 2011 for apprenticeships starting in August or September 2011,
- September 2011 for apprenticeships starting in August or September 2012 and
- May 2012 for apprenticeships starting in August or September 2012.

Due to the fact that different application periods were referred to, the study allows a comparison over time and addresses both firms that recruit rather early and offer new positions almost one year in advance (i.e., early recruiters) and firms that publish their job offers on a short notice and start selecting their applicants two to three months prior to

[46] A report by the BIBB (2011b) further outlines that the BA is the dominating recruiting channel among training companies. For a more detailed overview of recruitment channels, methods and strategies, see BIBB (2009b, 2011b).

the start of the apprenticeships (i.e., late recruiters).

The potential workplaces were located all over Germany both in the public and private sector.[47] In order to facilitate administration and keep costs to a minimum, job offers were only answered when the employer accepted email applications. This way of getting into touch with employers has been growing in popularity within the last decade and is more and more favored by both firms and applicants. Apart from that, email applications are accepted independent of firm size, sector and location (Weitzel et al., 2011a, 2011b).

Apart from job, time and contact restrictions, the advertisements had to require no prior work experience and no more than ten years of schooling (which implies that applicants were graduates from lower or middle school). Firms further encouraged the applicants to voluntarily provide additional credentials of internships, for instance. School certificates, on the other hand, were required and would have substantially reduced the response rate if left out. In addition to these formal requirements, most employers consider the applicant's passion for the respective profession as well as soft skills such as the ability to work in teams, having a high degree of intrinsic motivation and work accuracy as a necessary condition to apply for the job.

5.1.3 MATCHING PROCESS

Each application consisted of a CV, a cover letter and the last three school certificates. The CVs were matched according to age, the socio-economic area of residence, schooling, language skills and leisure time activities and only differed with respect to gender and ethnic background, respectively. Cover letters stated the candidates' motivation, skills and abilities for the job. Depending on the application period, the candidates were aged 15 or 16 and came from cities in the states of Lower Saxony (Brunswick, Hanover, Hildesheim), Hesse (Kassel) and North Rhine-Westphalia (Paderborn), respectively. The candidates were all German-born and stated German as their mother tongue as well as a good command in English. In addition, the résumés signaled the same IT skills which were altered depending on whether a white- or a blue-collar apprenticeship was addressed. Leisure time activities highlighted gender neutral sports such as handball and running and indicated a passion for hobbies that had a link to the corresponding profession such as, for

[47] Since the majority of apprentices still live with their families and most firms require applicants living in the company's neighborhood, the candidates stated that they were about to move with their family close to the location of the respective workplace. No statement about the relocation would have reduced the number of positive callbacks substantially and/or would have resulted in many inquiries on behalf of the employers.

example, the membership in the voluntary fire brigade for technical apprenticeships.

With respect to schooling, the applicants mentioned that they were currently in their last year of middle school. School certificates showed above-average grades in subjects that were considered as meaningful in the job offers such as math, technics and physics in technical occupations. A randomly chosen number of applications sent out in the second and third application period (i.e., in September 2011 and May 2012) also included information on a certified school internship in the respective industry. In Germany's lower and middle schools such internships are obligatory one year prior to graduation and usually last two to three weeks. Students use this opportunity to gather first practical experience. As mentioned above, in applications for apprenticeships, employers do generally not require such information. However, providing a certified internship might serve as an additional productivity device. Whenever attached, certificates on internships stated favorable information on candidates' working behavior and effort. They further outlined the intern's positive work attitude as well as his/her strong interest and intrinsic motivation. Due to the random allocation, certificates were provided by none, either or both of the candidates. This variation permits an isolation of the effect an additional signal has on the hiring outcome. In order to avoid any legal issues, the certificates were of fictitious schools and companies.[48]

To allow an unambiguous identification of employers' responses, all job candidates received individualized contact details: an email address, a cell phone number and a postal address. Phone calls were answered by voicemail which kindly asked the caller to leave a note with name and contact information. Postal mails were redirected to the researcher's address. In order to rule out any suspicion on behalf of the employers, pairs of applications were sent out with one to two days in between. In addition, the résumés, cover letters and certified internships slightly differed concerning layout and wording. Overall, three different designs of applications were prepared. By randomly altering the application forms across candidates and jobs, any bias due to differences in framing and dispatching orders could be controlled for.[49]

[48] Note that firms' responses did not indicate any suspicion due to fictitious certificates. Section 5.4 explicitly discusses any potential suspicion bias of the correspondence method and tests methodological variations.

[49] Examples of résumés and cover letters can be found in section B in the appendix.

5.1.4 NAMES AND PROFILE PICTURES

The correspondence method relies on applicants that only differ with respect to one feature. Here, differential treatment due to gender and ethnic differences is investigated. It is crucial to the study that these characteristics can unequivocally be identified by reading the applications. The identification of applicants' gender and ethnic origin is usually done by changing names and profile pictures (at least in case of gender studies and only where the attachment of profile pictures is common practice as is the case in Germany).

In both studies, the male candidate without a migration background is considered as the reference category and is given the name Jan Lange and Lukas Schmidt, respectively. The first names both belong to the 20 most frequently chosen first names in Germany at the beginning of the 1990s and the surnames can also be found among the 20 most common ones in Germany. Accordingly, the names of the female candidate, Anna Schneider and Laura Müller, are determined.[50] Like in prior correspondence studies on ethnic discrimination, names also serve as an indicator for ethnic background. Since the ethnicity study explicitly focuses on German born males who belong to the second or third generation of formerly immigrated Turks, the candidates' names are among the most common Turkish names in Germany, Kenan Yilmaz and Onur Öztürk.[51]

Applications also include profile pictures which all have a similar format and style concerning background colors, coiffures and facial expressions. The photos are characterized by a light background, candidates show a friendly smile, have a similar dress and the same hair color. In case of the matched pairs in the ethnicity study, the photos were also randomly varied across candidates to exclude any potential beauty bias. All in all, two different male and female profile pictures were used and controlled for in the multivariate analyses.

---

[50] For the selection of German-sounding first names, see http://www.beliebte-vornamen.de; for the selection of German-sounding last names, see http://www.bedeutung-von-namen.de/top50-nachnamen-deutschland.

[51] For the selection of Turkish-sounding first names, see http://www.baby-vornamen.de/Sprache_und_Herkunft/tuerkische_Vornamen.php; for the selection of Turkish-sounding last names, see http://www.herkunft-name.de/namensherkunft-familienname/nachnamen-international/tuerkische-nachnamen.htm. When choosing the names, those that are attached to prejudices or stereotypes were tried to be avoided. Name effects are tested by a subsample (see respectively tables C-3, C-4, C-10, and C-11 in the appendix), but are not found to be significant and meaningful for the results of the present studies. For a more elaborate empirical investigation of name effects, see e.g. Fryer and Levitt (2004).

5.1.5 Application Process and Response Documentation

Two applications (the German male as the reference category together with either the female or the ethnic minority candidate) were sent out in response to each job offer.[52] Cover letters, CVs and certificates were matched automatically using serial letters. As mentioned before, designs and emailing orders were randomly varied before the applications were dispatched. Across all application periods, firms were addressed only once although some offered different apprenticeships at the same time.

Employers' responses were then carefully reported for the consecutive three (in case of the applications sent out in May 2011 and May 2012) and nine months (for applications sent out in September 2011), respectively. The records included the date and the type of response (see below), as well as sex, name and position of the person responding (whenever possible) and were then complemented by information about the job offers such as job as well as firm characteristics. The firms replied via email, postal mail or phone. The answers can be classified into five different categories: either the applicant (i) did not receive any response, (ii) got an acknowledgement, (iii) was requested to provide additional information, (iv) was rejected or (v) was signaled interest on behalf of the employer which is subsequently referred to as a 'callback'.

A reminder was sent out to those companies that had not replied at all after three weeks. Acknowledgements mostly stated that the firm would check the documents and make a statement after having reviewed all incoming applications. Thus, some acknowledgements were followed by a response, i.e., either a rejection or a callback, on behalf of the employer. However, some firms never called back again and were therefore regarded as a case of no response. Rejections remained unanswered by the candidates whereas callbacks were politely withdrawn (with the note that the candidate already found another apprenticeship) within 48 hours to avoid any further inconvenience and costs to the companies. Callbacks, for instance, took the form `we would like to get to know more about you in a personal interview´ or `please call back so that we can arrange a job interview'. Overall, they are defined as either an invitation to a selection interview, a telephone interview, an assessment center or an offer for an internship. In the next section, the results from the gender study will be presented, analyzed and discussed.

---

52 In the remainder of the thesis, the female (Turkish-named) candidates are always considered and referred to as the minority group.

5.2 CORRESPONDENCE STUDY ON GENDER DISCRIMINATION

In what follows the correspondence study on gender discrimination in the labor market for apprenticeships in Germany is dealt with. First, the dataset (5.2.1) and descriptive results are presented (5.2.2). The subsequent section outlines the econometric method and conducts analyses on the employment outcomes for all job candidates (5.2.3). After that, the hypotheses developed in section 4.3 are tested and, finally, discussed (5.2.4). The discussion includes interpretations of the results and relates them to economic theories of discrimination as well as to previous findings on gender discrimination.

5.2.1 DATA

This section, on the one hand, presents the dataset generated by the field experiment and used for the empirical analyses (5.2.1.1) and, on the other hand, compares company characteristics of the dataset with those from the entire body of training companies in Germany (5.2.1.2).

5.2.1.1 THE DATASET FROM THE FIELD EXPERIMENT

Overall, 664 job offers were addressed which, due to the matched-pair setting, resulted in 1,328 individual applications. Since in case of 8 employers, dispatching errors were reported, the corresponding 16 applications were excluded from further analyses.

The main outcome variables show that in 81.6 percent of all applications, firms informed the candidates of whether or not they were invited. In other words, 1,070 times the applicants either received a rejection or a callback (subsequently denoted as a 'response'). Among these, 37.9 percent of all applications were answered by a callback. Whenever employers responded, it took them on average 23.8 working days with some answering immediately while the maximum waiting time was 178 working days. Employers used all three possible options to get in touch with the applicants. However, email responses dominated (65.3 percent).

Among the remaining 656 firms addressed, 52.7 percent were located in the South of Germany, 17.5 percent in Eastern Germany and 29.7 percent in the remaining states.[53] The

[53] The difference to 100% is due to rounding errors. The South of Germany includes the states of Bavaria, Baden-Wuerttemberg, Hesse, Rhineland-Palatinate, and Saarland. Eastern Germany covers the states of Berlin, Brandenburg, Mecklenburg-Western Pomerania, Saxony, Saxony-Anhalt, and Thuringia. Hence, the remaining states are Bremen, Hamburg, Lower Saxony, North Rhine-Westphalia, and Schleswig-Holstein.

majority of companies (76.7 percent) belonged to the industry and construction sector while 23.3 percent are in other sectors such as trade, services and public administration. The highest fraction of firms in the sample, around 51.5 percent, was of medium size, i.e., employed between 50 and 500 workers at the time of the study. The rest were either small companies with less than 50 employees (33.2 percent) or large companies with more than 500 employees (15.2 percent).

Applications were sent out at three different points in time. The first application period in May 2011 contained 246 (37.5 percent) distinct firms, the second period in September 2011 included 262 (39.8 percent) firms and the third period in May 2012 addressed 149 (22.7 percent) different employers. Thus, late recruiters as defined in section 5.1.1.1 made up 60.2 percent of the entire sample. While small firms accounted for the highest share among late recruiters (45.6 percent), they represented the lowest portion among the job offers already published in September (14.6 percent). In contrast, medium and large companies were overrepresented among early recruiters compared to the fraction they made up in May 2011 and 2012 (see table 5-3).

Table 5-3: Firm Size by Application Period

	Late (N=395)	Early (N=261)	Total
Small	45.57%	14.56%	33.23%
	(180)	(38)	(218)
Medium	45.06%	61.30%	51.52%
	(178)	(160)	(338)
Large	9.37%	24.14%	15.24%
	(37)	(63)	(100)

Notes: The table reports late and early recruiters by firm size in percent. Absolute numbers are in parentheses.
Source: Own dataset.[54]

The majority of apprenticeships the candidates applied for were technical occupations such as industrial mechanics. Recalling that men predominantly fill these kinds of jobs, they can be classified as male-dominated. Accordingly, those apprenticeships that have a higher fraction of women are considered as female-dominated. The latter represent 17.7 percent in the sample and were only referred to during the application period in May 2012 in order to be able to test for job stereotyping (‘$H_{job\ type}$’). The job offers also indicated the

[54] If not stated differently, the sources of all subsequent tables and figures are the datasets generated during the course of the correspondence studies.

number of apprenticeship positions the employers offered as well as the number of positions that were still available. The firms assigned up to 15 apprenticeships where on average 1.7 positions had not yet been filled at the time of application. In more than half of the cases (53.2 percent) the person responsible for the applications was female.

Even though the correspondence testing matches the candidates on relevant characteristics, names, profile pictures and contact data need to differ in order to avoid suspicion and to be able to unequivocally record companies' responses. However, name and beauty effects may bias the results on gender discrimination. Therefore, within a subsample two distinct male and female names as well as photos were chosen and incorporated. That is why about 5 percent of all applications contained alternative names (Lukas Schmidt and Laura Müller, respectively) and profile pictures (photo A and photo B, respectively). Apart from that, the places of residence were altered which allows controlling for the distance between applicants' current address and employers' workplace. On average, this distance was 286 kilometers where the range varied between 0 (residence and workplace are in the same city) and 556 kilometers. The random variation of additional certificates resulted in a fraction of 39.1 percent in which the candidates provided a credential on a school internship. In 273 cases no additional certificates were provided, in 130 cases both applicants attached a credential and in 123 (130) cases only the male (female) candidate handed in a complementary signal.

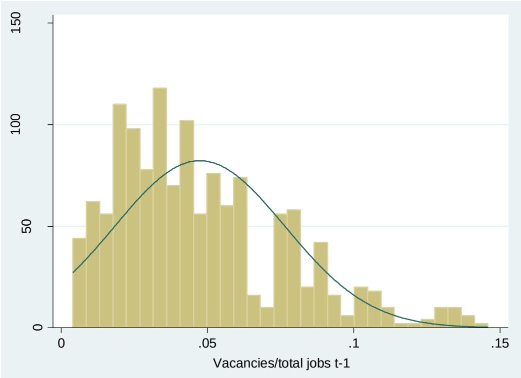
The information collected from companies' responses and their job offers was enriched by labor market data from the BA.[55] Since the workplace of every employer was known, detailed statistics on the regional labor market could be matched with firms. Thus, both the '$H_{scarcity}$' and the '$H_{share\ of\ females}$' hypotheses can be tested. With respect to the former, the variable 'vacancies/total jobs t-1' is constructed by dividing the number of unstaffed apprenticeship positions by the number of registered positions in the previous year. This ratio represents the degree of labor market scarcity employers had to face in the preceding application period and is restricted to the range between 0 and 1. Figure 5-5 shows the frequency distribution of the non-standardized scarcity measure.

Compared to current labor market data, the scarcity measure in t-1 proves to be superior

[55] The data contain information on the number of registered and unstaffed apprenticeship positions as well as on the number of registered and unemployed applicants. Even though registration for both employers and applicants is not obligatory, the BA (2012l) reports a high coverage that is especially dependent on the situation in the job market. If the demand for apprenticeship positions increases relative to supply, applicants are more likely to register, and vice versa.

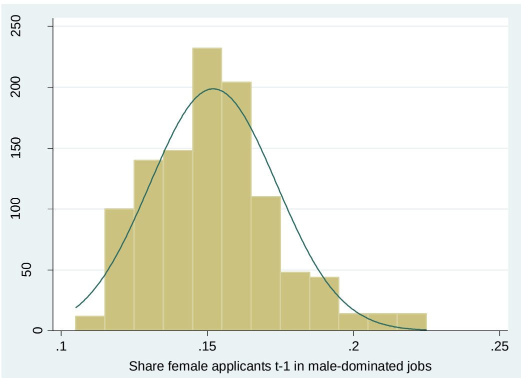
because it takes into account that employers only know ex post whether the quality and quantity of the applications received were sufficient to fill the vacancies. The mean ratio in this sample was 0.047 and ranged from 0.004 to 0.146. On average 4.7 percent of all apprenticeship jobs in the previous year could not be staffed.

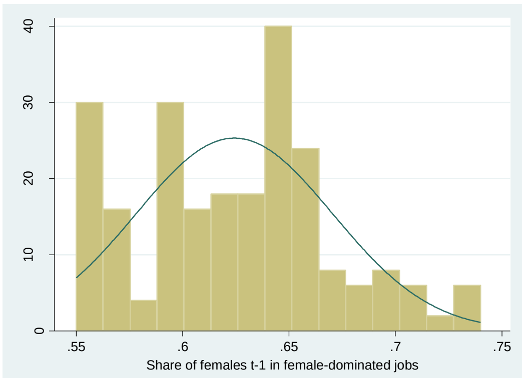
Figure 5-5: Frequency Distribution of Non-Standardized Vacancies/Total Jobs t-1



The share of female applicants in t-1 as another ratio collected from the data of the BA proxies employers' past experience with female applicants. Creating a regional female-total-applicant ratio and matching it with employer data yielded an average of 0.236 in the current sample. However, this ratio varied considerably depending on the nature of the job. While in male-dominated jobs on average 15.2 percent of all applicants were female, the share of female applicants averaged 62.4 percent in female-dominated jobs. Figure 5-6 shows the frequency distribution of the non-standardized 'share of females' measure separated by job type.[56] Table 5-4 provides an overview of the descriptive statistics for the entire dataset.

Figure 5-6: Frequency Distribution of Non-Standardized Share of Females t-1 Separated by Job Type





[56] Note that for the empirical analyses both variables reflecting labor market conditions are standardized.

Table 5-4: Descriptive Statistics of the Correspondence Study on Gender Discrimination

Variable	Operationalization	# of Obs.	Mean	SD	Min	Max
DEPENDENT VARIABLES						
Response	Dummy: Equals 1 if the applicant receives a response (either invitation or rejection) by the employer, 0 otherwise	1312	0.816	-	0	1
Callback	Dummy: Equals 1 if the applicant receives a callback (e.g. invitation) by the employer, 0 otherwise	1312	0.379	-	0	1
INDEPENDENT VARIABLES						
Response information						
Response time	Response time of employers in working days	1070	23.83	27.90	0	178
Type of response						
Email	Dummy: Equals 1 if employer responded by email, 0 otherwise	1070	0.653	-	0	1
Postal mail	Dummy: Equals 1 if employer responded by postal mail, 0 otherwise	1070	0.196	-	0	1
Phone	Dummy: Equals 1 if employer responded by phone, 0 otherwise	1070	0.150	-	0	1
Applicant information						
Female	Dummy: Equals 1 if the applicant is female, 0 otherwise	1312	0.500	-	0	1
Name						
Jan Lange	Dummy: Equals 1 if the applicant is named 'Jan Lange', 0 otherwise	1312	0.447	-	0	1
Lukas Schmidt	Dummy: Equals 1 if the applicant is named 'Lukas Schmidt', 0 otherwise	1312	0.053	-	0	1
Anna Schneider	Dummy: Equals 1 if the applicant is named 'Anna Schneider', 0 otherwise	1312	0.447	-	0	1
Laura Müller	Dummy: Equals 1 if the applicant is named 'Laura Müller', 0 otherwise	1312	0.053	-	0	1
Photo						
Male photo A	Dummy: Equals 1 if the applicant is male and has photo A, 0 otherwise	1312	0.446	-	0	1
Male photo B	Dummy: Equals 1 if the applicant is male and has photo B, 0 otherwise	1312	0.054	-	0	1
Female photo A	Dummy: Equals 1 if the applicant is female and has photo A, 0 otherwise	1312	0.444	-	0	1
Female photo B	Dummy: Equals 1 if the applicant is female and has photo B, 0 otherwise	1312	0.056	-	0	1
Design						
Design A	Dummy: Equals 1 if the application has design A, 0 otherwise	1312	0.370	-	0	1
Design B	Dummy: Equals 1 if the application has design B, 0 otherwise	1312	0.370	-	0	1
Design C	Dummy: Equals 1 if the application has design C, 0 otherwise	1312	0.260	-	0	1
Rank						
Rank 1	Dummy: Equals 1 if the application was sent out first, 0 otherwise	1312	0.500	-	0	1
Rank 2	Dummy: Equals 1 if the application was sent out second, 0 otherwise	1312	0.500	-	0	1
Certificate	Dummy: Equals 1 if the applicant provides an additional certificate, 0 otherwise	1312	0.391	-	0	1
Distance	Linear distance between applicant's home and location of employer (in km)	1312	285.74	123.66	0	556
Information on jobs and application period						
Application period						
May 2011	Dummy: Equals 1 if the application was sent out in May 2011, 0 otherwise	1312	0.375	-	0	1
Sep 2011	Dummy: Equals 1 if the application was sent out in September 2011, 0 otherwise	1312	0.398	-	0	1
May 2012	Dummy: Equals 1 if the application was sent out in May 2012, 0 otherwise	1312	0.227	-	0	1

Job						
Electronics technician	Dummy: Equals 1 if the candidate applies as an electronics technician, 0 otherwise	1312	0.105	-	0	1
Geriatric nurse	Dummy: Equals 1 if the candidate applies as a geriatric nurse, 0 otherwise	1312	0.037	-	0	1
Industrial clerk	Dummy: Equals 1 if the candidate applies as an industrial clerk, 0 otherwise	1312	0.066	-	0	1
Industrial mechanic	Dummy: Equals 1 if the candidate applies as an industrial mechanic, 0 otherwise	1312	0.264	-	0	1
Management assistant for office communication	Dummy: Equals 1 if the candidate applies as a management assistant for office communication, 0 otherwise	1312	0.075	-	0	1
Mechanic in plastics and rubber processing	Dummy: Equals 1 if the candidate applies as a mechanic in plastics and rubber processing, 0 otherwise	1312	0.143	-	0	1
Mechatronics fitter	Dummy: Equals 1 if the candidate applies as a mechatronics fitter, 0 otherwise	1312	0.155	-	0	1
Milling machine operator	Dummy: Equals 1 if the candidate applies as a milling machine operator, 0 otherwise	1312	0.105	-	0	1
Warehouse logistics operator	Dummy: Equals 1 if the candidate applies as a warehouse logistics operator, 0 otherwise	1312	0.050	-	0	1
Female-dominated job	Dummy: Equals 1 if the majority in the respective apprenticeship is female, 0 otherwise (i.e., the majority is male)	1312	0.177	-	0	1
Firm characteristics						
Size						
Small	Dummy: Equals 1 if the employer has less than 50 employees, 0 otherwise	1312	0.332	-	0	1
Medium	Dummy: Equals 1 if the employer has between 50 and 500 employees, 0 otherwise	1312	0.515	-	0	1
Large	Dummy: Equals 1 if the employer has more than 500 employees, 0 otherwise	1312	0.152	-	0	1
Location						
Other	Dummy: Equals 1 if the employer is not located in the South or East of Germany, 0 otherwise	1312	0.297	-	0	1
South	Dummy: Equals 1 if the employer is located in the South of Germany, 0 otherwise	1312	0.527	-	0	1
East	Dummy: Equals 1 if the employer is located in Eastern Germany, 0 otherwise	1312	0.175	-	0	1
Industry	Dummy: Equals 1 if the employer operates in the industry sector, 0 otherwise (i.e., service sector)	1312	0.767	-	0	1
Late recruiter	Dummy: Equals 1 if the employer recruits in May, 0 otherwise (i.e., September)	1312	0.602	-	0	1
Female responsible	Dummy: Equals 1 if the person responsible for recruiting as mentioned in the job offer is female, 0 otherwise	1312	0.532	-	0	1
Open positions	Number of open positions for an apprenticeship as indicated by the employer's job offer	1312	1.68	1.59	1	15
Labor market data						
Vacancies/total jobs t-1	Ratio of vacancies and total apprenticeships in the previous year (i.e., in the reporting period 2009/2010 and 2010/2011, respectively) and in the corresponding employment agency region of the employer	1312	0.047	0.029	0.004	0.146
Share of females t-1	Share of female applicants in the previous year (i.e., in the reporting period 2009/2010 and 2010/2011, respectively) and in the corresponding employment agency region of the employer	1312	0.236	0.182	0.110	0.740

5.2.1.2 COMPARISON WITH THE OVERALL POPULATION OF TRAINING COMPANIES

A comparison of firm characteristics in the present sample and the overall population of employers having registered their apprenticeship position at the BA in 2010/2011 is displayed in table 5-5. The figures reveal that small firms are underrepresented while medium-sized firms make up a higher share in the field experiment than in the actual population of training companies. A possible explanation is that the majority of small firms still rely on postal applications because they are less likely to use the Internet and have a relatively low number of incoming applications which keeps the administrative requirements for the hiring procedures within a reasonable range.

Table 5-5: Firm Characteristics in Field Experiment and Entire Population of Training Companies

	Field experiment	Entire population of training companies
Size		
Small	33.23%	45.97%
Medium	51.52%	36.39%
Large	15.24%	17.64%
Location		
South	52.74%	45.32%
East	17.53%	17.60%
Other	29.73%	37.02%

Notes: Data on firm size as of 2010. Data on location as a weighted average of 2010/2011 and 2011/2012.
Source: BA (2010a, 2011, 2012b), BIBB (2010a).

Apart from differences in firm size, employers from the South are slightly overrepresented in the present sample whereas those located in the northern and western states make up a lower share compared to the entire population. This might be due to the fact that particularly in the South of Germany where labor market competition for talent is particularly fierce, firms offer their vacancies via various channels and for a longer period of time which in turn increases the probability of appearing in the current sample. Whether or not the representativeness of the dataset influences the outcome on gender discrimination will be discussed in section 5.2.4.

5.2.2 DESCRIPTIVE RESULTS

According to Heckman and Siegelman (1993: 198), not any differential treatment on firm level can be regarded as discrimination, but “discrimination exists whenever two testers in a matched pair are treated differently in the aggregate or on average.” The results of the field experiment on apprenticeship applications suggest that these average differences exist.

Table 5-6: Firms' Detailed Responses by Gender

	Male (N=656)	Female (N=656)	Total (N=1,312)	Difference
No response	19.51%	17.38%	18.45%	2.13 pps
	(128)	(114)	(242)	(14)
Rejection	40.24%	47.10%	43.67%	-6.86 pps**
	(264)	(309)	(573)	(45)
Callback	40.24%	35.52%	37.88%	4.72 pps*
	(264)	(233)	(497)	(31)

Notes: The table reports detailed responses by gender as a fraction of overall applications in percent. Absolute numbers are in parentheses. * denotes 10% significance level and ** denotes 5% significance level of a chi-squared test (H_0: The male and female candidates are equally likely to receive a callback/a rejection at any matched-pair application).

Table 5-6 shows a detailed overview of employers' responses by gender for the whole dataset. Overall, 497 applications resulted in a callback by employers. Comparing callbacks by gender shows that the male candidate was invited 264 times (40.24 percent) whereas the female candidate received 233 positive responses (35.52 percent). Moreover, the male (female) applicant was rejected in 264 (309) cases while, accordingly, 128 (114) applications remained unanswered. Due to the nature of the correspondence method, these results indicate that the male candidate has a 4.72 percentage points higher probability of being called back than the female applicant. Conducting a chi-squared test shows that these gender differences in callbacks are statistically significant at the 10 percent level. It thus seems that hiring discrimination by gender exists.

Table 5-7: Firms' Callbacks Conditional on Job Type

	Male	Female	Difference
Male-dominated	40.93%	34.44%	6.49 pps**
	(221/540)	(186/540)	
Female-dominated	37.07%	40.52%	-3.45 pps
	(43/116)	(47/116)	

Notes: The table reports callbacks by gender as a fraction of applications in male- and female-dominated jobs, respectively, in percent. Absolute numbers of callbacks and applications are in parentheses. ** denotes 5% significance level of a chi-squared test (H_0: The male and female candidates are equally likely to receive a callback at any matched-pair application).

Looking more closely at where the differences in callbacks might stem from reveals that job type seems to be a moderator. Although female-dominated jobs were only considered in a rather small subsample, it becomes obvious that the lower callback rate of the female applicant is limited to male-dominated jobs. Table 5-7 highlights that the male candidate has a 6.49 percentage points higher probability of being invited. This difference is statistically significant at the 5 percent level. With respect to female-dominated jobs, however, the female applicant's disadvantage disappears.

Table 5-8: Firms' Callbacks Conditional on the Provision of an Additional Certificate

	Male	Female	Difference
No certificate	37.47%	33.84%	3.63 pps
	(151/403)	(134/396)	
Certificate	44.66%	38.08%	6.58 pps
	(113/253)	(99/260)	
Difference	7.19 pps*	4.24 pps	

Notes: The table reports callbacks by gender as a fraction of applications with and without an additional certificate in percent. Absolute numbers of callbacks and applications are in parentheses. * denotes 10% significance level of a chi-squared test (H_0: Applications with and without an additional certificate are equally likely to receive a callback).

Furthermore, the inclusion of a certified school internship seems to influence the candidates' callback rates (see table 5-8). If a credential is attached, the share of invitations to both the male and the female applicant increases. While the male candidate benefits by 7.19 percentage points, his female counterpart only realizes a 4.24 percentage points increase in positive responses with only the former difference being statistically significant at conventional levels.

Table 5-9: Firms' Callbacks Conditional on Application Period

	Male	Female	Difference
Late recruiters	38.99%	33.16%	5.83 pps*
	(154/395)	(131/395)	
Early recruiters	42.15%	39.08%	3.07 pps
	(110/261)	(102/261)	

Notes: The table reports callbacks by gender as a fraction of applications to late and early recruiters in percent. Absolute numbers of callbacks and applications are in parentheses. * denotes 10% significance level of a chi-squared test (H_0: The male and female candidates are equally likely to receive a callback at any matched-pair application).

With regard to the different application periods, it becomes obvious that differential treatment is somewhat higher if the sample is restricted to late recruiters (see table 5-9). A chi-squared test of equal callback distributions across gender indicates that the difference of 5.83 percentage points is statistically significant at the 10 percent level. In contrast, the callback rates for the male and female candidate do not significantly differ for applications dispatched to early recruiters.

Focusing on differential treatment at firm level, four scenarios can be observed, i.e., (i) mutual rejection or no response, (ii) invitations to both of the candidates or a callback to either the (iii) majority or (iv) minority group member. Table 5-10 compares employers' responses between the male and the female applicant conditional on job type (male- versus female-dominated), the provision of a certified internship, firm characteristics and labor market scarcity (split at its mean). Column (1) displays the number of employers referred to in each stratum. Columns (2) and (3) distinguish between employers that did not respond to or rejected both candidates and employers that invited at least one of them. Columns (4)–(6) separate the observations of column (3) into those cases where both candidates received a positive response (4) and those where either the male (5) or the female candidate (6) was favored. The callback rates for both the male and female applicant are presented in columns (7) and (8). Deducting column (8) from column (7) finally yields the difference in overall callback rates (9).[57]

Table 5-10: Firms' Responses of Correspondence Testing by Gender, Job Type, Certificate, Firm Characteristics and Labor Market Data

	Firms' responses					Callback rates		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
No. of paired applications	Rejection/ no response	At least one callback	Both	Only male	Only female	Male (4+5)/(1)	Female (4+6)/(1)	Difference (7)-(8)
All firms (656)	52.29 (343)	47.71 (313)	58.79 (184)	25.56 (80)	15.65 (49)	0.402	0.355	0.047* (p=0.078)
Job type								
Male-dominated job (540)	52.78 (285)	47.22 (255)	59.61 (152)	27.06 (69)	13.33 (34)	0.409	0.344	0.065** (p=0.028)
Female-dominated job (116)	50.00 (58)	50.00 (58)	55.17 (32)	18.97 (11)	25.86 (15)	0.371	0.405	-0.034 (p=0.590)
Additional certificate								
None provides additional certificate (273)	54.58 (149)	45.42 (124)	54.84 (68)	30.65 (38)	14.52 (18)	0.388	0.315	0.073* (p=0.073)
Both provide additional certificate (130)	48.46 (63)	51.54 (67)	64.18 (43)	28.36 (19)	7.46 (5)	0.477	0.369	0.108* (p=0.079)

57 As already discussed in section 3.1.2.2, some of the literature relies on the restricted sample (where mutual rejections and cases of no response, i.e., all observations as of column (2), are considered as non-observations) because it inter alia drops those job offers where the position has already been filled and no assessment on the candidates' applications has taken place. If there was a substantial number of these cases, regression results would probably underestimate the extent of discrimination, if any. In order to overcome any potential bias some researchers take into account both the full and the restricted sample. In subsequent econometric analyses, including the restricted sample always increases the magnitude of the coefficients and their significance level, but does not provide further insights on gender discrimination. Also, excluding the cases where employers note that the position has already been filled does not change much in the results. In fact, taking into account the full sample for the calculation of any gender effects is the more conservative way (for a thorough discussion, see Riach and Rich (2002)). Results using the restricted sample only are available from the author upon request.

Only male provides additional certificate (123)	51.22 (63)	48.78 (60)	65.00 (39)	20.00 (12)	15.00 (9)	0.415	0.390	0.024 (p=0.697)
Only female provides additional certificate (130)	52.31 (68)	47.69 (62)	54.84 (34)	17.74 (11)	27.42 (17)	0.346	0.392	-0.046 (p=0.441)
Timing								
Late recruiter (395)	53.16 (210)	46.84 (185)	54.05 (100)	29.19 (54)	16.76 (31)	0.390	0.332	0.058* (p=0.088)
Early recruiter (261)	50.96 (133)	49.04 (128)	65.63 (84)	20.31 (26)	14.06 (18)	0.421	0.391	0.031 (p=0.476)
Firm Size								
Small (<50) (218)	57.80 (126)	42.20 (92)	50.00 (46)	27.17 (25)	22.83 (21)	0.326	0.307	0.018 (p=0.680)
Medium (50-500) (338)	49.11 (166)	50.89 (172)	61.05 (105)	26.74 (46)	12.21 (21)	0.447	0.373	0.074* (p=0.051)
Large (>500) (100)	51.00 (51)	49.00 (49)	67.35 (33)	18.37 (9)	14.29 (7)	0.420	0.400	0.020 (p=0.774)
Location								
South (346)	56.07 (194)	43.93 (152)	59.21 (90)	25.66 (39)	15.13 (23)	0.373	0.327	0.046 (p=0.202)
East (115)	47.83 (55)	52.17 (60)	66.67 (40)	18.33 (11)	15.00 (9)	0.443	0.426	0.017 (p=0.790)
Other (195)	48.21 (94)	51.79 (101)	53.47 (54)	29.70 (30)	16.83 (17)	0.431	0.364	0.067 (p=0.179)
Sector								
Services (153)	46.41 (71)	53.59 (82)	54.88 (45)	26.83 (22)	18.29 (15)	0.438	0.392	0.046 (p=0.417)
Industry (503)	54.08 (272)	45.92 (231)	60.17 (139)	25.11 (58)	14.72 (34)	0.392	0.344	0.048 (p=0.117)
Person responsible for recruiting								
Male (299)	54.18 (162)	45.82 (137)	56.20 (77)	26.28 (36)	17.52 (24)	0.378	0.338	0.040 (p=0.306)
Female (340)	50.59 (172)	49.41 (168)	62.50 (105)	24.40 (41)	13.10 (22)	0.429	0.374	0.056 (p=0.137)
Vacancies/total jobs t-1 (Mean=0.047)								
Above mean (272)	56.25 (153)	43.75 (119)	59.66 (71)	18.49 (22)	21.85 (26)	0.342	0.357	-0.015 (p=0.718)
Below mean (384)	49.48 (190)	50.52 (194)	58.25 (113)	29.90 (58)	11.86 (23)	0.445	0.354	0.091*** (p=0.010)

Notes: This table shows the distribution of firms' responses. Absolute numbers are in parentheses. Column (1) displays the number of employers in each stratum. Column (2) reports the fraction of firms that gave none of the candidates a callback, so the remainder in column (3) called back at least one applicant. Firms that gave both candidates a positive answer, column (4), are considered as equal treatment, while the rest preferred either the male or the female candidate (columns (5) and (6)). Columns (7) and (8) contain the callback rate for the male and female applicant, respectively, while column (9) computes the difference in callback rates between the two candidate groups. 'Person responsible for recruiting' excludes those employers that did not name a recruiter in their job offers. In column (9), p-values of a chi-squared test that the male and female candidates are equally likely to receive a callback at any matched-pair application are in parentheses. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

Table 5-10 shows that approximately 48 percent (313 of 656) of the firms invited at least one candidate. While both candidates were invited by 184 employers, there was differential treatment in 129 companies. Among these observations, the female applicant was favored in 15.65 percent (49) whereas her male counterpart was invited in 25.56 percent (80) of the cases. While the application of the male candidate was successful in 40.2 percent, the overall callback rate for the female applicant was 35.5 percent only. This yields a difference of 4.7 percentage points which is statistically significant at the 10 percent level. Put differently, men are 13 percent (=0.402/0.355) more likely to receive a callback than their female counterparts.[58]

Differential treatment turns out to be most prominent in male-dominated jobs where the callback differences add up to 6.5 percentage points and is statistically significant at the 5 percent level.[59] Focusing on the provision of a certified internship shows that discrimination remains statistically significant only when either none or both of the candidates provide an extra credential. If either of the candidates has done an internship, differential treatment fully disappears. This is particularly surprising if only the male candidate provides a certificate. Here, the differences in callback rates would have been expected to be even larger. In contrast, the reverse (though non-significant) gap in callback differentials indicates that the female candidate seems to benefit if only she offers a certified internship. A detailed discussion on the role of certificates will be postponed to the next section.

Referring to firm characteristics, descriptive statistics reveal that differential treatment is particularly influenced by the timing of employers. While gender discrimination does not exist in case of early recruiters, companies that staff their positions rather late seem to discriminate the female candidate who was 17 percent less likely to receive an invitation to a job interview. Apart from that, discrimination is statistically significant at the 10 percent level only for medium-sized companies.

Callback differentials also vary if the sample is divided at the mean of the ‘vacancies/total jobs’ ratio. Whenever labor market scarcity is above the mean, gender discrimination seems to disappear. On the other hand, if the situation on the job market from an employer’s perspective is rather relaxed, the female candidate is 26 percent less likely (on

---

[58] In terms of the aforementioned net discrimination rate, i.e., the fraction of callbacks to the male applicant minus the fraction of callbacks to the female candidate as a share of overall callbacks to at least one of the applicants, the callback difference is 9.90% ($\frac{80-49}{313} \cdot 100$).

[59] Pairwise comparisons of callbacks separate for male- and female-dominated jobs can be requested from the author.

a 1 percent significance level) to be called back.[60]

Overall, descriptive results at group and firm level suggest that gender discrimination is affected by the job type, the provision of additional productivity signals, the application period and regional labor market scarcity. In order to assess any confounding effects and to test the aforementioned hypotheses on the sources of differential treatment, that is statistical and taste-based discrimination, econometric analyses are required.

Before that, however, more indirect ways of differential treatment are discussed. In fact, employers might process the applications differently conditional on group membership resulting in, for example, more cases of no response and longer callback or rejection times for the applicants of one group as opposed to the candidates of the other. Such behavior describes what Fibbi et al. (2006) call "equal but different treatment". Informing one candidate on his/her rejection and simultaneously not responding to the other one would be a first means of discrimination. Even though the actual hiring outcome could eventually be the same, i.e., both would turn up in column (2) of table 5-10, a case of no reply might further discourage the candidates and make them hope for a positive answer where in fact they will not receive any at all. The results of the present study, however, do not point at any gender differences with respect to the no response rate. Both candidates face statistically the same proportion of firms' responses, i.e., number of cases in which the companies either rejected or invited the applicants (see table 5-11). Applications of the male candidate remained unanswered slightly more often than those of his female counterpart. This seems quite odd in view of the fact that he was able to realize significantly more callbacks. However, the difference is insignificant so that further considerations of firms' response behavior as a source for gender differences can be neglected.[61]

60 Note that a comparison of callbacks separated by the share of female applicants (with a threshold at the mean) produces identical results as the division by job types (and is therefore not reported). This, of course, is somewhat plausible by definition as male-dominated (female-dominated) jobs have a relatively low (high) share of female applicants.

61 Additional multivariate regressions investigating firms' response behavior indicate that the probability of receiving a response is independent of gender (see table C-2 in the appendix).

Table 5-11: Firms' Responses by Gender

	Male	Female	Total	Difference
No response	19.51%	17.38%	18.45%	2.13 pps
	(128)	(114)	(242)	(14)
Response	80.49%	82.62%	81.55%	-2.13 pps
	(528)	(542)	(1070)	(14)

Notes: The table reports employers' responses by gender as a fraction of overall applications in percent. Absolute numbers are in parentheses.

In the same vein, equal but different treatment may occur within a positive scenario. Whenever an applicant is invited only after his/her counterpart has declined an invitation, it seems that he/she is the employer's second best option.[62] Table 5-12 considers all cases of mutual callbacks and shows that in respectively 14 and 19 percent of all callbacks, applicants are informed only after rejection on behalf of the matched counterpart. Again, it was rather the female than the male candidate who was slightly favored. In 35 (26) cases, the male (female) applicant received a callback after the counterpart declined the firm's interest. Nevertheless, the differences are not statistically signficant.

Table 5-12: Firms' Callbacks only after the Counterpart Has Declined an Invitation

Callbacks…	Fraction (Absolute number)
… to both candidates	100.00% (184)
… to the male candidate only after the female candidate has declined an invitation	19.02% (35)
… to the female candidate only after the male candidate has declined an invitation	14.13% (26)

Notes: The table reports cases of equal but different treatment by gender as a fraction of mutual callbacks. Absolute numbers are in parentheses.

Even though there are no systematic gender differences with respect to cases of 'second best options' as described above, the likelihood that a candidate voluntarily resigns increases with more time elapsing until the callback or rejection is announced. Thus, systematic differences with respect to average callback and rejection times, respectively, might be an additional indicator for differential treatment by gender. Table 5-13 displays the callback and rejection times, respectively, by gender and firm size. On average, firms

62 Duguet et al. (2012) show both theoretically and empirically that accounting for the response order allows for a more detailed understanding of whether discrimination can be considered as "weak" or "strong".

invite (reject) the candidates after 17.5 (29.4) working days. While no significant differences for the male and female applicants are revealed, there is variation across firms. Small companies react faster than medium-sized and large employers. This finding is not surprising since the latter on average have more apprenticeship positions to staff and in turn probably face a higher number of incoming applications that have to be administered. Moreover, decision processes tend to last longer as they involve more decision makers.

Table 5-13: Average Callback and Rejection Times in Working Days by Gender

	Callback			Rejection		
	Male	Female	Average	Male	Female	Average
All	17.6	17.3	17.5	29.5	29.2	29.4
Small	14.4	14.8	14.6	23.2	22.6	22.9
Medium	18.4	18.6	18.5	30.9	30.9	30.9
Large	20.5	17.4	18.9	37.0	36.5	36.7

5.2.3 ECONOMETRIC ANALYSES

In this section the estimation technique used for the empirical analyses is described (5.2.3.1), an empirical model is derived (5.2.3.2) and probit regressions are estimated to test the hypotheses developed in section 4.3 (5.2.3.3).

5.2.3.1 ESTIMATION TECHNIQUE

In the field experiments on both gender and ethnic discrimination, differential treatment occurs whenever the male (German-named) or the female (Turkish-named) applicant on average receives fewer callbacks from firms. The firm's callback is a binary outcome variable that equals 1 if the applicant receives a callback and is 0 otherwise.

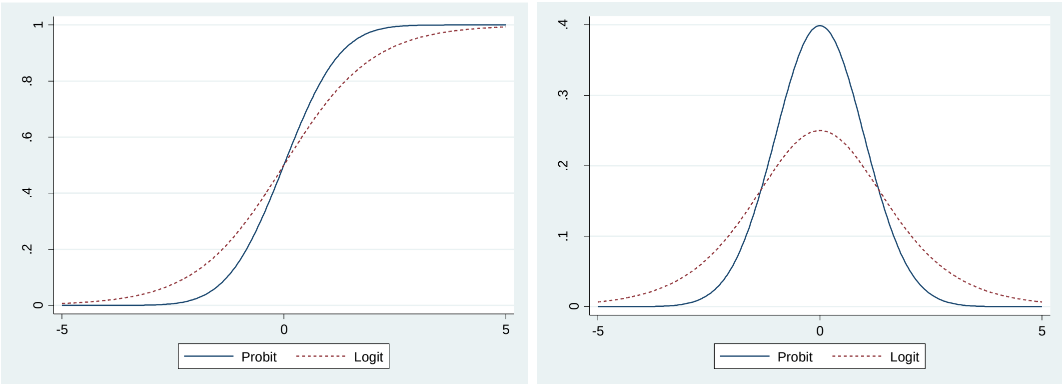
Analyzing binary outcome variables requires a modification of the classical linear regression technique that pays attention to the fact that for an observation i only two outcomes exist, i.e., an event (such as a callback) can either occur ($Y_i = 1$) or not occur ($Y_i = 0$). As for estimations with a continuous dependent variable, the probability P ($Y_i = 1$) can be modeled as a linear combination of X_k independent variables. Thus,

$$E\left(Y_i \middle| X_{i,k}\right) = P_i = \beta_0 + \sum_i^I \beta_k * X_{k,i} + \mu_i,$$

where β_0 represents the intercept with the y-axis, β_k denotes regression coefficient β of independent variable k and μ_i is a random error term with $E(\mu_i) = 0$. Due to its functional form, this relationship is also referred to as the linear probability model (LPM). The LPM allows P_i to take values between $-\infty$ and $+\infty$. However, the probability of an event to occur by definition needs to fall in the range between 0 and 1 for all values of the parameters β_k and the X_k. Moreover, the probabilities $P(Y_i = 0)$ and $P(Y_i = 1)$ have to add up to 1 which does clearly not hold for the LPM. In other words, a linear relationship between a dependent dummy variable and a set of independent variables like in the LPM violates crucial probability assumptions. As a consequence, a nonlinear functional form is required that satisfies these assumptions and thus enables the researcher to draw plausible inferences on the probability P_i. Here, econometricians rely on either the logistic or the standard normal cumulative distribution function (cdf). The former are referred to as logit and the latter as probit models. Both are superior to the LPM since they produce probability outcomes that are in accordance with the assumptions mentioned above (Gujarati and Porter, 2009).

Probit and logit regressions yield similar results since calculations of marginal effects and discrete changes are conducted analogously. In fact, the major difference is the underlying distribution which leads to slightly different solutions at the tails (see figure 5-7).

Figure 5-7: Cumulative Distribution and Density Functions of Probit and Logit Models



Probit and logit coefficients are not directly comparable. The reason is that the standard normal and logistic distributions have the same mean value of zero, but different variances. While the former has a variance of 1, the variance of the latter is $\frac{\pi^2}{3} = 1.814$. Thus, multiplying the coefficients from a probit regression with 1.814 results in the logit coefficients. However, both models lead to identical conclusions and may therefore be used interchangeably (Liao, 1994). In this dissertation only probit models are estimated. Logistic regression results are available from the author upon request.

5.2.3.1.1 FORMAL DERIVATION OF THE PROBIT MODEL

As mentioned above, in probit models $P_i = F_N(Z_i)$, where F_N represents the standard normal cdf $F_N(Z_i) = \int_{-\infty}^{Z_i} f_N(z)\,dz$ with standard normal density $f_N(Z_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{Z_i^2}{2}}$ and Z_i is an unknown (latent) variable that denotes a utility index of observation i. This utility index, which goes back to the rational choice theory developed by McFadden (1974), is determined by a linear combination of the independent variables X_k and a stochastic term μ_i that is a normally distributed random variable (as opposed to the logistic regression where the error term μ_i is a standard logistic random variable). Hence, Z_i is calculated as follows:

$$Z_i = \beta_0 + \sum_i^I \beta_k * X_{k,i} + \mu_i.$$

It is further assumed that if Z_i exceeds a critical or threshold level Z_i^*, $Y_i = 1$ will occur. Accordingly,

$$Y_i = \begin{cases} 1, \text{if } Z_i \geq Z_i^* \text{ and} \\ 0, \text{otherwise.} \end{cases}$$

Thus,

$$P_i = P(Y_i = 1 | X_{k,i}) = P(Z_i \geq Z_i^*).$$

Rearranging this equation given the normality assumption yields:

$$P_i = F_N(\beta_0 + \sum_i^I \beta_k * X_{k,i}) = F_N(Z_i).$$ [63]

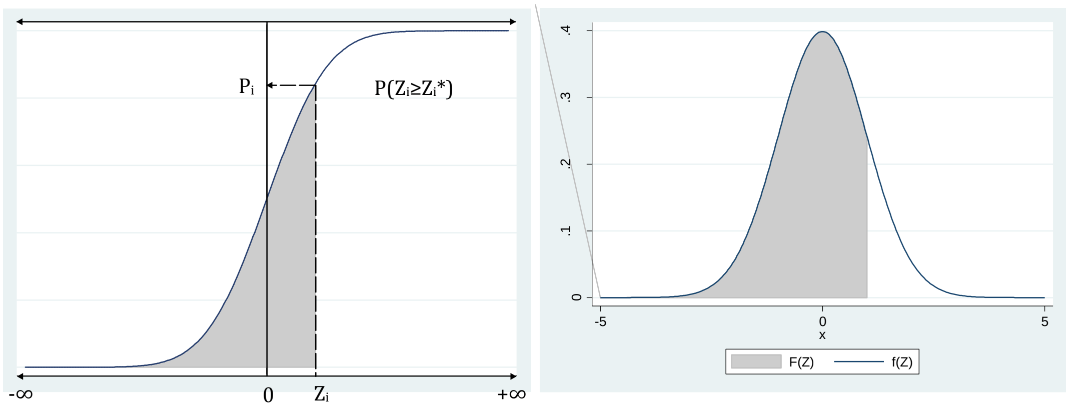
Hence, the probability $P(Z_i \geq Z_i^*)$ can be computed from the standard normal cdf $F_N(Z_i)$. Put in illustrative terms, the probability P_i is represented by the area under the standard normal cdf F_N that lies between $-\infty$ and Z_i and the area under the density curve f_N, respectively, and is thus increasing in Z_i (see figure 5-8) (Gujarati and Porter, 2009; Wooldridge, 2009).

Like previous derivations show, the latent variable Z_i connects the linear combination of independent variables with the normal cdf and therefore serves as a 'linking function'. In line with the name of the regression technique, Z_i is called a 'probit'. Since $P(Y_i = 1 | X_{k,i})$ violates the linearity assumption required for the use of Ordinary Least Squares (OLS), the parameters in probit (as well as in logit) regressions are estimated by the Maximum-Likelihood (ML) method which produces the most consistent and efficient

[63] Note that μ_i can be disregarded due to the normality assumption and its independence of X_k.

estimators.[64]

Figure 5-8: Illustration of the Probability P_i below the Normal Cumulative Distribution and Density Function



5.2.3.1.2 Probit Coefficients and Marginal Effects

In binary regression models, the primary goal is to identify and explain the effects of a set of independent variables X_k on the outcome probability $P\ (Y_i = 1)$. In the present context, particularly the effect of gender and any confounding factors on the callback probability of the applicants is evaluated. Due to the nonlinear nature of the standard normal cdf, the probit coefficients β_k only allow for drawing inferences on the direction and level of significance of an independent variable X_k on the probability P_i, but do not permit a plausible interpretation with respect to their magnitude. Furthermore, probit coefficients cannot be compared within and across estimation models as long as the empirical units and the set of regressors vary. For this reason, the partial effect of X_k on the response probability has to be derived. If the independent variable is continuous, the marginal effect, i.e., the effect of an infinitesimal change in X_k, is obtained as follows:

$$\Delta P_i = \frac{\partial[P(Y_i=1)]}{\partial[X_{i,k}]} = \beta_k * f_N(Z_i).$$

Given that F_N is a strictly increasing cdf, $f_N(Z_i) > 0$ (see figure 5-8) and thus ΔP_i always has the same sign as β_k. Unlike in linear regressions, the marginal effect of X_k differs depending on $f_N(Z_i)$, i.e., all other values of X_k and their parameters β_k. The largest effect occurs if $Z_i = 0$. Hence, $f_N(0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{0_i{}^2}{2}} \approx 0.4$ as illustrated in figure 5-8. According to the standard normal cdf, this results in a predicted probability $\widehat{P}\ (Y_i = 1)$ of 0.5. Consequently,

[64] For a discussion of the assumptions and the procedure of the ML method, see for example Aldrich and Nelson (1984).

any $Z_i \gtrless 0$ produces smaller (absolute) marginal effects compared to $Z_i = 0$. In fact, the marginal effects decrease if Z_i approaches $\pm\infty$ where $F_N(Z_i)$ approaches 0 and 1, respectively. For ease of interpretation, researchers calculate the partial effect at the average of all other explanatory variables by plugging in their means in Z_i. In case of categorical independent variables, however, the mean is often replaced by the mode as this makes interpretations less tedious. The partial effect of a categorical independent variable, e.g., the effect of being a woman (here: $X_G = 1$) versus being a man ($X_G = 0$) on the outcome probability, is ceteris paribus calculated as a discrete change: $\Delta P_i = F_N(Z_{X_G=1}) - F_N(Z_{X_G=0})$ (Gujarati and Porter, 2009; Wooldridge, 2009).

Moreover, the intuition of linear regression models also needs to be adapted for probit estimations if interaction terms are included. Ai and Norton (2003) show that the full interaction effect is not just the marginal effect of the interaction between two independent variables, but the cross-partial derivative of the predicted probabilities $\widehat{P}\,(Y_i = 1)$. This implies that (i) the interaction effect could be nonzero even if the average marginal effect is equal to zero, (ii) the significance level of the interaction effect varies depending on the predicted probabilities and (iii) the magnitude and direction of the interaction effect are conditional on the values of other covariates.

5.2.3.1.3 GOODNESS OF FIT MEASURES

Apart from the estimation technique and interpretation of the coefficients, the goodness of fit (GoF) measures in probit models also differ from those in linear regression models. The most prominent ones used for model comparisons are presented below (Aldrich and Nelson, 1984; Wooldridge, 2009; Backhaus et al., 2011).

Likelihood-ratio (LR) test: This measure tests the hypothesis that all coefficients except for the intercept are zero and is calculated as:

$$\text{LR test} = -2(LL_0 - LL_f),$$

where LL_0 is the log-likelihood of the null (intercept) model and LL_f is the log-likelihood of the fitted model. The computed LR chi-squared is compared with the critical value of the chi-squared distribution at significance level α with $k - 1$ degrees of freedom. Referring to linear regression models, the LR test is comparable to the overall F statistic.[65]

Pseudo R^2: Apart from the LR test, various pseudo R^2 measures that are somewhat related to each other can be calculated. For convenience, only McFaddens-R^2 is reported in the analysis. The rationale is similar to the coefficient of determination in OLS estimations. If the fit diminishes, the pseudo R^2 approaches 0 and if the fit improves, it approaches 1. McFaddens-R^2 is probably the most frequently used GoF measure for models with categorical dependent variables such as probit and logit models. Similar to the LR test, it computes the log-likelihood of the fitted and null (intercept) model and relates them to each other:

$$R_{\mathrm{McFadden}}{}^2 = 1 - \left(\frac{LL_f}{LL_0}\right).$$

Thus, if the estimated model has no explanatory power, it follows that the ratio $\left(\frac{LL_f}{LL_0}\right) = 1$ and the Pseudo $R_{\mathrm{McFadden}}{}^2 = 0$. In contrast to the LR test which indicates the overall significance of the estimation model, McFadden's R^2 is a measure that maps the estimation quality of the independent variables employed in the model and thus enables the researcher to compare the fit of different regression models. In contrast to linear regression models, however, the pseudo R^2 measure is usually fairly low. In fact, values of $0.2 \leq$ Pseudo $R^2 \leq 0.4$ can already be considered as a reasonable model fit (Urban, 1993).

5.2.3.2 EMPIRICAL MODEL

In the subsequent regressions, the response and callback dummy is modeled as a set of independent variables that include a dummy for gender, a vector of various firm characteristics, variables reflecting the situation on the regional labor market, a dummy that accounts for the provision of an additional certificate, a dummy for the type of job and a set of control variables. Since the empirical model puts its emphasis on the effect of gender on the callback probability $P_i\ (Y = 1)$, where $Y_i = 1$ if the candidate receives a callback, the regression model needs to be based on a probabilistic distribution. Here, probit regression analysis is used which follows the standard normal cdf.

Next, the full empirical model is presented. However, the empirical estimations include

65 Note that if standard errors are clustered (as will be the case in subsequent analyses (see footnote 33)) a Wald test rather than a LR test is performed. The Wald test and LR test, however, are shown to be asymptotically equivalent and usually yield similar conclusions (Engle, 2007). For a formal description of the Wald test, see Wooldridge (2010).

various model specifications as sensitivity checks and to document the robustness of the results. In particular, interaction effects that should test the aforementioned hypotheses on the factors influencing differential treatment, if any, are incorporated in the regression models. Overall,

$$P_i(\text{Callback}) = \beta_0 + \beta_1 \text{Female}_i + \beta_2 \overrightarrow{\text{Firm characteristics}_i} + \beta_3 \text{ Share of females t} - 1_i + \beta_4 \text{ Vacancies/total jobs t} - 1_i + \beta_5 \text{ Certificate}_i + \beta_6 \text{ Female} - \text{dominated job}_i + \beta_7 \overrightarrow{\text{Controls}_i} + \mu_i,$$

where β_0 is a constant, β_k denotes the regression coefficient β of regressor k, μ_i depicts a normally distributed error term of applicant i and the independent variables are as described in table 5-4. The vector of firm characteristics includes information on firm size, location, industry, whether the employer is a late recruiter and a dummy for the sex of the recruiter. The variables proxying the labor market situation, i.e., the share of females in t-1 and vacancies/total jobs in t-1, are standardized so that $\mu(X_k) = 0$ and $\sigma(X_k) = 1$. Further controls include a dummy for the apprenticeship year, the number of open positions, the distance to the workplace, as well as dummies for the dispatching order and the template (design) of the application.

5.2.3.3 PROBIT REGRESSIONS AND HYPOTHESES TESTING

First, the empirical analysis investigates the relationship between job type and callback probability by gender. Therefore, the data from the three application periods are pooled which results in an overall sample of 1,312 observations. Table 5-14 reports average marginal effects on the probability of receiving a callback. Model (I) only includes the female dummy, model (II) additionally includes firm characteristics, model (III) adds standardized labor market variables, model (IV) also incorporates a dummy for the job type and model (V) further controls for an interaction term that equals one if the female candidate applies for a female-dominated job. All models account for potential joint effects originating from the control variables. Additional photo and name effects have been tested but appeared insignificant as demonstrated by tables C-3 and C-4 in the appendix. They are thus excluded from further regression analyses.

Table 5-14: Marginal Effects from Probit Regressions on Callback Dummy and Test of Job Type Hypothesis

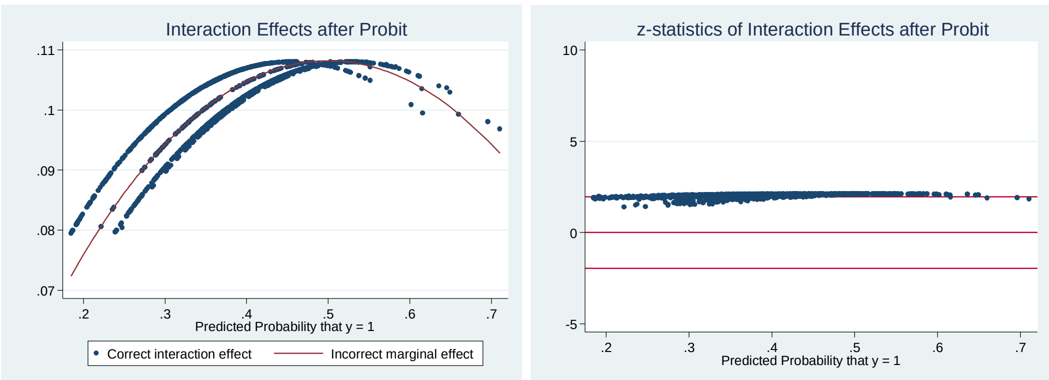
Callback	**(I)**	**(II)**	**(III)**	**(IV)**	**(V)**
Female	-0.050***	-0.051***	-0.051***	-0.051***	-0.070***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.019)
Medium		0.108***	0.108***	0.107***	0.107***
		(0.041)	(0.041)	(0.041)	(0.041)
Large		0.079	0.079	0.077	0.077
		(0.064)	(0.064)	(0.064)	(0.064)
South		-0.052	-0.043	-0.043	-0.040
		(0.054)	(0.057)	(0.057)	(0.057)
East		0.059	0.065	0.066	0.066
		(0.055)	(0.055)	(0.057)	(0.058)
Industry		-0.067	-0.068	-0.069	-0.069
		(0.053)	(0.053)	(0.053)	(0.053)
Late recruiter		-0.013	-0.001	-0.001	-0.001
		(0.058)	(0.083)	(0.084)	(0.084)
Female responsible		0.018	0.018	0.018	0.018
		(0.035)	(0.035)	(0.035)	(0.035)
Share of females t-1			-0.004	-0.015	-0.016
			(0.031)	(0.117)	(0.117)
Vacancies/total jobs t-1			-0.011	-0.011	-0.011
			(0.021)	(0.021)	(0.021)
Certificate				0.026	0.024
				(0.032)	(0.032)
Female-dominated job				0.032	-0.018
				(0.315)	(0.309)
Female x					0.105**
Female-dominated job					(0.051)
Controls	Yes	Yes	Yes	Yes	Yes
No. of obs.	1,312	1,312	1,312	1,312	1,312
Pseudo R^2	0.010	0.021	0.021	0.021	0.022
Log likelihood	-861.957	-852.607	-852.331	-852.064	-851.026
Wald chi-squared	17.315	29.007	29.341	30.279	35.429
P-value	0.015	0.010	0.022	0.035	0.012

Notes: Each model reports average marginal effects of a probit regression on the callback dummy (Y=1: employer calls back the job applicant). Marginal effects are calculated at the mean of all independent variables and denote an infinitesimal change in case of continuous variables and a discrete change in case of dummy variables. Standard errors clustered on firm level are in parentheses. Regressions consider the entire sample. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

The results indicate that the female applicant has a 5 percentage points lower callback probability than the male candidate. This effect is robust and statistically significant at the 1 percent level for the models (I) to (IV). Model (V) reveals slightly different results. In line with the '$H_{job\ type}$' hypothesis, the interaction term indicates that the likelihood of an invitation significantly increases by 10.5 percentage points if the female applicant addresses female-dominated jobs. As a consequence, the magnitude of the coefficient of the female dummy (denoting female's callback probability in male-dominated jobs) increases (-0.070). The inclusion of the interaction term further allows for drawing inferences on how the male candidate performs in female-dominated jobs. Yet, the results do not reveal differential treatment of men contingent on job type as the point estimate of the 'female-dominated job' dummy depicting men's callback probability net of female effects turns out to be insignificant.

Due to the fact that the underlying probability function in probit regression models is nonlinear, the effect size of the independent variables may vary as a function of all other independent variables included in the model. In table 5-14, average marginal effects are calculated at the mean of all other regressors. In order to represent a standard applicant addressing a standard employer, the discrete independent variables are alternatively fixed at their mode instead of their mean (see table C-5 in the appendix). This change produces minor differences in the magnitude of the effects, but neither influences their direction (sign) nor their significance level. Nevertheless, when only looking at marginal effects in case of interaction terms, misleading conclusions may be derived (see 5.2.3.1.2). Thus, the entire cross derivative (correct interaction effects) of the 'female x female-dominated job' interaction is calculated and displayed. Figure 5-9 outlines that the effect is positive and statistically significant independent of the predicted probabilities of the observations in the sample.

Figure 5-9: Interaction Effect between Female and Female-Dominated Job Dummy



Restricting the sample may be useful for analyzing whether the results are sensitive to employers not responding at all or by those having already completed their recruitment process. Especially in case of the latter, findings on differential treatment are likely to be biased since both applicants are rejected even though no actual evaluation on behalf of the recruiters has taken place. Thus, no statement on whether discrimination would have occurred can be made. Yet, both the effect of the female dummy and the interaction term remain robust if the sample is restricted to those employers that responded to (N=1,152) or called back (N=626) at least one of the candidates.[66] Thus, overall, '$H_{job\ type}$' stating that the female applicant is being discriminated in male-dominated jobs cannot be rejected.

Concerning the GoF measures of the regression models, the p-values indicate that all specifications predict the callback probability significantly better than the intercept model which estimates the outcome by pure chance. Nevertheless, even for probit analyses the pseudo R^2 are rather low varying between 0.01 and 0.022. This is due to the nature of the correspondence study which limits the difference between two applicants to one single attribute (such as gender) where all other things such as schooling and labor market experience are kept constant during the application process. As a consequence, the variance in independent variables is quite low. In a nutshell, experimental control comes at the expense of estimation quality in terms of model fit. The regression results and conclusions derived with regard to the hypotheses, however, do not seem to be affected as appears from alternative estimation methods, different model specifications and various robustness checks.

In order to evaluate the source of discrimination and to test the hypotheses on statistical ('$H_{certificate}$' and '$H_{share\ of\ females}$', respectively) as well as taste-based discrimination ('H_{timing}' and '$H_{scarcity}$', respectively), the sample is subsequently restricted to occupations employing a male majority which reduces the number of observations to 1,080. Table 5-15 depicts average marginal effects of regressions on the callback dummy. In particular, the joint and interaction effects of the independent variables are presented. Model (I) reports the single effects of gender, a certificate dummy, the share of female applicants in the previous year, a dummy for late recruiters as well as the ratio between vacancies and total jobs in the previous year. Models (IIa) to (IId) include an interaction term between the female dummy and either of these variables and model (III) takes into account all single and interaction effects. All other regressors are considered in the analysis, but not reported. The effects displayed below remain robust independent of the inclusion of additional controls (see table C-6 in the appendix).

---

66 Results for the restricted samples are available from the author upon request.

Table 5-15: Marginal Effects from Probit Regressions on Callback Dummy and Hypotheses Testing

Callback	**(I)**	**(IIa)**	**(IIb)**	**(IIc)**	**(IId)**	**(III)**
Female	-0.067***	-0.062**	-0.067***	-0.029	-0.067***	0.043
	(0.020)	(0.028)	(0.019)	(0.027)	(0.020)	(0.062)
Certificate	0.025	0.033	0.026	0.025	0.024	0.078
	(0.036)	(0.046)	(0.036)	(0.036)	(0.036)	(0.056)
Female x Certificate		-0.016				-0.100
		(0.057)				(0.078)
Share of females t-1	-0.021	-0.021	-0.047**	-0.021	-0.021	-0.049**
	(0.020)	(0.020)	(0.023)	(0.020)	(0.020)	(0.023)
Female x			0.052**			0.055**
Share of females t-1			(0.022)			(0.022)
Late recruiter	-0.021	-0.021	-0.021	0.017	-0.021	0.052
	(0.088)	(0.088)	(0.088)	(0.091)	(0.088)	(0.095)
Female x				-0.072*		-0.134**
Late recruiter				(0.038)		(0.059)
Vacancies/total jobs t-1	0.002	0.002	0.002	0.002	-0.016	-0.012
	(0.022)	(0.022)	(0.022)	(0.022)	(0.023)	(0.023)
Female x					0.037**	0.030
Vacancies/total jobs t-1					(0.018)	(0.018)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
No. of obs.	1,080	1,080	1,080	1,080	1,080	1,080
Pseudo R^2	0.026	0.026	0.028	0.027	0.027	0.031
Log likelihood	-696.980	-696.948	-695.456	-696.244	-696.198	-693.120
Wald chi-squared	31.831	32.142	42.605	32.828	35.728	49.631
P-value	0.016	0.021	0.001	0.018	0.008	0.000

Notes: Each model reports average marginal effects of a probit regression on the callback dummy (Y=1: employer calls back the job applicant). Marginal effects are calculated at the mean of all independent variables and denote an infinitesimal change in case of continuous variables and a discrete change in case of dummy variables. Standard errors clustered on firm level are in parentheses. Regressions consider only male-dominated jobs. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

Without any interaction, the callback probability of the female candidate is on average 6.7 percentage points lower compared to the male applicant (see model (I)). The effect size goes along with the results in model (V) of table 5-14 which reports a 7.0 percentage points lower chance of receiving an invitation for women if the effect from an interaction between the female dummy and female-dominated jobs is controlled for. The third column (model (IIa)) includes an interaction that denotes the hypothesized beneficial effect of the female applicant providing an additional productivity signal. However, the interaction is not statistically significant holding all other independent variables constant at their mean. The insignificant interaction remains the same independent of the predicted probability (see figure C-1 in the appendix). Hence, as already indicated by table 5-8, the additional certificate does not reduce gender discrimination and '$H_{certificate}$' can be rejected.

Next, model (IIb) explicitly investigates whether the callback probability for women is influenced by the share of female applicants in the previous year. According to the '$H_{share\ of\ females}$' hypothesis, employers should treat women more favorably the more they have previously been in contact with them. Indeed, the regression results support this assumption. The probability of a callback to the female applicant is on average 5.2 percentage points higher and statistically significant at the 5 percent level if the share of female applicants increases by one standard deviation. The statistical significance holds for all predicted probabilities across the sample (see figure C-2 in the appendix). In contrast, the callback probability for the male candidate decreases by almost 5 percentage points (as can be shown by the point estimate of the variable 'share of females t-1'). These findings lend support to the idea that an informational deficit reduces the minority (female) group's callback rate. In contrast, increasing experience obviously raises women's callback probability. Yet, the overall gender effect does not change, i.e., the female candidate is significantly disadvantaged independent of employers' previous experience.

Model (IIc) reveals somewhat surprising results. In contrast to 'H_{timing}', late recruiters do not react to time pressure by inviting both male and female job candidates equally often. While the female applicant on average suffers from a 7.2 percentage points lower callback rate when sending out applications to employers in May (2011 and 2012), the female dummy denoting differences in callback probabilities at early recruiters turns out to be statistically insignificant (p=0.290). This finding particularly contradicts 'H_{timing}' according to which firms being confronted with potential losses from not filling a vacancy are expected to discriminate less, if at all. The results turn out to be quite robust contingent on different predicted probabilities (see figure C-3 in the appendix).

Model (IId) provides additional insights on how the recruiting behavior of firms develops with a change in the supply of suitable apprentices ('$H_{scarcity}$'). The interaction term states that the callback probability for the female candidate increases by 3.7 percentage points if labor market scarcity (denoted by the ratio between vacancies and total jobs in t-1) increases by one standard deviation. This relationship turns out to be statistically significant (at the 5 percent level) across the entire probability distribution (see figure C-4 in the appendix). Again, however, the coefficient of the female dummy remains unchanged indicating that the effects from labor market scarcity do not eliminate discrimination.

Referring to the robustness of the interaction terms, the last column (model III) reflects the joint effect of all interactions. The results support the '$H_{share\ of\ females}$' hypothesis. Both the point estimate (share of females t-1) and the interaction (female x share of females t-1) do not differ with respect to their effect size and significance level compared to model (IIb). Focusing on the interaction between the female dummy and late recruiters reveals that the coefficient from model (IIc) becomes even more negative. Females who address job offers from late recruiters have a 13.4 percentage points lower probability of being called back. '$H_{scarcity}$', however, cannot fully be supported as the interaction coefficient becomes insignificant (though p=0.105).

Apart from the findings on differential treatment, not many effects from the probit estimates turn out to be statistically significant except for the ones of the firm size dummies. Table 5-14 reveals that applications arriving at medium-sized companies have on average an 11 percentage points higher success probability compared to the reference group, i.e., firms with less than 50 employees. A closer look reveals that these results are particularly affected by a higher fraction of small recruiters that do not respond to any of the candidates indicating that these firms have less formalized recruiting procedures. Since the firm size effect only proves to be significant for the entire sample, but becomes insignificant as soon as the sample is restricted to male-dominated occupations (results not displayed, but available upon request), further discussions should be extended towards the more interesting question on whether any firm characteristics interact with the female dummy and thus affect gender discrimination.

Table 5-16 displays average marginal effects of a probit regression with these interactions. Model (I) includes all observations while model (II) is restricted to male-dominated jobs. The direct effects of the variables interacted are included, but not reported for the sake of brevity. The results support the findings presented above. While all other interactions turn out to be statistically insignificant, female applicants have a lower callback probability when applying for male-dominated jobs at late recruiters. Apart from that, neither firm size, location and industry nor recruiters' sex significantly interact with the female dummy.[67]

[67] As the internal recruitment process is like a black box to the researcher, i.e., there is no possibility to find out whether the application is forwarded to the department in which the candidate is employed or directly decided upon in the HR department, any hypothesized relations between candidates' callbacks and recruiters' sex are speculative. Particularly in large firms the applications often address an HR official but are forwarded to the foreman or training officer who then makes the actual employment decision. Any effects of recruiter characteristics are thus likely to be biased and only have weak, if any, explanatory power. Previous research analyzing the effect of recruiters' sex identified whether the person responsible for hiring was a man or woman either due to personal audits or phone calls (see e.g. Carlsson, 2011).

Table 5-16: Marginal Effects from Probit Regressions on Callback Dummy and Interaction of Female Dummy and Firm Characteristics

Callback	(I)	(II)
Female x Medium	-0.060	-0.062
	(0.043)	(0.047)
Female x Large	-0.014	0.001
	(0.057)	(0.062)
Female x South	0.016	0.033
	(0.045)	(0.049)
Female x East	0.054	0.054
	(0.056)	(0.065)
Female x Industry	-0.007	0.098
	(0.047)	(0.065)
Female x Female responsible	-0.015	-0.001
	(0.036)	(0.039)
Female x Late recruiter	-0.054	-0.077**
	(0.037)	(0.037)
Controls	Yes	Yes
No. of obs.	1,312	1,080
Pseudo R^2	0.022	0.029
Log likelihood	-851.143	-695.062
Wald chi-squared	36.366	39.850
P-value	0.066	0.022

Notes: Each model reports average marginal effects of a probit regression on the callback dummy (Y=1: employer calls back the job applicant). Marginal effects are calculated at the mean of all independent variables. Standard errors clustered on firm level are in parentheses. Model (I) considers the full sample, model (II) is restricted to male-dominated jobs. Controls include all point estimates of the variables interacted. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

Thus far, the analyses have revealed three main findings. First, gender discrimination clearly depends on the job type. Second, the concepts of taste-based and statistical discrimination as proxied in the regression models cannot fully explain why women suffer from lower callback rates in male-dominated jobs. And third, firms' recruiting behavior affects discriminatory treatment, though in the opposite direction to what has been expected. Either of these results certainly requires a closer inspection.

5.2.4 Discussion

Next, the regression results reported above are discussed separately accounting for the potential sources of gender hiring discrimination.

5.2.4.1 Job Stereotyping and Gender Discrimination

Regression estimates from table 5-14 confirm that job stereotyping exists and disadvantages female applicants when applying for male-dominated apprenticeships. The difference in callback rates varies between 7 and 11 percentage points and thus oscillates around the lower end of what has been found in other matched-pair studies reporting callback differences between 5 and 35 percentage points (see table A-1 in the appendix; note also that some of these studies do not find statistically significant callback differences by sex). One reason why the extent of discrimination is rather low can be identified when looking at the labor market situation of the jobs addressed. Choosing technical occupations where current and future labor demand is expected to be high, on the one hand, increases the probability to observe a sufficient number of mutual and one-sided rejections and callbacks allowing the researcher to carry out statistical tests. On the other hand, the extent of discrimination may be affected by the job referred to, in particular when employers respond to scarcity. Thus, assuming that the matched-pair applicants only address jobs where competition for talent is intense (relaxed), the magnitude of differential treatment is expected to be lower (higher) as compared to other occupations. Yet, without a control group, i.e., correspondence tests using the same pair of applicants in less demanded jobs, no final judgment can be made whether the callback difference is influenced by the job offers referred to or any other impact factors. In case of the former, the line of argument is closely related to the theory of taste discrimination which will be addressed in section 5.2.4.3 (even though not job type, but regional labor supply is used to find out more about employers' preferences).

In contrast to the present study, previous research also yields significantly fewer callbacks for males in female-dominated occupations where the differences fall in a range between 3 and 44 percentage points. The reasons why these results cannot be reproduced in this field experiment are quite obvious. As the main purpose was to investigate the sources of discrimination in clearly male-dominated professions, varying the job type only served as a control limiting the number of observations to a minimum. Hence, gender equality in callbacks might predominately stem from the relatively low share of female-dominated jobs addressed in the experiment (roughly 18%). Moreover, the selected jobs have two more peculiarities. First, the market for industrial clerks is not as gender segregated as other labor market segments. In fact, the difference between the share of men and women working in this field is relatively low compared to e.g. the industrial mechanic profession (see section 5.1.1.2). Thus, the classification as being female-stereotyped can well be contested. In fact, denoting this type of job as 'gender-neutral' or 'gender-integrated' might be more suitable. Second, the demand for apprentices in Germany's health care sector currently exceeds the demand in any other industry. This in turn may have led to gender 'callback equality'. Indeed, callback rates for either candidate were above 60 percent (62.5 percent for the male and 66.7 percent for the female candidate) and thus significantly higher than in all other occupations addressed (see table C-7 in the appendix for a detailed overview of callbacks by type of apprenticeship). Conclusions with regard to (the absence of) discriminatory treatment of men in female-dominated jobs should therefore be drawn only carefully.

With regard to theory, the confirmation of '$H_{job\ type}$' could somewhat be regarded as an indicator of statistical discrimination. Classifying jobs as either male- or female-stereotyped simply stems from segregated labor markets and an overrepresentation of either gender in some occupations. Segregation in turn produces differences in employers' accumulated experience where productivity information is expected to be superior or more precise for majority workers. Consequently, employers would have an economic rationale to favor men over women and vice versa. However, neither previous evidence (Booth and Leigh, 2010), nor the data from this study directly support this relationship. That is, the share of women working in different male-dominated occupations does not correlate with callback differences.

5.2.4.2 GROUP EXPERIENCE AND THE ROLE OF ADDITIONAL SIGNALS

The results from model (IIb) in table 5-15 suggest that employers discriminate somewhat less with an increasing proportion of female candidates in the previous application period. Apparently, as postulated, increasing experience with women, denoted as the share of female applicants for technical apprenticeships in the previous year and respective labor market region, allows employers to evaluate their quality more precisely. In turn, they invite women equally often as their male counterparts. A closer look, however, challenges this interpretation. Even though the effect of the interaction term is positive and statistically significant, discrimination against women as proxied by the (negative) point estimate of the female dummy does not disappear. The increasing likelihood of women being called back comes at the costs of men whose callback probability declines with a rising female applicant ratio, but does not compensate the gender callback gap.

Still, the main findings turn out to be robust. This is particularly highlighted if the sample is split at the mean of the standardized 'share of females t-1' variable, i.e., zero (see table 5-17). For the above mean sample (model (I)), the gender coefficient turns out to be insignificant (so does the whole model), whereas for the below mean sample (model (II)), the difference in callback rates is statistically significant at the 1 percent level and amounts to 10.3 percentage points. Hence, '$H_{share\ of\ females}$' as a test for statistical discrimination finds weak support, although it may well be assumed that it does not explain the entire gender gap in hiring.

Table 5-17: Marginal Effects from Probit Regressions on Callback Dummy with Sample Split at the Mean Share of Females t-1

Callback	(I)	(II)
Female	-0.024	-0.103***
	(0.033)	(0.026)
Certificate	Yes	Yes
Late recruiter	Yes	Yes
Vacancies/total jobs t-1	Yes	Yes
Firm characteristics	Yes	Yes
Controls	Yes	Yes
No. of obs.	448	632
Pseudo R^2	0.045	0.052
Log likelihood	-279.321	-400.662
Wald chi-squared	18.733	41.334
P-value	0.283	0.000

Notes: Each model reports average marginal effects of a probit regression on the callback dummy (Y=1: employer calls back the job applicant). Marginal effects are calculated at the means of all independent variables and denote an infinitesimal change in case of continuous variables and a discrete change in case of dummy variables. Standard errors clustered on firm level are in parentheses. Model (I) considers all observations where the standardized share of females in t-1 is above the average, i.e., zero, model (II) reports results for all applications in areas below the average. Either model includes only male-dominated jobs. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

As an alternative indicator of statistical discrimination, additional certificates on job-related internships have been attached to the applications. Yet, unlike in e.g. Heilman et al. (1988), the provision of these credentials does not influence gender differences in callbacks. Neither does the effect of the female dummy change, nor does the female-certificate interaction turn out to have a statistically significant impact on callback probabilities (see model (IIa) in table 5-15). However, the rejection of '$H_{certificate}$' does not necessarily speak against the prevalence of statistical discrimination. Two alternative explanations are equally plausible.

On the one hand, employers might consider the provision of a certified internship as a weak productivity signal compared to school credentials and thus assign them only a minor role when assessing applicants' future productivity. As a result, callbacks to both male and female applicants do not significantly increase and affect gender differences. On the other hand, attaching an additional certificate may put one group at an advantage, but disadvantage the other. This would produce two scenarios: either the gap in callbacks increases between groups because additional information strengthens the market position of the established group, i.e., the male candidate benefits while the female does not, or the group difference in callbacks declines because the reduction of information asymmetries benefits the minority group. Descriptive statistics suggest that the provision of additional productivity information significantly increases the callback probability for the male applicant ($p=0.067$), but leaves callbacks to the female candidate unaffected ($p=0.267$) (see table 5-8). The beneficial effect for men also holds if the sample is restricted to male-dominated jobs (not displayed, but available upon request). Consequently, the hiring gap rather widens than decreases. This is in line with research by Neumark (1999) and Pinkston (2003) who show that employers' perception of credentials may differ by gender where majority candidates benefit relative to minority candidates at the beginning of the employer-employee relationship. However, multivariate analyses do not corroborate these results. As model (IIa) in table 5-15 indicates, signaling professional expertise in technical occupations leaves the callback difference unaffected in either way.

5.2.4.3 LABOR MARKET SCARCITY AND RECRUITER EFFECTS

Thus far, statistical discrimination has been shown to explain some of the findings from the correspondence test. Nevertheless, as demonstrated above, the study also finds evidence for taste-based discrimination. A tighter labor market in the previous year works in favor of women and induces an increase in callback rates (see model (IId) in table 5-15).[68] Yet, this increase does not affect the male-female callback gap which remains stable at around 6.7 percentage points. A sample split at the mean of the 'vacancies/total jobs t-1' variable and a probit regression on callbacks (controlling, inter alia, for recruiter type) yields no differential treatment if the standardized scarcity ratio exceeds zero (see model (I) of table 5-18), but an 11.6 percentage points callback difference in disfavor of women (on a 1 percent significance level) if it is below zero (see model (II)). Put differently, discrimination is restricted to employers that face little if any labor market scarcity and can thus 'afford' neglecting minority group candidates. On the other hand, firms that are confronted with fierce competition for suitable apprentices would incur higher costs for not recruiting women due to e.g. additional search activities and productivity losses. They therefore respond rationally by employing women. This in turn

[68] Note that alternative scarcity measures have also been tested, but were found not to be significant.

is consistent with Becker's taste for discrimination approach (Becker, 1971).

Table 5-18: Marginal Effects from Probit Regressions on Callback Dummy with Sample Split at the Mean Vacancies/Total Jobs t-1

Callback	**(I)**	**(II)**
Female	0.002	-0.116***
	(0.033)	(0.026)
Certificate	Yes	Yes
Late recruiter	Yes	Yes
Females/total applicants t-1	Yes	Yes
Firm characteristics	Yes	Yes
Controls	Yes	Yes
No. of obs.	446	634
Pseudo R^2	0.044	0.047
Log likelihood	-277.282	-404.814
Wald chi-squared	16.211	39.858
P-value	0.438	0.001

Notes: Each model reports average marginal effects of a probit regression on the callback dummy (Y=1: employer calls back the job applicant). Marginal effects are calculated at the means of all independent variables and denote an infinitesimal change in case of continuous variables and a discrete change in case of dummy variables. Standard errors clustered on firm level are in parentheses. Model (I) considers all observations where the standardized vacancies/total jobs variable is above the average, i.e., zero, model (II) reports results for all applications in areas below the average. Either model includes only male-dominated jobs. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

Another interesting finding on taste discrimination has recently been published by Kuhn and Shen (2013). They show that gender-targeted job advertisements decrease with skill requirements. They interpret this as a sign for taste discrimination since the supply of more qualified labor is scarce and thus distastes become more costly. Fortunately, the job offers used for the present field experiment also include information on job requirements. Two different types of employers could be identified where about one half requires at least a degree from middle school (N=556) while the other half accepts a school degree lower than middle school (N=524). When splitting the sample by school degree, however, the results do not differ from each other, i.e., the female candidate is significantly discriminated independent of skill level (results available upon request). Thus, in the present context, employers either do not face labor-supply differences by school degree or do not respond to supply differences by inviting the minority candidate equally often than her majority counterpart. The absence of the postulated effect, though, may also stem from a different operationalization of labor market discrimination. While Kuhn and Shen (2013) investigate the statements of employers by observing gender-targeted wording in job offers, the present study assesses how employers actually react. As has been shown in chapter 3, stated and revealed preferences may indeed differ with regard to employment outcomes.

Referring to the three types of preference-based discrimination, i.e., employer, coworker and customer discrimination, the data do unfortunately not provide enough information to separate the effects inherent in any of these concepts. However, anecdotal evidence from firms' responses particularly points in two directions suggesting that employer and coworker discrimination might play a meaningful role. The former type can be exemplified by an email that, even though apparently written to foster internal decision making, was accidently forwarded to the female applicant. In this email, the potential supervisor states that from his point of view the female candidate looks too young and dainty for the job. Here, the gender and profile picture serve as a pre-selection device that is clearly linked to employers' prejudices. But the mechanisms in the hiring process might also indicate coworker discrimination. In another case where an employer involuntarily attached internal email correspondence, it was disclosed that the recruiters expect coworker discrimination against the female candidate. In particular, they doubted that a young woman would be able to handle the occasionally very rough tone in a work environment where male colleagues dominate. Interestingly, the female applicant was still invited which, of course, does not exclude that other employers rejected her for exactly the same reason. The persistence of customer discrimination as a third component, e.g. shown by Neumark (1996), can be disregarded in the present context. Firstly, technical apprentices do usually not get in contact with firms' customers and, secondly, discrimination does not significantly vary across firms that operate in the industry and service sector, respectively (see insignificant female-industry interaction in table 5-16).

Another response outlines the whole dilemma when attempting to distinguish between different forms of taste-based discrimination. One employer offered a position as an industrial clerk rather than as a warehouse logistics operator to the female applicant while the same employer invited the male candidate for the job that he originally applied for. The email sent to the female applicant included favorable statements on the fit of her profile and the company's products and customers. Yet, it indirectly recommended that administrative tasks might suit her better than technical ones (which is also referred to as "job channeling" in the literature). This could imply at least two considerations. On the one hand, the recruiter might have anticipated coworker discrimination in the respective department and thus looked for alternative options or, on the other hand, firms' representatives could have used this argument as a means of covering their own personal distaste.[69] Either way, the interpretations of firms' responses refer to single observations and can, of course, not be generalized. In fact, more research is required that leads to a better understanding of how these three components affect the hiring decision. For the purpose of this thesis (though not for policy implications in general), further differentiations are disregarded as they yield the same hiring outcome in the end.

Table 5-19: Marginal Effects from Probit Regressions on Callback Dummy with Sample Split by Recruiter Type

Callback	**(Ia)**	**(Ib)**	**(IIa)**	**(IIb)**
Female	-0.031	-0.048	-0.097***	-0.100***
	(0.025)	(0.031)	(0.027)	(0.028)
Certificate	No	Yes	No	Yes
Females/total applicants t-1	No	Yes	No	Yes
Vacancies/total jobs t-1	No	Yes	No	Yes
Firm characteristics	No	Yes	No	Yes
Controls	No	Yes	No	Yes
No. of obs.	522	522	558	558
Pseudo R^2	0.001	0.092	0.008	0.031
Log likelihood	-352.315	-319.982	-358.209	-349.849
Wald chi-squared	1.456	36.445	12.764	21.785
P-value	0.228	0.002	0.000	0.150

Notes: Each model reports average marginal effects of a probit regression on the callback dummy (Y=1: employer calls back the job applicant). Marginal effects are calculated at the means of all independent variables and denote an infinitesimal change in case of continuous variables and a discrete change in case of dummy variables. Standard errors clustered on firm level are in parentheses. Models (Ia) and (Ib) consider early recruiter sample, models (IIa) and (IIb) late recruiter sample. Either model includes only male-dominated jobs. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

Apart from statistical and preference-based discrimination, the regression estimates have revealed that firms' response behavior towards women varies systematically by recruiter type where gender discrimination in male-dominated jobs is restricted to late recruiters as demonstrated by model (IIc) in table 5-15. While the 'female-late recruiter interaction term turns out to be statistically significant and negative, the female coefficient becomes insignificant. To circumvent problems resulting from interaction effects in probit models and to check the robustness of the recruiter effect, the probit regression on the callback dummy is conducted separately for late and early recruiters. Results of the latter are displayed in models (Ia) and (Ib) of table 5-19, results of the former can be found in models (IIa) and (IIb). While the female candidate is not treated differently in the early-

[69] In fact, in the present case, the employer did not invite the female applicant to a job interview while her male counterpart received a callback.

recruiter sample, she has a 9.7 to 10.0 percentage points lower callback probability when applying at late recruiters. Either effect persists independent of controls (though the inclusion of controls apparently affects the model fit). Hence, quite surprisingly, the results of both the regression model with interaction effect as well as the robustness checks with sample split by recruiter type suggest exactly the opposite to what has been hypothesized in 'H_{timing}'. Recruiter type does not reflect the need to hire apprentices and thus offers clear evidence for taste discrimination, but may signal management quality.

Table 5-20: Marginal Effects from Probit Regression on Late Recruiter Dummy

Late recruiter	**(I)**
Medium	-0.24***
	(0.05)
Large	-0.31***
	(0.07)
South	0.12**
	(0.06)
East	0.37***
	(0.06)
Industry	-0.16**
	(0.07)
Female responsible	-0.08*
	(0.05)
Share female applicants t-1	0.01
	(0.03)
Vacancies/total jobs t-1	-0.10***
	(0.03)
Open positions	-0.03
	(0.02)
No. of obs.	1,080
Pseudo R^2	0.137
Log likelihood	-645.623
Wald chi-squared	90.558
P-value	0.000

Notes: Table reports average marginal effects of a probit regression on the late recruiter dummy (Y=1: firm offers vacancy in May). Standard errors clustered on firm level are in parentheses. Results are restricted to male-dominated jobs. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

Table 5-20 reveals systematic differences between late and early recruiters with respect to firm and labor market characteristics. It denotes average marginal effects from the probability $P_i\ (Y = 1)$ of being a late recruiter versus $P_i\ (Y = 0)$ of being an early recruiter. Probit regression estimates show that late recruiters (i) are more likely to be small, (ii) are overrepresented in the East and the South of Germany, (iii) operate in the service sector and (iv) more often have a female responsible for the recruitment of apprentices.

Moreover, late recruiters find themselves in areas where the situation in the labor market is rather relaxed, while early recruiters face a higher degree of labor market scarcity.[70] More precisely, a one standard deviation increase in labor market scarcity (vacancies/total jobs t-1) significantly (at the 1 percent level) reduces the probability that the employer is a late recruiter by 10 percentage points. This relationship may also explain why the significant 'female x vacancies/total jobs t-1' interaction disappears if the female dummy is additionally interacted with recruiter type (see model (III) of table 5-15).

Previous analyses have already demonstrated that (even if) accounting for labor market conditions and other firm characteristics, the recruiter effect persists. Consequently, the question arises why late and early recruiters treat the female candidate differently. Several explanations seem equally plausible. The first deals with management quality. Late recruiters may employ less professional recruitment processes that systematically disadvantage minority workers. The data compiled provide a possibility to proxy and thus to empirically test the lack of managerial expertise.

Table 5-21 reports average marginal effects of a probit regression on (i) the response dummy and (ii) a dummy for the employer's reaction after being reminded by the job candidate given (i). Both dependent variables should serve as an indicator on how reliable and organized firms' recruiting processes are. The results do not reveal significant differences by recruiter type concerning the response probability, but show systematic variations with respect to the reminder dummy. The probability that late recruiters answer only after having been reminded by the applicant is 15.8 percentage points higher than in case of early recruiters. This, indeed, can be considered as evidence for (poor) management quality affecting gender inequality in recruiting decisions. Relating these findings to the large-scaled survey data on management practices presented by Bloom and van Reenen (2007) and Bloom et al. (2012) indicates that firm size moderates the effects. They find that the average management score with respect to how human capital is attracted, managed and retained increases with company size. These quality indicators, in turn, are shown to have a positive and significant effect on firm performance. As the recruiter type in the present studies correlates with firm size, the argumentation outlined above finds support in the Bloom and van Reenen data.

---

70 Note that the regression coefficients hardly change if the entire sample rather than the sample with male-dominated jobs only is considered.

Table 5-21: Marginal Effects from Probit Regressions on Response and Reaction to Reminder Dummy

	(Response)	(Reaction to reminder)
Late recruiter	0.001	0.158***
	(0.032)	(0.039)
Firm characteristics	Yes	Yes
No. of obs.	1,080	877
Pseudo R^2	0.040	0.071
Log likelihood	-501.160	-468.946
Wald chi-squared	27.796	48.075
P-value	0.000	0.000

Notes: Table reports average marginal effects of a probit regression on the response (Y=1: applicant receives a response on behalf of the employer) and reacting to reminder (Y=1: firm responds only after being reminded given that a firm responds at all) dummy, respectively. Standard errors clustered on firm level are in parentheses. Results are restricted to male-dominated jobs. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

Secondly, late recruiters may simply fail to find adequate staff even though their job offers had been published a long time ago.[71] On the one hand, the threshold level for potential apprentices could be too high. This idea turns out to be rather unlikely as the majority of jobs only mention quite moderate scholastic requirements (see above). Also, the overall callback rates do not differ between applications sent out in May and September.[72] Alternatively, employers' reputation could differ between late and early recruiters. It may well be that the former do not find adequate staff as a sanction of the labor market to discriminating behavior in the past. Being a late recruiter would then be the result of a negative selection effect. Unfortunately, no panel data are available to test this assumption.

Third, late recruiters may treat the male and female applicant differently as a result of statistical discrimination. As they are under pressure to find apprentices in time, they select members of the majority group in order to minimize the probability of inviting an unsuitable person. Moreover, what is generally referred to as "rough sorting" might be involved (see e.g. Carlsson and Rooth, 2008). In the context of male-dominated jobs, gender might serve as a (first) screening device without looking more closely at the information provided by the applications which again would result in the minority candidate being rejected to a larger extent. In contrast, early recruiters have enough time

71 Unfortunately, the length of time the vacancy had already been published could not be recorded.

72 Note that the applicant pool may differ across application periods. Assuming that the better qualified candidates are more likely to apply for a job at early recruiters, the quality of the applicant pool would be lower in the late recruiter sample. As applicants' quality remained constant for the entire experiment, this on average should have led to a lower callback rate for the applications sent out in September. However, no support for significant callback differences can be found in the data.

and probably a multilevel hiring process to carefully select the candidates with the best fit implying that they give men and women equal opportunities. This can be supported by comparing waiting periods conditional on recruiter type. While late recruiters on average give a callback (rejection) after 9.7 (18.9) working days, early recruiters need 27.3 (45.6) working days to make a decision.[73]

Overall, the results discussed in this section suggest that taste and statistical discrimination in conjunction with a recruiter effect are responsible for gender discrimination in the labor market for apprenticeships.

5.2.4.4 THE ROLE OF SOCIETAL ATTITUDES

The discussion about where a taste for discrimination might stem from has revealed that societal attitudes may affect employers' response behavior towards minorities. Previous research has shown that, for example, the treatment of women varies conditional on how people in different regions vote on gender issues (Fortin, 2005; Backes-Gellner et al., 2013). If the majority votes in favor of policies promoting gender equality, employers are found not to discriminate. Conversely, in regions where the public opinion challenges affirmative action fostering gender-equal employment outcomes, employers seem to adapt regional tastes in their hiring and pay practices. Whereas former studies use natural experiments originating from national referendums or the results of social surveys, no such information is available for Germany.

However, what might reflect regional attitudes on the role of men and women in the labor market is the share of votes different parties receive in general elections. While some parties like the Christian Democratic Union (CDU) and the Christian Socialist Union (CSU) are considered to be more conservative with a traditional understanding of the role of men and women in society (which, very simplified, reflects the 'breadwinner' versus 'housekeeper' discussion), others, like the Social Democratic Party (SPD), the Green Party (Die Grünen) and the Free Democratic Party (FDP), represent a more liberal way promoting women's labor market participation. Following these assumptions, the

[73] Including applicants' waiting period in the regression model does not qualitatively affect the results (estimations not displayed but available upon request). Furthermore, interacting the waiting period with the female dummy does not reveal any gender differences with respect to response times. However, the waiting period turns out to have a U-shaped relationship on callback probabilities if the sample is restricted to late recruiters whereas the relationship is linear if only the early recruiter sample is considered. These results somewhat support the assumption that recruitment processes differ by recruiter type.

probability that the female candidate is discriminated in male-dominated jobs should increase with the proportion of votes accumulated by the CDU/CSU and decrease with a rise in popularity of SPD, Die Grünen and FDP. To empirically investigate this relationship, the regional results from the last federal elections in 2009 are matched with employer data.

Table 5-22: Marginal Effects from Probit Regressions on Callback Dummy and Interaction of Female Dummy and Share of CDU/CSU Votes

Callback	**(Ia)**	**(Ib)**	**(IIa)**	**(IIb)**
Female	-0.214*	-0.220*	-0.076***	-0.080***
	(0.122)	(0.124)	(0.028)	(0.029)
Share CDU/CSU votes	-0.007**	-0.003	-0.005*	-0.001
	(0.003)	(0.004)	(0.003)	(0.004)
Female x	0.004	0.004		
Share CDU/CSU votes	(0.004)	(0.004)		
Female x			0.023	0.027
Share CDU/CSU votes above average			(0.043)	(0.044)
Controls	No	Yes	No	Yes
No. of obs.	1,080	1,080	1,080	1,080
Pseudo R^2	0.007	0.026	0.006	0.026
Log likelihood	-710.847	-696.561	-711.136	-696.824
Wald chi-squared	16.678	34.327	14.810	32.520
P-value	0.001	0.017	0.002	0.027

Notes: Table reports average marginal effects of a probit regression on the callback dummy (Y=1: employer calls back the job applicant). Marginal effects are calculated at the mean of all independent variables and denote an infinitesimal change in case of continuous variables and a discrete change in case of dummy variables. Standard errors clustered on firm level are in parentheses. Samples are restricted to male-dominated jobs. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

In federal elections voters have two votes, the first going towards the regional representative and the second determining the number of seats in the German Federal Parliament. The sample average of the first (second) CDU/CSU vote is 40.9 percent (34.8 percent). Table 5-22 reports average marginal effects of a probit regression on the callback dummy where models (Ia) and (Ib) include an interaction of the female dummy and the share of second CDU/CSU votes while models (IIa) and (IIb) add an interaction between the female and above-average CDU/CSU dummy. All models are restricted to male-dominated jobs and either include (models (Ia) and (IIa)) or exclude (models (Ib) and (IIb)) control variables. Comparing the regression estimates, however, does not support the hypothesized effect, i.e., the coefficient of the female dummy turns out to be negative and significantly different from zero independent of the inclusion of an interaction effect. In other words, the results do not suggest a correlation between voting behavior and the extent of discrimination towards women. Using alternative measures such as the proportion of CDU/CSU first votes or electoral results of other parties (expecting a reverse effect) do not help explaining why gender differences in hiring can be observed.

This might have two reasons. On the one hand, voting behavior may not be an adequate proxy for societal attitudes, especially because the profiles and programs of the major parties in Germany are hard to disentangle, so are their gender role models. This, in turn, makes assumptions on the electorate and their attitudes concerning gender equality in the labor market very speculative. On the other hand, employers might not adapt regional attitudes when forming personal tastes.

5.3 Correspondence Study on Ethnic Discrimination

This section presents the results of the correspondence testing for ethnic discrimination. The structure is very similar to the gender study presented above. In section 5.3.1, the dataset is described, sections 5.3.2 and 5.3.3 present descriptive and empirical results and section 5.3.4 concludes with a discussion of the findings.

5.3.1 Data

Analogously to the presentation of the results on gender hiring discrimination, the dataset is described (5.3.1.1) before the characteristics of the employers addressed in the field experiment are compared with those from the entire body of training companies in Germany (5.3.1.2).

5.3.1.1 The Dataset from the Field Experiment

All in all, 1,246 applications were sent out to 623 different employers of which 15 were disregarded due to dispatching errors. The remaining 1,216 applications produced a response rate of 79.1 percent and a callback rate of 37.2 percent. The firms on average responded within 25 working days where the preferred way of responding was by email (63.4 percent). Concerning company characteristics, the majority of firms were medium-sized (53.8 percent), located in the South of Germany (56.6 percent) and operating in the manufacturing sector (90 percent). Across the sample, 57.1 percent of all firms were referred to in May 2011 or 2012 and are therefore classified as late recruiters. Similar to the gender study, small firms are clearly underrepresented among early recruiters (14.6 percent) while the opposite holds true for medium- and large-sized companies (see table 5-23). On average, employers offered 1.71 open positions while, again, this number correlates with firm size. According to the job advertisements, around half of the people dealing with the applications were female.

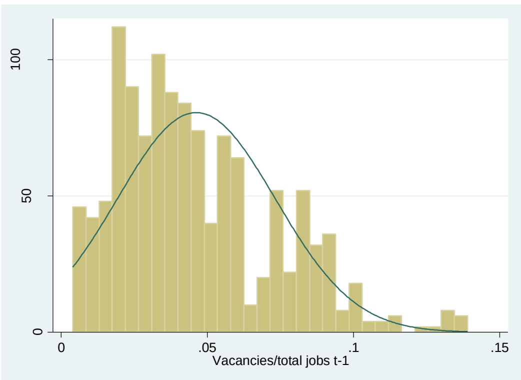
Table 5-23: Firm Size by Application Period

	Late (N=347)	Early (N=261)	Total
Small	41.21%	14.56%	29.77%
	(143)	(38)	(181)
Medium	48.13%	61.30%	53.78%
	(167)	(160)	(327)
Large	10.66%	24.14%	16.45%
	(37)	(63)	(100)

Notes: The table reports late and early recruiters as a fraction of firm size in percent. Absolute numbers are in parentheses.

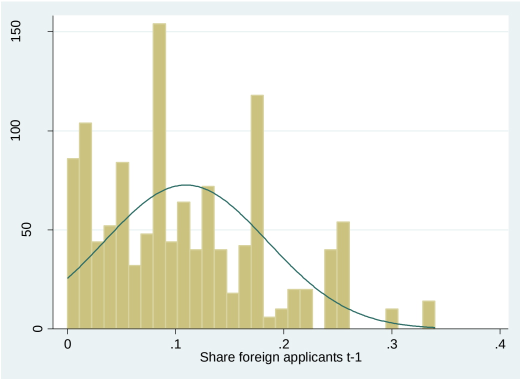
As any confounding effects between the callback rate and the ethnic background should be excluded, names, profile pictures, template designs, dispatching orders and places of origin were altered. The latter was controlled for including the distance between the workplace and the applicant's home (286 kilometers on average). Moreover, the last application period in May 2012 included alternative names ('Lukas Schmidt' for the German-named and 'Onur Öztürk' for the Turkish-named candidate). Apart from that, 37.5 percent of all candidates were equipped with an additional certificate documenting an internship in a technical occupation.

Figure 5-10: Frequency Distribution of Non-Standardized Vacancies/Total Jobs t-1



Data on labor market scarcity and the share of foreign applicants in the previous year were taken from the reports of the BA and matched with employers' respective labor market region. Analogous to the study on gender discrimination, scarcity is reflected by a ratio that divides the number of vacancies by the number of total apprenticeships reported. On average, 4.6 percent of the jobs remained unstaffed with the ratio varying between 0.4 to 13.9 percent. Figure 5-10 illustrates the non-standardized vacancies/total jobs t-1 variable as a frequency distribution.

Figure 5-11: Frequency Distribution of Non-Standardized Share of Foreigners t-1



The share of foreigners in t-1 proxies the fraction of applicants with non-German citizenship in the pool and thus reflects employers' likelihood of getting in touch with job candidates from minority groups. Since neither detailed information on the number of applicants with a migration background, nor on those with a Turkish migration background was available, this ratio serves as a proxy for employers' previous experience with other than German ethnicities. The fraction of foreigners averaged 11 percent in the entire sample, but varied between 0 and as much as 34 percent. An illustration of its non-standardized frequency distribution is provided in figure 5-11.

For the regression analyses, both measures reflecting the labor market situation are standardized in order to control for potential outlier effects and to facilitate the interpretation of the estimation coefficients. After all, table 5-24 provides an overview of the descriptives of the ethnicity study.

Table 5-24: Descriptive Statistics of the Correspondence Study on Ethnic Discrimination

Variable	Operationalization	# of Obs.	Mean	SD	Min	Max
DEPENDENT VARIABLES						
Response	Dummy: Equals 1 if the applicant receives a response (either invitation or rejection) by the employer, 0 otherwise	1216	0.791	-	0	1
Callback	Dummy: Equals 1 if the applicant receives a callback (e.g. invitation) by the employer, 0 otherwise	1216	0.372	-	0	1
INDEPENDENT VARIABLES						
Response information						
Response time	Response time of employers in working days	962	25.33	30.04	0	179
Type of response						
Email	Dummy: Equals 1 if employer responded by email, 0 otherwise	962	0.634	-	0	1

Postal mail	Dummy: Equals 1 if employer responded by postal mail, 0 otherwise	962	0.223	-	0	1
Phone	Dummy: Equals 1 if employer responded by phone, 0 otherwise	962	0.142	-	0	1
Applicant information						
Turkish name	Dummy: Equals 1 if the applicant has a Turkish-sounding name, 0 otherwise	1216	0.500	-	0	1
Name						
Jan Lange	Dummy: Equals 1 if the applicant is named 'Jan Lange', 0 otherwise	1216	0.457	-	0	1
Lukas Schmidt	Dummy: Equals 1 if the applicant is named 'Lukas Schmidt', 0 otherwise	1216	0.043	-	0	1
Kenan Yilmaz	Dummy: Equals 1 if the applicant is named 'Kenan Yilmaz', 0 otherwise	1216	0.461	-	0	1
Onur Öztürk	Dummy: Equals 1 if the applicant is named 'Onur Öztürk', 0 otherwise	1216	0.039	-	0	1
Photo						
Photo A	Dummy: Equals 1 if the applicant provides photo A, 0 otherwise	1216	0.500	-	0	1
Photo B	Dummy: Equals 1 if the applicant provides photo B, 0 otherwise	1216	0.500	-	0	1
Design						
Design A	Dummy: Equals 1 if the application has design A, 0 otherwise	1216	0.361	-	0	1
Design B	Dummy: Equals 1 if the application has design B, 0 otherwise	1216	0.376	-	0	1
Design C	Dummy: Equals 1 if the application has design C, 0 otherwise	1216	0.263	-	0	1
Rank						
Rank 1	Dummy: Equals 1 if the application was sent out first, 0 otherwise	1216	0.500	-	0	1
Rank 2	Dummy: Equals 1 if the application was sent out second, 0 otherwise	1216	0.500	-	0	1
Certificate	Dummy: Equals 1 if the applicant provides an additional certificate, 0 otherwise	1216	0.375	-	0	1
Distance	Linear distance between applicant's home and location of employer (in km)	1216	286.25	116.87	22	553
Information on jobs and application period						
Application period						
May 2011	Dummy: Equals 1 if the application was sent out in May 2011, 0 otherwise	1216	0.405	-	0	1
Sep 2011	Dummy: Equals 1 if the application was sent out in September 2011, 0 otherwise	1216	0.429	-	0	1
May 2012	Dummy: Equals 1 if the application was sent out in May 2012, 0 otherwise	1216	0.166	-	0	1
Job						
Electronics technician	Dummy: Equals 1 if the candidate applies as an electronics technician, 0 otherwise	1216	0.150	-	0	1
Industrial mechanic	Dummy: Equals 1 if the candidate applies as an industrial mechanic, 0 otherwise	1216	0.313	-	0	1
Mechanic in plastics and rubber processing	Dummy: Equals 1 if the candidate applies as a mechanic in plastics and rubber processing, 0 otherwise	1216	0.178	-	0	1
Mechatronics fitter	Dummy: Equals 1 if the candidate applies as a mechatronics fitter, 0 otherwise	1216	0.211	-	0	1
Milling machine operator	Dummy: Equals 1 if the candidate applies as a milling machine operator, 0 otherwise	1216	0.150	-	0	1
Firm characteristics						
Size						
Small	Dummy: Equals 1 if the employer has less than 50 employees, 0 otherwise	1216	0.298	-	0	1
Medium	Dummy: Equals 1 if the employer has between 50 and 500 employees, 0 otherwise	1216	0.538	-	0	1
Large	Dummy: Equals 1 if the employer has more than 500 employees, 0 otherwise	1216	0.164	-	0	1

Location						
Other	Dummy: Equals 1 if the employer is not located in the South or East of Germany, 0 otherwise	1216	0.262	-	0	1
South	Dummy: Equals 1 if the employer is located in the South of Germany, 0 otherwise	1216	0.566	-	0	1
East	Dummy: Equals 1 if the employer is located in Eastern Germany, 0 otherwise	1216	0.173	-	0	1
Industry	Dummy: Equals 1 if the employer operates in the industry sector, 0 otherwise (i.e., service sector)	1216	0.900	-	0	1
Late recruiter	Dummy: Equals 1 if the employer recruits in May, 0 otherwise (i.e., September)	1216	0.571	-	0	1
Female responsible	Dummy: Equals 1 if the person responsible for recruiting as mentioned in the job offer is female, 0 otherwise	1216	0.508	-	0	1
Open positions	Number of open positions for an apprenticeship as indicated by the employer's job offer	1216	1.71	1.59	1	15
Labor market data						
Vacancies/total jobs t-1	Ratio of vacancies and total apprenticeships in the previous year (i.e., in the reporting period 2009/2010 and 2010/2011, respectively) and in the corresponding Employment Agency region of the employer	1216	0.046	0.027	0.004	0.139
Share of foreigners t-1	Share of foreign applicants in the previous year (i.e., in the reporting period 2009/2010 and 2010/2011, respectively) and in the corresponding Employment Agency region of the employer	1216	0.110	0.076	0.000	0.340

5.3.1.2 Comparison with the Overall Population of Training Companies

This section puts the dataset from the field experiment into perspective with the entire population of training companies in Germany. Table 5-25 shows that small employers are underrepresented relative to medium-sized firms. The reason for that may be the more frequent use of the job platform of the BA as a recruiting channel by the latter. Concerning companies' location, firms from the South of Germany are overrepresented in the sample. This may directly be linked to regional labor market constraints. As employers from the South experience fiercer competition for suitable apprentices, they probably use a multi-channel strategy (including the job platform of the BA) to publish their job offers and face longer staffing periods which both increasing the probability of being part of the sample.

Even though firm characteristics slightly differ between the current sample and the overall population, this should neither affect the generalizability of the results nor does it indicate firm selection. The latter would be an issue if firms advertising their jobs via the BA systematically differed from other companies.

Table 5-25: Firm Characteristics in Field Experiment and Entire Population of Training Companies

	Field experiment	Entire population of training companies
Size		
Small	29.77%	45.97%
Medium	53.78%	36.39%
Large	16.45%	17.64%
Location		
South	56.58%	45.32%
East	17.27%	17.60%
Other	26.15%	37.02%

Notes: Data on firm size as of 2010; data on location as a weighted average of 2010/2011 and 2011/2012.
Source: BA (2010a, 2011, 2012b), BIBB (2010a).

5.3.2 DESCRIPTIVE RESULTS

Regarding the hiring outcome, descriptive results indicate a preferential treatment of the applicant with the German-sounding name. Table 5-26 shows that while the German-named candidate received 257 callbacks (42.27 percent of all applications), the Turkish-named applicant was invited in 195 (32.07 percent) of all cases. This yields a difference of 10.20 percentage points which is statistically significant at the 1 percent level. Recalling that the correspondence method implements the ceteris paribus condition with respect to all other applicant characteristics, these findings indicate discrimination against the Turkish-named candidate.

Table 5-26: Firms' Detailed Responses by Name

	German name (N=608)	Turkish name (N=608)	Total	Difference
No response	19.57%	22.20%	20.89%	-2.63 pps
	(119)	(135)	(254)	(16)
Rejection	38.16%	45.72%	41.94%	-7.56 pps**
	(232)	(278)	(510)	(46)
Callback	42.27%	32.07%	37.17%	10.20 pps***
	(257)	(195)	(452)	(62)

Notes: The table reports detailed responses by name as a fraction of overall applications in percent. Absolute numbers are in parentheses. ** denotes 5% significance level and *** denotes 1% significance level of a chi-squared test (H_0: The German- and Turkish-named candidates are equally likely to receive a callback/a rejection at any matched-pair application).

Focusing on the importance of an additional certificate for the hiring outcome, the results indicate that both candidates equally benefit with an increase in callbacks of 8.16 percentage points and 8.33 percentage points (both statistically significant at the 5 percent level), respectively. Consequently, the extent of differential treatment remains constant and statistically significant (see table 5-27).

Table 5-27: Firms' Callbacks Conditional on the Provision of an Additional Certificate

	German name	Turkish name	Difference
No certificate	39.21%	28.95%	10.26 pps***
	(149/380)	(110/380)	
Certificate	47.37%	37.28%	10.09 pps**
	(108/228)	(85/228)	
Difference	8.16 pps**	8.33 pps**	

Notes: The table reports callbacks by name as a fraction of applications with and without an additional certificate in percent. Absolute numbers of callbacks and applications are in parentheses. ** denotes 5% and *** denotes 1% significance level of a chi-squared test (H_0: The German- and Turkish-named candidates are equally likely to receive a callback at any matched-pair application (in rows) and H_0: Applications with and without an additional certificate are equally likely to receive a callback (in columns), respectively).

Considering the different application periods and dividing the sample into late and early recruiters further reveals that discrimination seems to be somewhat higher if applications were dispatched in 'late' application periods (12.39 percentage points compared to 7.28 percentage points). However, in both cases the Turkish-named candidate received significantly fewer callbacks than the German-named counterpart (see table 5-28).

Table 5-28: Firms' Callbacks Conditional on Application Period

	German name	Turkish name	Difference
Late recruiters	42.36%	29.97%	12.39 pps***
	(147/347)	(104/147)	
Early recruiters	42.15%	34.87%	7.28 pps*
	(110/261)	(91/261)	

Notes: The table reports callbacks by name as a fraction of applications to late and early recruiters in percent. Absolute numbers of callbacks and applications are in parentheses. * denotes 10% and *** denotes 1% significance level of a chi-squared test (H_0: The German- and Turkish-named candidates are equally likely to receive a callback at any matched-pair application).

Table 5-29 displays the pairwise treatments by name, certificate, firm characteristics and labor market data rather than the aggregate outcomes. In column (1) the number of paired applications for each subsample is displayed. Column (2) shows the number of firms that neither replied nor rejected both of the applicants, leaving those employers that invited at least one of the candidates in column (3). The next three columns divide the firm-level observations from column (3) into cases of both-sided callbacks (column 4) and callbacks to either the German-named (column 5) or the Turkish-named applicant (column 6). Columns (7) and (8) calculate the callback rates, i.e., the share of callbacks among the total number of applications, for either candidate. Subtracting column (8) from column (7) yields the percentage points difference in callbacks (column (9)). Whether this difference is statistically different from zero is then tested by a standard chi-squared significance test (H_0: Callbacks to résumés with the German and Turkish name are equally distributed at any matched-pair application).

Table 5-29: Firms' Responses of Correspondence Testing by Name, Certificate, Firm Characteristics and Labor Market Data

	Firms' responses					Callback rates		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
No. of paired applications	Rejection/ no response	At least one callback	Both	Only German name	Only Turkish name	German name (4+5)/(1)	Turkish name (4+6)/(1)	Difference (7)-(8)
All firms (608)	55.26 (336)	44.74 (272)	66.18 (180)	28.31 (77)	5.51 (15)	0.423	0.321	0.102** (p=0.000)
Additional certificate								
None provides additional certificate (267)	60.30 (161)	39.70 (106)	67.92 (72)	28.30 (30)	3.77 (4)	0.382	0.285	0.097** (p=0.017)
Both provide additional certificate (115)	47.83 (55)	52.17 (60)	65.00 (39)	33.33 (20)	1.67 (1)	0.513	0.348	0.165** (p=0.011)
Only German-named candidate provides additional certificate (113)	53.10 (60)	46.90 (53)	56.60 (30)	35.85 (19)	7.55 (4)	0.434	0.301	0.133** (p=0.038)
Only Turkish-named candidate provides additional certificate (113)	53.10 (60)	46.90 (53)	73.58 (39)	15.09 (8)	11.32 (6)	0.416	0.398	0.018 (p=0.787)
Timing								
Late recruiter (347)	55.62 (193)	44.38 (154)	62.99 (97)	32.47 (50)	4.55 (7)	0.424	0.300	0.124*** (p=0.001)
Early recruiter (261)	54.79 (143)	45.21 (118)	70.34 (83)	22.88 (27)	6.78 (8)	0.421	0.349	0.073* (p=0.087)
Firm Size								
Small (<50) (181)	60.77 (110)	39.23 (71)	59.15 (42)	36.62 (26)	4.23 (3)	0.376	0.249	0.127*** (p=0.009)
Medium (50-500) (327)	53.21 (174)	46.79 (153)	66.67 (102)	28.76 (44)	4.58 (7)	0.446	0.333	0.113*** (p=0.003)
Large (>500) (100)	52.00 (52)	48.00 (48)	75.00 (36)	14.58 (7)	10.42 (5)	0.430	0.410	0.020 (p=0.774)
Location								
South (344)	58.43 (201)	41.57 (143)	60.14 (86)	31.47 (45)	8.39 (12)	0.381	0.285	0.096*** (p=0.008)
East (105)	51.43 (54)	48.57 (51)	76.47 (39)	21.57 (11)	1.96 (1)	0.476	0.381	0.095 (p=0.163)
Other (159)	50.94 (81)	49.06 (78)	70.51 (55)	26.92 (21)	2.56 (2)	0.478	0.358	0.119** (p=0.031)

			Sector					
Services	40.98	59.02	80.56	16.67	2.78	0.574	0.492	0.082
(61)	(25)	(36)	(29)	(6)	(1)			(p=0.364)
Industry	56.86	43.14	63.98	30.08	5.93	0.406	0.302	0.104***
(547)	(311)	(236)	(151)	(71)	(14)			(p=0.000)
			Person responsible for recruiting					
Male	58.28	41.72	57.85	40.50	1.65	0.410	0.248	0.162***
(290)	(169)	(121)	(70)	(49)	(2)			(p=0.000)
Female	51.17	48.83	73.29	18.49	8.22	0.448	0.398	0.050
(299)	(153)	(146)	(107)	(27)	(12)			(p=0.214)
			Share of foreigners t-1 (Mean=0.110)					
Above mean	55.00	45.00	58.12	33.33	8.55	0.412	0.300	0.112***
(260)	(143)	(117)	(68)	(39)	(10)			(p=0.009)
Below mean	55.46	44.54	72.26	24.52	3.23	0.431	0.336	0.095**
(348)	(193)	(155)	(112)	(38)	(5)			(p=0.011)
			Vacancies/total jobs t-1 (Mean=0.046)					
Above mean	61.98	38.02	66.30	25.00	8.70	0.347	0.285	0.062
(242)	(150)	(92)	(61)	(23)	(8)			(p=0.140)
Below mean	50.82	49.18	66.11	30.00	3.89	0.473	0.344	0.128***
(366)	(186)	(180)	(119)	(54)	(7)			(p=0.000)

Notes: This table shows the distribution of firms' responses. Absolute numbers are in parentheses. Column (1) displays the number of employers in each stratum. Column (2) reports the fraction of firms that gave none of the candidates a callback, so the remainder in column (3) called back at least one applicant. Firms that gave both candidates a positive answer, column (4), are considered as equal treatment, while the rest preferred either the German- or Turkish-named candidate (columns (5) and (6)). Columns (7) and (8) contain the callback rate for the German- and Turkish-named applicant, respectively, while column (9) computes the difference in callback rates between the two candidate groups. Person responsible for recruiting excludes those employers that did not name a recruiter in their job offers. In column (9), p-values of a chi-squared test that the German- and Turkish-named candidates are equally likely to receive a callback at any matched-pair application are in parentheses. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

In line with the descriptive results displayed above, table 5-29 shows that across the entire sample differential treatment occurred in 92 cases in which the majority candidate benefited the most (77 times). Dividing the overall callbacks of the German-named applicant by the overall callbacks of his Turkish-named counterpart gives a success ratio of 1.32 (=0.423/0.321). In other words, the minority candidate is 32 percent less likely to receive a callback. Testing the hypothesis that callbacks are equally distributed across groups reveals that the null hypothesis can be rejected at the 5 percent level. Given that the candidates are carefully matched, these findings can directly be interpreted as discrimination. However, the extent of discriminatory treatment obviously varies across different subsamples. In particular, the distribution of callbacks does not statistically differ by name in case that (i) only the Turkish-named candidate hands in an additional credential, (ii) the employer is of large size, (iii) the firm is located in the East of Germany, (iv) the company operates in the service sector, (v) the recruiter is female and (vi) the scarcity measure is above its mean. On the other hand, discrimination is most prominent if

(vii) both applicants provide an extra credential (difference: 16.5 percentage points), (viii) the employer is a late recruiter (12.4 percentage points), (ix) the company has less than 50 employees (12.7 percentage points), (x) the person responsible for recruiting is male (16.2 percentage points) and (xi) the labor market situation is relatively relaxed (12.8 percentage points).

Before turning to the multivariate analyses, more subtle forms of differential treatment are considered. Table 5-30 reports firms' responses by name. A gap in companies' response behavior would give a first impression of discriminatory treatment. Even if the counterpart was rejected (which would result in the same overall employment outcome), not replying at all would discourage the applicant from sending out further applications. Regarding the descriptive results, no such differences can be found in the current sample. More precisely, the null hypothesis that firms' responses are equally distributed across names cannot be rejected.

Table 5-30: Firms' Responses by Name

	German name	Turkish name	Total	Difference
No response	19.57%	22.20%	20.89%	-2.63 pp.
	(119)	(135)	(254)	(16)
Response	80.43%	77.80%	79.11%	2.63 pps
	(489)	(473)	(962)	(16)

Notes: The table reports employers' responses by name as a fraction of overall applications in percent. Absolute numbers are in parentheses.

However, probit regressions on the response dummy with standard errors clustered on firm level suggest that the response probability is negatively correlated with the Turkish name dummy. The point estimate shows a 2.8 to 2.9 percentage points difference that is statistically significant at the 10 percent level and robust to various model specifications (see table C-8 in the appendix). On the one hand, this might be a first indicator of callback differences. On the other hand, though, it may leave the gap in callback rates unaffected as the majority candidate might still receive a rejection instead. Either way, the fact that firms' response behavior at least partly accounts for different invitation probabilities across the two demographic groups cannot completely ruled out.

In the same vein as the response behavior, cases in which one candidate receives a callback only after the other candidate has rejected the invitation can be considered another form of the so called "equal but different treatment". This phenomenon can be found in about one quarter of all cases of mutual callbacks, but benefits both applicant groups equally (see table 5-31).

Table 5-31: Firms' Callbacks only after the Counterpart Has Declined an Invitation

Callbacks…	**Fraction (Absolute number)**
… to both candidates	100.00% (180)
… to the German-named candidate only after the Turkish-named candidate has declined an invitation	14.44% (26)
… to the Turkish-named candidate only after the German-named candidate has declined an invitation	12.22% (22)

Notes: The table reports cases of equal but different treatment by name as a fraction of mutual callbacks. Absolute numbers are in parentheses.

Moreover, table 5-32 reports average reaction times, i.e., the time until the candidate either receives a callback or a rejection by the employer. The reason for a variation in reaction times might be twofold: companies either gather applications to be able to select from a larger pool of job candidates or they simply postpone their decision on purpose hoping that inadequate applicants withdraw. However, mean comparison tests of callback and rejection times do not reveal significant differences by group. In case of the former, it took the companies on average 18.3 days until the candidates were informed whereas rejections were sent out after 31.5 days. Longer callback times for medium and large corporations can be attributed to the fact that more recruiters are involved in the decision process, that more vacancies have to be filled and that the number of incoming applications is larger than in companies with less than 50 employees. Furthermore, the fraction of medium and large firms is higher among early recruiters (see table 5-23) which generally dedicate more time to decision making.

Table 5-32: Average Callback and Rejection Times in Working Days by Name

	Callback			Rejection		
	German name	Turkish name	Average	German name	Turkish name	Average
All	17.9	18.6	18.3	30.6	32.4	31.5
Small	14.0	11.0	12.5	23.3	25.2	24.3
Medium	19.0	21.1	20.0	31.7	34.7	33.2
Large	20.7	20.4	20.6	37.6	36.0	36.8

In order to provide further evidence for the reasons of ethnic discrimination, various probit estimations are conducted to disentangle the effects that originate from differences in the provision of certificates as well as firm and labor market characteristics.

5.3.3 ECONOMETRIC ANALYSES

The following section presents the empirical model (5.3.3.1) which is used for the subsequently performed econometric analyses (5.3.3.2).

5.3.3.1 EMPIRICAL MODEL

As the dependent variable (the callback dummy) is binary, the linearity assumption of the OLS method would be violated. Consequently, an alternative estimation technique based on a probabilistic distribution function is required. Probit regressions have, inter alia, proven to account for the nonlinear relationship between the covariates and the outcome variable and produce plausible results. Transforming the estimation coefficients into marginal effects further facilitates the interpretation of these results. The baseline model estimated below looks as follows:

$$
\begin{aligned}
\mathrm{P_i(Callback)} = \\
&\beta_0 + \beta_1 \text{Turkish Name}_i + \beta_2 \overrightarrow{\text{Firm characteristics}_i} + \\
&\beta_3 \text{ Share of foreigners t} - 1_i + \beta_4 \text{ Vacancies/total jobs t} - 1_i + \\
&\beta_5 \text{ Certificate}_i + \beta_6 \overrightarrow{\text{Controls}_i} + \mu_i,
\end{aligned}
$$

where β_0 is a constant, β_k denotes the regression coefficient β of regressor k and μ_i represents a normally distributed error term of applicant i. The name and the certificate dummy as well as company, job and framework controls serve as further independent variables (see table 5-24). Firm characteristics include size, location, industry and recruiter type. The vector of control variables includes the year the apprenticeship starts, the number of open positions, the distance between the applicant's home and the workplace, as well as dummies for dispatching order and résumé design.

5.3.3.2 PROBIT REGRESSIONS AND HYPOTHESES TESTING

Table 5-33 reports average marginal effects from a probit regression on the callback dummy together with their standard errors clustered on firm level. Model (I) only displays the effect of the Turkish name dummy, model (II) additionally accounts for firm characteristics, model (III) adds standardized labor market variables and model (IV) incorporates the certificate dummy. All models include the set of control variables as described above and use the entire sample, i.e., all 1,216 observations.

Table 5-33: Marginal Effects from Probit Regressions on Callback Dummy

Callback	(I)	(II)	(III)	(IV)
Turkish name	-0.108***	-0.110***	-0.110***	-0.109***
	(0.016)	(0.016)	(0.016)	(0.016)
Medium		0.077*	0.076*	0.073
		(0.044)	(0.044)	(0.045)
Large		0.086	0.084	0.079
		(0.066)	(0.066)	(0.067)
South		-0.045	-0.032	-0.032
		(0.058)	(0.059)	(0.059)
East		0.019	0.036	0.032
		(0.060)	(0.065)	(0.065)
Industry		-0.162**	-0.168***	-0.173***
		(0.063)	(0.064)	(0.064)
Late recruiter		0.078	0.084	0.091*
		(0.054)	(0.054)	(0.054)
Female responsible		0.082**	0.082**	0.083**
		(0.038)	(0.038)	(0.038)
Share of foreigners t-1			0.002	0.001
			(0.022)	(0.022)
Vacancies/total jobs t-1			-0.025	-0.025
			(0.022)	(0.022)
Certificate				0.077**
				(0.034)
Controls	Yes	Yes	Yes	Yes
No. of obs.	1,216	1,216	1,216	1,216
Pseudo R^2	0.023	0.044	0.045	0.048
Log likelihood	-783.842	-767.369	-766.136	-764.143
Wald chi-squared	58.024	76.345	78.194	81.306
P-value	0.000	0.000	0.000	0.000

Notes: Each model reports average marginal effects of a probit regression on the callback dummy (Y=1: employer calls back the job applicant). Marginal effects are calculated at the mean of all independent variables and denote an infinitesimal change in case of continuous variables and a discrete change in case of dummy variables. Standard errors clustered on firm level are in parentheses. Regressions consider the entire sample. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

Regression results of the name dummy support the overall findings from section 5.3.2 and lend support to '$H_{minority}$'. The applicant with the Turkish-sounding name has a 10.8-11.0 percentage points lower callback probability compared to his German-named counterpart. This effect is statistically significant at the 1 percent level and robust across all model specifications. Moreover, the influence of the name dummy remains almost unaffected if calculated at the mode rather than the mean of all other categorical covariates (see table C-9 in the appendix), that is, for a standard applicant at a standard employer the coefficients vary between -0.108 and -0.116. Tables C-10 and C-11 in the appendix further demonstrate that the effect of the Turkish name dummy is independent of any confounds that are based on different names and photos. The coefficients of the alternative name ('Jan Lange' versus 'Lukas Schmidt' and 'Kenan Yilmaz' versus 'Onur Öztürk') and photo ('Photo A' versus 'Photo B') dummies turn out to be insignificant for either demographic group. Lower callbacks can thus only be attributed to the candidate's ethnicity.

Concerning firm characteristics, regression results show weak evidence for medium-sized employers recruiting the job candidates significantly more often in comparison to small-sized firms. The reason for that might be that small firms have less formalized decision processes and therefore tend to recruit people who have been recommended by coworkers or who have already worked for the company (e.g. during a school internships or summer vacation). In addition, table 5-33 reveals that applications sent out to firms operating in the manufacturing sector on average yield 17 percentage points lower callbacks. Across the model specifications, this effect is statistically significant at the 1 and 5 percent level, respectively, and might account for the fact that graduates interested in technical apprenticeships rather focus on the industry sector which increases the number of applications and, consequently, competition among applicants. Alternatively, firms in the service sector might simply invite a higher fraction from their pool of applicants in order to screen their service orientation in a face-to-face interview. If that were the case, hiring probabilities across both sectors would converge over all stages of the recruitment process which, unfortunately, cannot be investigated with data from this study. Moreover, if a woman is responsible for recruiting, the overall callback probability increases by 8 percentage points. This effect is robust, but does not allow a causal interpretation since the researcher cannot observe whether other recruiters were involved in the decision-making processes. Finally, the inclusion of the certificate dummy in model (IV) highlights the beneficial effect of the provision of additional productivity relevant information. If an additional credential is attached, employers respond with a 7.7 percentage points higher callback rate that is statistically different from zero at the 5 percent level.

With respect to the GoF measures, all model specifications predict the outcome variable better than the intercept model. However, similar to the study on gender discrimination, the pseudo R^2 is rather low which can be attributed to the ceteris paribus condition of the correspondence method, i.e., the fact that apart from firm and labor market characteristics only applicants' names as a proxy for ethnic background differ.

Even though the findings from above provide evidence that ethnic discrimination in technical occupations seems to persist, no conclusions on the sources of differential treatment can be derived. Therefore, table 5-34 investigates whether the name dummy interacts with the covariates as mentioned in the hypotheses section. The model specifications yield average marginal effects at the mean of all other independent variables. Model (I) only includes point estimates, models (IIa) to (IId) interact the Turkish name dummy with either covariate and model (III) additionally tests the joint effects. The full regression table with and without control variables can be found in the appendix (table C-12).

Table 5-34: Marginal Effects from Probit Regressions on Callback Dummy and Hypotheses Testing

Callback	(I)	(IIa)	(IIb)	(IIc)	(IId)	(III)
Turkish name	-0.109***	-0.117***	-0.110***	-0.079***	-0.109***	-0.070
	(0.016)	(0.025)	(0.016)	(0.023)	(0.016)	(0.053)
Certificate	0.077**	0.067	0.077**	0.077**	0.076**	0.083*
	(0.034)	(0.043)	(0.034)	(0.033)	(0.034)	(0.050)
Turkish name x		0.021				-0.013
Certificate		(0.053)				(0.070)
Share of foreigners t-1	0.001	0.001	0.014	0.001	0.001	0.016
	(0.022)	(0.022)	(0.024)	(0.022)	(0.022)	(0.024)
Turkish name x			-0.027			-0.031*
Share of foreigners t-1			(0.018)			(0.018)
Late recruiter	0.091*	0.091*	0.091*	0.117**	0.091*	0.121**
	(0.054)	(0.054)	(0.054)	(0.056)	(0.054)	(0.060)
Turkish name x				-0.052*		-0.060
Late recruiter				(0.031)		(0.049)
Vacancies/total jobs t-1	-0.025	-0.025	-0.025	-0.025	-0.033	-0.034
	(0.022)	(0.022)	(0.022)	(0.022)	(0.023)	(0.023)
Turkish name x					0.017	0.020
Vacancies/total jobs t-1					(0.015)	(0.015)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
No. of obs.	1,216	1,216	1,216	1,216	1,216	1,216
Pseudo R^2	0.048	0.048	0.048	0.048	0.048	0.049
Log likelihood	-764.143	-764.080	-763.698	-763.729	-763.960	-762.972
Wald chi-squared	81.306	81.789	80.762	81.164	83.031	82.739
P-value	0.000	0.000	0.000	0.000	0.000	0.000

Notes: Each model reports average marginal effects of a probit regression on the callback dummy (Y=1: employer calls back the job applicant). Marginal effects are calculated at the mean of all independent variables and denote an infinitesimal change in case of continuous variables and a discrete change in case of dummy variables. Standard errors clustered on firm level are in parentheses. Regressions consider the entire sample. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

As model (IIa) indicates, the interaction between the Turkish name and the certificate dummy does not significantly increase the minority candidates' callback probability relative to his majority counterpart. In contrast to '$H_{certificate}$', but in line with the descriptive findings from above, the provision of a certified internship equally benefits both applicants. As a result, the gap in callbacks is not reduced by this additional ability signal. Even for different values of the predicted callback probability, the interaction term remains statistically insignificant which underlines the absence of the postulated effect (see figure C-5 in the appendix).

According to '$H_{share\ of\ foreigners}$', previous contact with other members of a group increases employers' ability of predicting future productivity. Consequently, a higher share of foreign employees should increase the likelihood of a callback for job candidates with a migration background. However, model (IIb) does not support this assumption since the interaction effect between the Turkish name and the share of foreigners in t-1 turns out to be insignificant while the point estimate of the name dummy remains unchanged. Figure C-6 in the appendix further shows that the significance level of the interaction effect is independent of different combinations of other independent variables included in the model.

Focusing on different hiring behavior between late and early recruiters shows that, similar to the gender study, the former tend to discriminate somewhat more which contradicts 'H_{timing}'. While the positive point estimate of the late recruiter variable, i.e., the callback probability of the German-named candidate for applications sent out in May, increases, the interaction term becomes negative and statistically significant at the 10 percent level. Thus, in addition to the negative point estimate of the name dummy, the minority applicant has a 5.2 percentage points lower chance of being called back from late recruiters than the majority candidate. However, recruiter type does not fully explain the callback gap as the Turkish name coefficient remains statistically significant. This means that even early recruiters discriminate against the minority candidate.

Model (IId) considers the joint effect that stems from labor market scarcity. Contrasting the corresponding hypothesis and preliminary descriptive results (see table 5-29), the regression estimates do not indicate any statistically significant relationship between the scarcity measure and the name dummy. This is also supported by figure C-8 in the appendix which displays the predicted probability at different points of the probability distribution. Thus, '$H_{scarcity}$' can be rejected.

Overall, the findings do not support any of the hypotheses reflecting statistical and taste-based discrimination. Instead, a rather weak late-recruiter effect can be found which, in contrast to the timing hypothesis, turns out to be significantly negative. Reasons for these ambiguous results as well as alternative explanations will be discussed in the next section.

5.3.4 Discussion

In the following, the main results presented above will be discussed while additional estimates and references to the existing empirical literature are used to put the findings into perspective.

5.3.4.1 Relation to Prior Findings

The findings on ethnic discrimination mainly support results from previous correspondence and audit studies showing that ethnic minorities are systematically disadvantaged with respect to access to employment (e.g. Riach and Rich, 2002). Here, the German applicant, on average, can expect 4 callbacks for every 10 applications whereas his Turkish-named counterpart only receives 3 positive responses for every 10 attempts. The average callback differential oscillates around 10 percentage points and thus falls into the lower range of what other researchers have reported so far (3 to 43 percentage points; see table A-2 in the appendix). However, if the focus is restricted to ethnic Turks, the ethnic penalty found in the present context is located at the upper end. While prior evidence from Belgium and the Netherlands suggests that Turkish immigrants have a 7 to 11 percentage points lower callback probability than observationally similar natives (Andriessen et al., 2012; Baert et al., 2013), callback gaps found in the German labor market are somewhat smaller. Goldberg et al. (1996) on average find a 1 pps gap between first generation Turkish immigrants and native Germans whereas Kaas and Manger (2012) report a 5 percentage points gap between second generation Turks and their German counterparts. In fact, the results indicate that the extent of differential treatment turns out to be higher in labor market segments where employees are on average less qualified. In other words, minority apprenticeship applicants seem to suffer more than e.g. business and economics students that were used as job candidates in the Kaas and Manger (2012) study.

Qualitatively, the findings from present and prior research support what has explicitly been tested in a matched-pair experiment by Carlsson (2010). That is, hiring discrimination persists for first and second generation immigrants. However, drawing any conclusions from the treatment of Turks to other ethnic minorities can only be speculative. Former studies suggest that compared to other immigrant groups Turks suffer most with respect to both hiring probabilities and wages (e.g. BIBB, 2006; Uhlendorff and Zimmermann, 2006; Lehmer and Ludsteck, 2011). Hence, the results presented may rather overestimate the actual effect of discrimination faced by the entire population with migration background. Still, the findings may explain some of the stylized facts on native-immigrant labor market differences, in particular occupational segregation and the gap in (youth) unemployment rates. Furthermore, firms' discriminatory behavior may have caused the share of foreigners participating in dual training to decrease over the last decade. Not surprisingly, this reduction has been most noticeable in technical and industry apprenticeships such as electronic technician, mechatronic and industrial mechanic, all of which have been addressed in the present field experiment (BIBB, 2006).

Table 5-35: Marginal Effects from Probit Regressions on Callback Dummy and Interaction of Turkish Name Dummy and Firm Characteristics

Callback	(I)
Turkish name x Medium	-0.002
	(0.041)
Turkish name x Large	0.080
	(0.053)
Turkish name x South	0.013
	(0.036)
Turkish name x East	0.040
	(0.047)
Turkish name x Industry	-0.026
	(0.047)
Turkish name x Female responsible	0.110***
	(0.034)
Turkish name x Late recruiter	-0.039
	(0.033)
Controls	Yes
No. of obs.	1,216
Pseudo R^2	0.052
Log likelihood	-760.980
Wald chi-squared	90.103
P-value	0.000

Notes: The model reports average marginal effects of a probit regression on the callback dummy (Y=1: employer calls back the job applicant) for the entire sample. Marginal effects are calculated at the mean of all independent variables. Standard errors clustered on firm level are in parentheses. Controls include all point estimates of the variables interacted. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

Table 5-35 compares whether ethnic discrimination varies with respect to employer characteristics. In particular, interactions between the Turkish name and firm dummies are tested. The only effect that turns out to be statistically significant originates from recruiters' sex. In line with the findings from e.g. Carlsson and Rooth (2007) and Carlsson (2010), the minority candidate has a ceteris paribus 11 percentage points higher callback probability if the person responsible for administrating incoming applications is female. Put differently, male recruiters tend to discriminate more. However, as has already been noted in section 5.2.3.3, the sex of actual decision makers is unobservable so that a causal relationship can only be assumed.

5.3.4.2 GROUP EXPERIENCE AND THE ROLE OF ADDITIONAL SIGNALS

The regression estimates presented in table 5-34 indicate that both '$H_{certificate}$' and '$H_{share\ of\ females}$' need to be rejected. This may imply three possible explanations, i.e., (i) statistical discrimination does indeed not affect employers' rationale to treat majority and minority applicants differently, (ii) the operationalization does not adequately reflect group differences in asymmetric information and (iii) the information provided helps sufficiently assessing the candidates' future productivity and thus already captures the effect originating from statistical discrimination. Explanation (iii) can be supported by looking at what applications in Germany generally include. Unlike in most other countries, it is obligatory to attach school certificates when officially getting in touch with an employer for the first time. In the U.S., for example, such credentials are normally handed in at a later stage of the recruitment process (see previous correspondence studies presented in chapter 3). In case of labor market entrants, however, school certificates serve as a very strong and credible signal which, from an employer's perspective, leads to a reduction of information asymmetries. The larger this reduction, the lower are employers' perceived group differences in unobserved productivity. Consequently, any other variables proxying statistical discrimination become insignificant.

Another argument concerns the operationalization. It assumes that room for statistical discrimination exists even in the presence of school certificates. No matter whether these credentials reduce asymmetric information or not, minority applicants are still significantly disadvantaged if employers are not equipped with further devices (such as reference letters) that help assessing applicants' productivity. Yet, both the share of foreign applicants in t-1 as well as the inclusion of extra credentials may simply not serve as adequate devices in the context of apprenticeship applications. Concerning the former, employers may not care about whom they have evaluated in previous recruiting processes as is denoted by the variable 'share of foreign applicants in t-1', but use personal work experience with members of a group to proxy future performance of an applicant who belongs to that same group. Thus, the share of minority workers employed by the firm addressed in the field experiment might have led to a better understanding of whether differences in group experience affect employment outcomes. Unfortunately, no such data were available and, hence, could not be matched with job offers.

The analysis further indicates that the provision of an additional credential does not reduce the gap in callbacks. This is somewhat in contrast to the results by Kaas and Manger (2012). They show that the Turkish-named candidate on average has a 14 percent lower callback probability compared to his German-named counterpart, but that differential treatment becomes insignificant if reference letters by university professors are attached. Interestingly, the provision of these references leaves callbacks to the majority candidate unaffected while the minority applicant significantly benefits. The latter obviously has to present more credentials to signal the same productivity. This can be interpreted as evidence for statistical discrimination (see also Heilman, 1984; Biernat and Kobrynowicz, 1997). Other studies, however, challenge these results. Among others, Bertrand and Mullainathan (2004) show that blacks realize inferior returns to skills as opposed to whites as callback differences increase if high-quality résumés are dispatched. Now, employers' responses in the present study indicate a beneficial effect of extra credentials, but do not reveal group differences in their returns (see model (IIa) of table 5-34 as well as tables C-10 and C-11 in the appendix). As a consequence, the callback gap persists and '$H_{certificate}$' cannot be supported. This, of course, does not rule out the possibility that additional productivity signals lead to a decrease of the callback differential in other labor market segments where e.g. evaluations by former employers provide more information on applicants' abilities.[74]

5.3.4.3 Labor Market Scarcity and Recruiter Effects

As model (IId) in table 5-34 demonstrates, labor market scarcity reflected by the fraction of vacancies among all apprenticeships offered in t-1 does not affect the extent of ethnic discrimination. In other words, employers do not discriminate significantly less if they are confronted with competition for suitable job candidates and are therefore willing to incur extra costs due to increased search activities and forgone productivity potentials. Previous research provides evidence that employers indeed respond to labor market tightness by changing callback behavior, in particular in favor of ethnic minorities (Kalter, 2002; Baert et al., 2013). Yet, these findings refer to the occupational and qualificatory rather than the regional labor-supply structure. The former two cannot be assessed in the present context since neither jobs addressed nor applicants' résumé quality (except for additional

[74] Other effects associated with the provision of extra credentials have been tested, but found to be insignificant. Zibrowius (2012), for instance, finds that returns to skills are largest where the share of immigrants is lowest. Interacting the certificate dummy with the share of foreign applicants in t-1, however, does not yield different effects by demographic groups (results not displayed but available upon request).

credentials) produce sufficient variation. With regard to regional scarcity, previous empirical evidence highlights that the level of employers' prejudice differs contingent on societal attitudes. However, this somewhat contrasts with the idea that employers reveal their true tastes only if they face economic pressure in terms of labor market competition for talents. The lack of statistically significant results originating from the scarcity measure may indicate that taste discrimination either is absent or that alternative proxies (some of which have been tested but neither proved to be statistically significant) are required.

In case preference-based discrimination persists, the assumption that it originates from customers' distastes can be neglected for two reasons. On the one hand, apprentices in technical occupations do hardly get in touch with customers and, on the other hand, differential treatment is not statistically significant in the service sector where customer contact is most likely. Regarding the impact of the remaining two forms, i.e., employer and coworker discrimination, however, the data do not allow an unambiguous distinction. This point is thus left open for future research.

Table 5-36: Marginal Effects from Probit Regressions on Callback Dummy with Sample Split by Recruiter Type

Callback	**(Ia)**	**(Ib)**	**(IIa)**	**(IIb)**
Turkish name	-0.073***	-0.104***	-0.124***	-0.130***
	(0.022)	(0.027)	(0.021)	(0.022)
Certificate	No	Yes	No	Yes
Foreigners/total applicants t-1	No	Yes	No	Yes
Vacancies/total jobs t-1	No	Yes	No	Yes
Controls	No	Yes	No	Yes
No. of obs.	522	522	694	694
Pseudo R^2	0.004	0.111	0.013	0.048
Log likelihood	-346.444	-309.127	-448.344	-432.460
Wald chi-squared	10.646	42.830	35.191	51.666
P-value	0.001	0.000	0.000	0.000

Notes: Each model reports average marginal effects of a probit regression on the callback dummy (Y=1: employer calls back the job applicant). Marginal effects are calculated at the means of all independent variables and denote an infinitesimal change in case of continuous variables and a discrete change in case of dummy variables. Standard errors clustered on firm level are in parentheses. Models (Ia) and (Ib) consider early recruiter sample, models (IIa) and (IIb) late recruiter sample. Either model includes only male-dominated jobs. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

Lastly, the analyses indicate that recruiter type at least weakly affects callback differentials. However, the estimated effect contradicts what has been hypothesized by 'H_{timing}'. As indicated by model (IIc) in table 5-34, the gap in callback rates is 5.2 percentage points higher if applicants address late recruiters. Yet, unlike in the experiment on gender discrimination, the negative and statistically significant interaction does not cause the effect of the Turkish name dummy to become insignificant. In other words, discrimination can also be found among early recruiters. This can further be demonstrated by splitting the sample across recruiter types (see table 5-36). While average marginal effects of the ethnicity dummy vary between 7.3 and 10.4 percentage points in the early-recruiter sample (model (Ia) and (Ib)), they range from 12.4 to 13.0 percentage points if only late recruiters are considered (model (IIa) and (IIb)).

Again, a plausible explanation for the late-recruiter effect may be based on systematic differences in management quality or, more specifically, in recruitment practices. Table 5-37 tries to capture these differences by conducting two separate probit regressions on (i) the probability that the applicant receives any response on behalf of the employer and (ii) the probability that the employer reacts after being reminded conditional on that he has responded at all. If, ceteris paribus, the late recruiter dummy turns out to be statistically significant in any of these specifications, at least some evidence on the management quality argument is provided. In fact, it seems that late recruiters lack proficiency in administrating applications. They are 15.3 percentage points more likely to postpone any reaction unless the job candidate inquires. Recruiter type thus somehow acts as a proxy for management quality which in turn seems to affect the extent of ethnic discrimination.

Table 5-37: Marginal Effects from Probit Regressions on Response and Reaction to Reminder Dummy

	(Response)	**(Reaction to reminder)**
Late recruiter	0.005	0.153***
	(0.032)	(0.038)
Firm characteristics	Yes	Yes
No. of obs.	1,216	962
Pseudo R^2	0.054	0.064
Log likelihood	-589.430	-527.519
Wald chi-squared	42.912	47.930
P-value	0.000	0.000

Notes: Table reports average marginal effects of a probit regression on the response (Y=1: applicant receives a response on behalf of the employer) and reacting to reminder (Y=1: Firm responds only after being reminded given that a firm responds at all) dummy. Standard errors clustered on firm level are in parentheses. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

Overall, support for the postulated hypothesis from the empirical analyses is rather poor. Apart from a weak recruiter effect, taste-based and statistical discrimination do not seem to deliver further insights into the systematic patterns of ethnic hiring discrimination.

5.3.4.4 The Role of Societal Attitudes

Perceptions of the role of ethnic minorities in the labor market and in society may vary across regions. People living in the Eastern federal states and rural areas, for example, are said to be more prejudiced towards foreigners and fellow citizens with migration background. Sociological and psychological approaches assume that tastes prevailing in society may shape employers' attitudes and, as a result, their recruiting behavior (Charles and Guryan, 2008). Previous research links employers' implicit attitudes as well as differences in a population's explicit (i.e., revealed) attitudes to ethnic discrimination. Recall that the study by Rooth (2010) finds a 5 percentage points decrease in the minority candidate's callback probability with a one standard deviation increase in recruiters' implicit association test score. Moreover, Carlsson and Rooth (2011) merge results from a social survey on attitudes towards ethnic minorities with data from a correspondence test. They show that regional variations in people's opinions on these minorities affect hiring probabilities of Middle Eastern-named job candidates significantly.

To reflect and quantify regional differences in tastes in the course of the present study, voting results from the last German Federal Elections in 2009 are used. Fortunately, these results can be broken down to areas in which the firms addressed by the applicants are located. The parties involved in the election represent different attitudes towards ethnic minorities. In this respect, the electorates of the major parties do not substantially differ from each other. Some parties may be considered as more devoted to issues on integration, but, in general, all of them have tried to establish a foreigner-friendly culture in Germany in the recent past. However, one exception known beyond regional levels remains. The National Democratic Party of Germany (NPD) is a neo-fascist party which, in a nutshell, means that they encounter ethnic minorities with extreme prejudice. The share of votes assigned to the NPD may thus be considered a proxy for regional distastes. If these distastes affect employers' recruiting decisions, the extent of ethnic discrimination should increase with the share of NPD votes. The respective percentage averages 1.86 percent and ranges from 0 to 5.8 percent in the sample.

Table 5-38 shows average marginal effects of probit regressions on the callback dummy. Models (Ia) and (Ib) add an interaction between the Turkish name dummy and the share of NPD votes excluding and including firm and labor market controls, respectively. In turn, models (IIa) and (IIb) include an interaction between the name dummy and a dummy that equals one if the share of NPD votes exceeds its average. Again, the former does not include controls while the latter does. Surprisingly, the results indicate the opposite to what has been expected. In the first two models, only the name dummy turns out to be statistically significant. However, in models (IIa) and (IIb) also the interaction effect is positive and statistically significant. Depending on the model specification, the minority candidate has an 8.6 to 10.6 higher callback probability in regions where the share of NPD votes exceeds the sample average. Even though differential treatment remains, unlike expected, the callback gap is substantially reduced in potentially less foreigner-friendly areas. This effect persists even if labor market scarcity and the share of foreign applicants are controlled for (see model (IIb)). Hence, the present findings seem to contradict the results by Rooth (2010) and Carlsson and Rooth (2011) and suggest that societal attitudes proxied by electoral results foster a convergence rather than a divergence of the majority-minority hiring gap.

Table 5-38: Marginal Effects from Probit Regressions on Callback Dummy and Interaction of Name Dummy and Share of NPD Votes

Callback	(Ia)	(Ib)	(IIa)	(IIb)
Turkish name	-0.130***	-0.141***	-0.137***	-0.153***
	(0.033)	(0.035)	(0.024)	(0.025)
Share NPD votes	-0.015	-0.044	-0.023	-0.053*
	(0.020)	(0.031)	(0.020)	(0.031)
Turkish name x	0.015	0.017		
Share NPD votes	(0.015)	(0.016)		
Turkish name x			0.086**	0.106**
Share NPD votes above average			(0.041)	(0.042)
Controls	No	Yes	No	Yes
No. of obs.	1,216	1,216	1,216	1,216
Pseudo R^2	0.009	0.049	0.011	0.052
Log likelihood	-795.331	-762.751	-793.874	-760.695
Wald chi-squared	44.721	83.312	46.789	88.473
P-value	0.000	0.000	0.000	0.000

Notes: Table reports average marginal effects of a probit regression on the callback dummy (Y=1: employer calls back the job applicant). Marginal effects are calculated at the mean of all independent variables and denote an infinitesimal change in case of continuous variables and a discrete change in case of dummy variables. Standard errors clustered on firm level are in parentheses. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

5.4 METHODOLOGICAL VARIATIONS

This section focuses on the effect of methodological variations, i.e., dispatching only a single versus matched-pair applications, on aggregate response and callback rates. In particular, it is tested whether competition in correspondence testing systematically leads to different hiring outcomes for the majority and minority candidates. Such a comparison also enables the researcher to fully exclude any bias from deception on behalf of employers which, on the one hand, would result in significantly lower callback rates in case of the correspondence method and, on the other hand, would underestimate the extent of discrimination against the minority candidate as a higher fraction of employers would treat the candidates equally.

Therefore, in the last application period (May 2012) in both the study on gender and ethnic discrimination not only paired applications were dispatched, but the same set of résumés was also sent out individually. The latter is subsequently referred to as the 'single application method' while the former is either called 'pairwise application' or 'correspondence method'. Table C-14 in the appendix shows the descriptive statistics of the method comparison in the gender study. Apart from the use of the correspondence approach where 149 employers were addressed, the male and the female candidate applied individually in 73 cases resulting in an overall number of 444 single applications. Put differently, around 67 percent of companies' responses were generated within the correspondence setting while the remaining 33 percent arose from single applications. All in all, the candidates received a response in 80 percent and were promoted to the next stage of the recruitment process in around 40 percent of the cases. Analogously to the study presented in section 5.2, men and women dispatched their applications equally often. Apart from that, it has to be noted that the majority of the jobs considered here can be classified as female-dominated. Hence, the results from above suggest that no systematic differences between the two candidate groups should be expected.

In a similar way as the dataset generated by the résumés for the study on gender discrimination, the data for the method comparison with respect to ethnic discrimination were collected. In addition to the 101 cases where employers received an application from both candidates, in, respectively, 49 and 51 cases either the German- or the Turkish-named candidate applied. Thus, overall, 302 applications were dispatched of which around 80 percent ended with a response and 45.4 percent with a callback. As indicated by the correspondence variable, two thirds of the responses originate from pairwise application testing while one third goes back to the single application method (see table C-15 in the appendix).

For either subsample, expectations are very similar, i.e., overall response and callback rates should not differ conditional on the method chosen. In the same vein, results on gender and ethnic discrimination should neither qualitatively nor quantitatively vary. If they do, it cannot be excluded that the application method impacts on differential treatment.

Next, the aggregate results from the two methodological approaches are compared for both datasets. Table 5-39 reveals no statistically significant differences between the single and pairwise application method for any response type, i.e., no response (3.80 percentage points), rejection (3.24 percentage points) and callback (0.56 percentage points). Moreover, the differences between the methods separated by gender do also not turn out to be statistically significant. The same holds true for a comparison in the context of the ethnicity study (see table 5-40). Again, chi-squared tests on method-specific outcome differences indicate that neither the overall results nor employers' responses by name do significantly differ as a function of the application method.

Table 5-39: Firms' Responses by Method and Gender

	Single application (N=146)		Pairwise application (N=298)		Total (N=444)		Differences between methods	
	Male (N=73)	Female (N=73)	Male (N=149)	Female (N=149)	Male (N=222)	Female (N=222)	Male	Female
No response	21.92		18.12		19.37		3.80 pps	
	27.40	16.44	18.79	17.45	21.62	17.12	8.61 pps	-1.01 pps
Rejection	39.04		42.28		41.22		-3.24 pps	
	35.62	42.47	40.94	43.62	39.19	43.24	-5.32 pps	-1.15 pps
Callback	39.04		39.60		39.41		-0.56 pps	
	36.99	41.10	40.27	38.93	39.19	39.64	-3.28 pps	2.17 pps

Notes: The table reports detailed responses by method and gender as a fraction of the overall number of applications in percent. Overall as well as gender-specific differences between methods are reported in percentage points. Chi-squared tests cannot reject the null hypothesis (H_0: The single and pairwise application methods are equally likely to produce a case of no response, rejection and callback, respectively).

Table 5-40: Firms' Responses by Method and Name

	Single application (N=100)		Pairwise application (N=202)		Total (N=302)		Differences between methods	
	German name (N=49)	Turkish name (N=51)	German name (N=101)	Turkish name (N=101)	German name (N=150)	Turkish name (N=152)	German name	Turkish name
No response	16.00		21.78		19.87		-5.78 pps	
	10.20	21.57	17.82	25.74	15.33	24.34	-7.62 pps	-4.17 pps
Rejection	36.00		34.16		34.77		1.84 pps	
	34.69	37.25	29.70	38.61	31.33	38.16	4.99 pps	-1.36 pps
Callback	48.00		44.06		45.36		3.94 pps	
	55.10	41.18	52.48	35.64	53.33	37.50	2.62 pps	5.54 pps

Notes: The table reports detailed responses by method and name as a fraction of the overall number of applications in percent. Overall as well as gender-specific differences between methods are reported in percentage points. Chi-squared tests cannot reject the null hypothesis (H_0: The single and pairwise application methods are equally likely to produce a case of no response, rejection and callback, respectively).

Subsequently, multivariate analyses are conducted to further investigate what has already been suggested by the descriptive results. The full empirical models for the probit regressions conducted below look as follows:

$$P_i(\text{Response}) = \beta_0 + \beta_1 \text{Correspondence}_i + \beta_2 \text{ Female/Turkish name}_i + \beta_3 \text{ Correspondence}_i * \text{Female/Turkish name}_i + \beta_4 \overrightarrow{\text{Controls}_i} + \mu_i$$

and

$$P_i(\text{Callback}) = \beta_0 + \beta_1 \text{ Correspondence}_i + \beta_2 \text{ Female/Turkish name}_i + \beta_3 \text{ Correspondence}_i * \text{Female/Turkish name}_i + \beta_4 \overrightarrow{\text{Controls}_i} + \mu_i,$$

where β_0 is a constant, β_k denotes the regression coefficient β of regressor k and μ_i represents a normally distributed error term of applicant i. The correspondence variable is a dummy that equals 1 if two matched applications were sent out in response to the same job offer. The dummy representing the minority group equals 1 either if the candidate was female or Turkish-named (depending on the dataset). In order to test the effect on the probability of receiving a response or a callback by the firms, two regressions with these two dependent variables were estimated separately for each sample. Controls include firm characteristics, regional labor market data, the certificate dummy, the job type (only in the gender study), the number of open positions, the distance as well as the dispatching order (which always equals 1 if only a single application is sent out) and the résumé design.

It could further be argued that, for instance, the minority candidate disproportionally benefits from not being in competition with an equally qualified applicant from the majority group for reasons discussed in the previous sections. The reference group, i.e., the German-named male candidate, might suffer if employers are unable to compare his application with someone being equipped with similar human capital endowments. Hence, the positive effects from direct competition for one candidate may outweigh the negative impact for the other candidate and vice versa. Consequently, an interaction term is included in the model that equals 1 if the minority group, i.e., the female or Turkish-named candidate, applies within the correspondence setting. The estimated coefficient should then account for any method-specific differences across groups.

Models (I) to (III) in tables 5-41 and 5-42 report average marginal effects from probit regressions on the response dummy. Both estimations indicate that the selection of the application method does not affect the likelihood of whether the employer contacts the job candidate or not. The point estimates of the correspondence dummy turn out to be statistically insignificant independent of the inclusion of an interaction term. Thus, there is no difference in employers' response behavior between the correspondence and single application method. In addition, no gender and name effects can be observed as neither interaction coefficient turns out to be statistically significant. Due to the insignificant effects, not surprisingly, the explanatory power of the regression models is rather low. This especially applies to the estimates in table 5-41 that do not predict employers' responses any better than the intercept model.

More convincingly and additionally supportive of the nonexistence of a correspondence effect are the results from probit analyses on the callback dummy. Focusing on the gender study, models (IV) to (VI) of table 5-41 show that the marginal effects of the correspondence dummy are insignificant. Apart from that, the insignificant interaction term in model (VI) does not lend support to any gender-specific effect.

Table 5-41: The Effects of the Correspondence Dummy on Response and Callback Rates in the Gender Study

	(I)	**(II)**	**(III)**	**(IV)**	**(V)**	**(VI)**
Correspondence	0.05	0.05	0.09	0.04	0.04	0.08
	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.07)
Female		0.04	0.09		-0.00	0.05
		(0.03)	(0.06)		(0.04)	(0.08)
Correspondence x Female			-0.08			-0.09
			(0.08)			(0.09)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
No. of obs.	444	444	444	444	444	444
Pseudo R^2	0.045	0.047	0.050	0.059	0.059	0.060
Log likelihood	-208.442	-207.889	-207.374	-280.117	-280.113	-279.750
Wald chi-squared	13.808	14.887	15.463	29.121	29.111	29.648
P-value	0.540	0.533	0.562	0.016	0.023	0.029

Notes: Each model reports average marginal effects of a probit regression on the response dummy (Y=1: employer gives the applicant either a rejection or a callback) (models (I) to (III)) and the callback dummy (Y=1: employer calls back the job applicant) (models (IV) to (VI)), respectively. Marginal effects are calculated at the mean of all independent variables and denote an infinitesimal change in case of continuous variables and a discrete change in case of dummy variables. Standard errors clustered on firm level are in parentheses. Regressions consider the entire sample as of table C-14 in the appendix. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

In line with this finding, the effect of the correspondence variable also turns out to be insignificant if the sample of the study on ethnic discrimination is considered (see models (IV) to (VI) of table 5-42). Both, the point estimate and interaction term do not significantly affect the hiring outcome. The systematic disadvantage of the Turkish-named applicant, however, remains. The minority candidate has an 18 percentage points lower chance of being invited to a job interview. If the name is interacted with the correspondence dummy, the effect becomes statistically insignificant which is most likely due to the small number of observations causing an increase in standard errors.

Table 5-42: The Effects of the Correspondence Dummy on Response and Callback Rates in the Ethnicity Study

	(I)	(II)	(III)	(IV)	(V)	(VI)
Correspondence	-0.06	-0.07	-0.08	-0.06	-0.07	-0.06
	(0.05)	(0.05)	(0.07)	(0.07)	(0.07)	(0.09)
Turkish name		-0.10***	-0.12		-0.18***	-0.16
		(0.04)	(0.08)		(0.05)	(0.10)
Correspondence x Turkish name			0.03			-0.02
			(0.09)			(0.11)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
No. of obs.	302	302	302	302	302	302
Pseudo R^2	0.074	0.088	0.089	0.034	0.056	0.056
Log likelihood	-139.484	-137.261	-137.213	-200.854	-196.429	-196.414
Wald chi-squared	17.185	23.550	25.760	10.219	22.480	24.982
P-value	0.246	0.073	0.058	0.746	0.096	0.070

Notes: Each model reports average marginal effects of a probit regression on the response dummy (Y=1: employer gives the applicant either a rejection or a callback) (models (I) to (III)) and the callback dummy (Y=1: employer calls back the job applicant) (models (IV) to (VI)), respectively. Marginal effects are calculated at the mean of all independent variables and denote an infinitesimal change in case of continuous variables and a discrete change in case of dummy variables. Standard errors clustered on firm level are in parentheses. Regressions consider the entire sample as of table C-15 in the appendix. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

Overall, the regression estimates indicate that the study design, i.e., whether single or matched pairs of applications are dispatched, neither affects joint hiring outcomes, nor callback probabilities by gender and name. These findings are robust across various model specifications and for two different datasets. Beyond that, the interaction effects all remain statistically insignificant for different combinations of the other independent variables.[75] The implications are thus twofold.

First, the presence and extent of discrimination demonstrated by the correspondence studies in section 5.2 and 5.3 are unbiased from any method-specific effects. Even though there has been increasing media coverage as a result of the Kaas and Manger (2012) study and the pilot project on anonymous applications (Krause et al., 2012b), the deceptive nature of the matched-pair testing apparently has not been revealed. This is further supported by findings reported in Carlsson and Rooth (2012) who neither find a relationship between public attention and employers' discriminatory behavior. Second, the evidence presented above supports the use of the single application method as an alternative to the correspondence testing. On the one hand, it further reduces the (involuntary) effort on behalf of employers which may increase acceptance of this

[75] Graphics illustrating the interaction effects are available from the author upon request.

experimental approach. On the other hand, using the single application method eliminates any remaining criticism associated with the correspondence method claiming that evidence of discrimination may be biased if employers reveal the deceptive nature of the study. For multivariate analyses of firms' responses, candidates could then be matched according to employer characteristics. At least the aggregate results for each demographic group should not significantly differ if the candidates apply individually.

So far, a ceteris paribus comparison of the single and pairwise application method has not been conducted. Only Gringart and Helmes (2001) use both approaches simultaneously. However, they investigate whether paired and single applications produce the same hiring outcomes if unsolicited applications rather than applications addressing publicly available job offers are dispatched. They draw the same conclusion with respect to aggregate hiring outcomes, but do not focus on any group-specific differences. Thus, to the best of the author's knowledge, the present study is the first showing that both procedures come to equivalent results. In fact, the single application method proves to be advantageous relative to correspondence testing in terms of lower costs to employers and higher feasibility on behalf of the researcher.

6 Conclusion

The last chapter begins with a summary of the main findings (6.1). Section 6.2 outlines where the present thesis has made a contribution to the academic literature before a special focus is laid upon policy implications and a discussion under which conditions any policy measures are likely to eliminate hiring discrimination (6.3). Finally, the thesis concludes by highlighting limitations and suggesting directions for future research (6.4).

6.1 Summary of Overall Findings

This thesis has presented results of two large-scaled field experiments designed to investigate gender and ethnic discrimination in predominantly male-dominated jobs in the German labor market for apprenticeships. Apprenticeships matter for both the labor market's demand and supply side. In Germany, a significant number of school graduates start their working careers as apprentices and quite often use dual training as a doorstep into regular employment. Employers, in contrast, either satisfy their current labor demand with apprentices or strategically invest in apprenticeship training to guarantee the supply of qualified labor in the future.

Firms offering apprenticeship positions in the years 2011 and 2012 were addressed by two equally equipped applicants that only differed with respect to one demographic characteristic such as gender in the first and ethnic background in the second study. The matched-pair design allows separating a treatment effect based on these characteristics from any other factors driving labor market differences. In particular, the pre-selection stage in the recruitment process, i.e., employers' callbacks to written applications, were reported for either candidate and compared between the control and minority group.

The study on gender discrimination, first of all, highlights that differential treatment mainly depends on the job type. Discrimination against the female candidate can only be observed in male-dominated occupations where women have a 19 percent (6.5 percentage points) lower callback probability as compared to men. A closer look at the factors influencing differential treatment shows that prior experience with female applicants as well as above-average labor market scarcity in the previous year have a statistically significant and positive impact on women's callback probabilities, but are economically marginal at best. In other words, the overall disadvantage of the female candidate neither disappears nor substantially decreases. Instead, the point in time when women apply for an apprenticeship affects their hiring probabilities relative to men. While male applicants have statistically the same callback rates independent of the application period, discrimination against the female candidates is restricted to late recruiters that publish their job offers shortly before the scheduled start of the contract.

With regard to the correspondence test investigating discrimination against Turkish-named applicants, the prevalence of discriminatory treatment has been found, although its sources remain rather suggestive. In fact, the minority candidate has a 32 percent (10.2 percentage points) lower chance of receiving a positive response from the firms addressed. Recruiter-type weakly affects the magnitude of this effect whereas late recruiters discriminate somewhat more. Hypotheses directly reflecting taste-based and statistical discrimination, however, are not supported. The inclusion of an additional credential equally benefits the majority and minority candidate and thus does not reduce the callback gap. Similarly, employers' behavior does not systematically change with a one standard deviation increase in the share of foreign applicants and in the ratio of unfilled to total apprenticeships.

Lastly, a subsample of both studies has been used to assess whether the results produced with the correspondence method persist if single rather than pair-wise applications are dispatched. The analyses here indicate that the findings are independent of methodological variations and yield the same outcomes.

6.2 Contribution to Academic Research

To the best of the author's knowledge, this is the first study that uses the correspondence method to investigate gender discrimination in access to employment in the German labor market. The study design not only allows identifying the prevalence of discriminatory treatment, but (also) provides direct evidence of its sources, none of which has been considered in the context of apprenticeship training thus far. The general findings are in line with similar field experiments from other countries and suggest that females are discriminated in male-dominated jobs. Yet, the involvement of both taste-based and statistical discrimination in employers' decision making process has not been found to exist to date. Most strikingly is the fact that the market seems to be divided into discriminators and non-discriminators where evidence is provided that links recruiter-type to managerial proficiency. Whether recruiter-type is endogenous, i.e., proves to be a result of inferior labor market reputation through systematic discriminatory treatment in the past, cannot be answered with the data at hand. Moreover, the cross-sectional character does not permit any conclusions on whether discriminators are driven out of the market in the long run, which would be a direct consequence of Becker's taste for discrimination approach.

With regard to the study on ethnic discrimination, results from prior research are qualitatively supported. Quantitatively, the extent of discrimination oscillates around the lower end of what has been found in foreign labor markets, but turns out to be higher compared to other studies conducted in Germany (see Goldberg et al., 1996; Kaas and Manger, 2012). The latter is in line with the predictions by Kaas and Manger (2012) who expect discrimination to be more prominent in low-qualified jobs. The evidence presented, however, goes along with the taste-based discrimination approach, given that misplacement of high-qualified positions is more costly and high-qualified employees are more difficult to find. Conversely, in relation to the findings from other labor markets, the relatively small hiring gap can be related to the increasing importance of apprentices to satisfy employers' future labor demand and their exposed position compared to other entry-level and low-qualified jobs predominantly analyzed in previous research.

Overall, the role of taste-based and statistical discrimination seems to be arguable. In fact, most of the hypotheses reflecting any of these approaches cannot be supported. Undoubtedly, further research studying and disentangling the effects from economic motives of discrimination is required. When designing future field experiments, though, results from methodological variations have shown that researchers should consider using (previously matched) single applications to approach employers as a suitable alternative to pair-wise testing.

6.3 POLICY IMPLICATIONS

Regarding the situation in the German labor market, the results presented are somewhat surprising. Even though employers continuously claim that their demand for qualified labor, especially in technical occupations, cannot, or at least not sufficiently, be satisfied, minority workers still face disadvantages in access to these jobs. This particularly counteracts initiatives with the goal to increase, for example, the fraction of women in male-dominated jobs and contradicts statements in job offers that prompt female candidates to apply. Given this affirmative environment, one would expect that women are, all other things equal, even favored when applying for male-dominated jobs. Selecting into these jobs may signal additional skills (e.g. assertiveness and ambition) which are not directly conveyed by written applications. Yet, the opposite holds true so that, as a result, labor market segregation persists with far-reaching consequences, inter alia, for wages, career profiles and even pre-market investment decisions. The results also quite convincingly outline the discrepancy between what employers state and how they actually (re)act. Reconsidering the ongoing discussion on voluntary and obligatory female quotas in top-management positions, similar developments can obviously be observed in other labor market segments, i.e., employers claim their good will, but lack revealing consequences.

From a policy-maker perspective, the discussion should rather emphasize how the callback differences found in the data can be eliminated or, at least, reduced. On a macroeconomic level, researchers have investigated the impact of changes in the legislation on equal opportunities in access to employment and have found that the introduction of anti-discrimination laws has been beneficial to females as well as racial minorities (e.g. Beller, 1982; Heckman and Payner, 1989). On firm level, though, the evidence is quite heterogeneous (Pager and Shepherd, 2008). The effects of Equal Employment Opportunity Laws are often hard to separate from any convergences that go back to increasing human capital endowments and improved schooling quality. Not surprisingly, differential treatment unrelated to productivity may still prevail as the present study shows.

One way to overcome intended and unintended discriminatory behavior is the implementation of some forms of blinding measures. While blind auditions indeed have raised the share of females in U.S. orchestras (Goldin and Rouse, 2000), a much more frequently used procedure in regular recruitment settings are anonymous applications. With this method, any information that allows inferences on applicants' demographic characteristics such as names, profile pictures and dates is made inaccessible to recruiters. In this way, the focus is solely upon productivity-driving factors that can actively be influenced by applicants. Unlike in the German labor market, highlighting human capital endowments and covering characteristics pre-determined by birth is very common in other countries (Krause et al., 2010). However, empirical evidence of its success in promoting minorities' employment opportunities is very limited and has only produced spurious results in favor of this procedure (Åslund and Nordström Skans, 2012; Krause et al., 2012a). In fact, reports following a pilot project that has tested anonymous applications in Germany show only marginally improved hiring opportunities for minority groups (Krause et al., 2012b). A thorough analysis of the costs and benefits associated with this procedure is, yet, missing.

Another way to address differential treatment is the implementation of affirmative action policies that actively promote the recruitment of minority applicants and may reach as far as exerting reverse discrimination, i.e., favoring minorities, all other things being equal (Holzer and Neumark, 2000a). Previous evidence shows that affirmative action policies increase the number of employers' recruiting and screening practices as well as their actual hires of ethnic minorities and females without suffering from a decrease in applicants' and employees' quality (Holzer and Neumark, 2000b).

As an alternative to measures that are embedded in the formal and organizational structure of the firm, results from audit and correspondence tests can be used simply to raise recruiters' consciousness on the prevalence of discrimination and its sources (Greenwald and Banaji, 1995). Understanding the latter is particularly crucial when deciding upon the implementation of a particular measure or a set of measures. Given the prevalence of taste-based discrimination, anonymizing applications would only postpone discriminatory treatment to the next stage of the recruitment process where, for example, in a face-to-face interview most demographic characteristics are revealed. As a consequence, discrimination persists while, simultaneously, both employers and applicants are confronted with higher costs from e.g. the time spent for preparing, travelling and conducting job interviews. On the other hand, in the presence of statistical discrimination, anonymous applications may well serve as a means to not only increase minorities' callbacks, but also their hiring probabilities. Having passed the first stage of the recruitment process, minorities have the possibility to convince employers of their individual quality and thus discard any negative perceptions based on group membership. If only statistical discrimination prevailed, the treatment effects would have been more prominent than actually reported. This, in turn, gives rise to the current results from the gender study indicating the presence of both statistical and taste-based discrimination. Initial blinding measures would therefore only eliminate differential treatment at workplaces where employers, coworkers and customers have neutral preferences.

Any recommendations with respect to diversity initiatives on firm level originating from the present findings remain somewhat suggestive. Previous evidence, for instance, finds that minority hires increase if the person responsible for the recruitment process belongs to the same minority group (e.g. Stoll et al., 2004; Giuliano and Levine, 2009). However, whether these effects reflect prejudices and information asymmetries or can be explained by sociological approaches such as similarity attraction (Byrne, 1971) or social identity theory (Tajfel, 1982) remains unanswered. Unfortunately, in the current context, gender and ethnic background of the actual decision maker cannot be retraced with certainty which makes any inferences on e.g. in-group favoritism sensitive to bias. Similarly, no information on demographic characteristics of employers' workforces is available which makes empirical investigations on the role of workforce diversity on discriminatory practices impossible.

Undoubtedly, the present findings stimulate the discussion on inequalities in access to employment. Policy makers may use the results to raise awareness among employers. Employers, in turn, may check their current recruitment practices for any group bias and, if necessary, establish more formalized procedures that leave less room for personal preferences and productivity misperceptions based on group membership. Besides, it seems worthwhile for employers to assess how coworkers' distastes influence hiring discrimination and what can be done to decrease the costs associated with group preferences.

From an applicant's perspective, the results from both the gender and ethnicity study strongly suggest minority candidates to address job offers from early recruiters as this significantly narrows the gap in callback rates between them and equally qualified majority applicants. After all, policy implications should be closely linked to the type of discrimination.

6.4 LIMITATIONS AND OUTLOOK

The field experiments entail a couple of limitations concerning the methodological approach, the data collected and the generalizability of results. First, matched-pair testing with fictitious applicants only allows observing the initial stage of the recruitment process. While previous research suggests that discrimination is most prominent when first contact takes place, it cannot be ruled out that the actual hiring gap is abolished, reduced or increased. Second, the market for apprenticeships only represents a snapshot of the German labor market as a whole. The prevalence and magnitude of discrimination may thus vary depending on the labor market segment investigated which calls for the inclusion of other industries and occupational positions. Third, the results unveiling the presence of ethnic discrimination refer to ethnic Turks but should not be regarded as evidence for discrimination against other ethnic minorities. According to previous empirical findings from the German labor market, callback rates of other minority groups are very likely to deviate from those of second generation Turks (see literature review in chapter 3.2.2). The fourth limitation concerns data availability. Unfortunately, no information on the entire applicant pool, the quality of applications as well as firms' training portfolios and threshold levels is available. As a consequence, no evidence on the relationship between recruitment standards, labor-supply competition, employers' reputation as a training company and differential treatment could be produced.

Finally, some problems are associated with the use of regional labor market data. Since companies all over Germany were referred to, while, at the same time, administrative constraints restricted the number of observations, for some regions employers' responses to only one correspondence pair exist. This, in turn, may result in small observation biases. Besides, the number of job offers in a region may be endogenous to the attractiveness of employers operating in that area. Employers' attractiveness, on the other hand, may be based upon their reputation in the labor market which, as has been argued in the empirical section, may negatively correlate with (the extent of) discriminatory behavior.

Future research should address the issues outlined above and continue focusing on the separation of taste-based and statistical discrimination. The design of further correspondence tests should permit more in-depth analyses of differences in returns to schooling and additional credentials. Bertrand and Mullainathan (2004), for example, create low and high quality résumés and find different return rates between white and black applicants, ceteris paribus. In the same vein, future research may vary years of schooling, school grades as well as the number and quality of work certificates. Thus, analyzing whether both, either or none of the applicant groups benefit from an increase in skill levels and amount of information provided is made possible. If the callback gap diminishes with supplemental credentials, the prevalence of statistical discrimination would be supported. In contrast, if not only informational deficits are abolished, but more ability is signaled, a decrease in discrimination would be a rational response to higher costs associated with ongoing and more intense search efforts and thus signal taste-based discrimination. Thus, the inclusion of credentials may be used as a proxy for different types of discrimination which should be considered carefully if matched-pair studies are set up. In order to clearly identify the effect of labor market scarcity, the extent of discrimination between a small number of a priori selected regions and occupations that differ with respect to labor supply and demand need to be compared (one such example is presented by Baert et al. (2013)).

Enhancing the number of observations by repeating matched-pair tests (retaining the experimental design) at the same employers in subsequent years would enable the researcher to build a (balanced) panel and allow further analyses of the recruiter effects by using fixed and random effects regressions. In particular, this would enable the researcher to observe whether recruiters respond to increasing/ decreasing labor market scarcity by shifting from late to early job offers and vice versa.

Even though the present studies empirically confirm the prevalence of hiring discrimination, it also remains subject to forthcoming research whether and under which conditions these systematic differences persist. In light of the demographic changes, higher skill requirements, voluntary and obligatory affirmative action policies and the increasing importance of employer branding, discrimination in the labor market may disappear in the long run. However, other trends might hinder or stop the discrimination-driven convergence of employment gaps. Investigating these trends promises further insights and therefore is a fruitful ground for future empirical research.

References

Abraham, Martin and Thomas Hinz (Eds.) (2008): Arbeitsmarktsoziologie. Wiesbaden: Verlag für Sozialwissenschaften.

Agerström, Jens and Dan-Olof Rooth (2011): The Role of Automatic Obesity Stereotypes in Real Hiring Discrimination. In: *Journal of Applied Psychology,* 96 (4): 790–805.

Ahmed, Ali M.; Andersson, Lina and Mats Hammarstedt (2009): Ethnic Discrimination in the Market Place of Small Business Transfers. In: *Economics Bulletin,* 29 (4): 3050–3058.

Ahmed, Ali M. and Mats Hammarstedt (2009): Detecting Discrimination against Homosexuals: Evidence from a Field Experiment on the Internet. In: *Economica,* 76 (303): 588–597.

Ai, Chunrong and Edward C. Norton (2003): Interaction Terms in Logit and Probit Models. In: *Economics Letters,* 80 (1): 123–129.

Aigner, Dennis J. and Glen G. Cain (1977): Statistical Theories of Discrimination in Labor Markets. In: *Industrial and Labor Relations Review,* 30 (2): 175–187.

Akerlof, George A. (1970): The Market for "Lemons": Quality Uncertainty and the Market Mechanism. In: *The Quarterly Journal of Economics,* 84 (3): 488–500.

Akerlof, George A. (1980): The Theory of Social Customs, of Which Unemployment May Be One Consequence. In: *The Quarterly Journal of Economics,* 94 (4): 749–775.

Albert, Rocío; Escot, Lorenzo and José A. Fernández-Cornejo (2011): A Field Experiment to Study Sex and Age Discrimination in the Madrid Labour Market. In: *The International Journal of Human Resource Management,* 22 (2): 351–375.

Aldashev, Alisher; Gernandt, Johannes and Stephan L. Thomsen (2007): Earnings Prospects for People with Migration Background in Germany. Zentrum für Europäische Wirtschaftsforschung (ZEW). Mannheim. Discussion Paper, No. 07-031.

Aldrich, John H. and Forrest D. Nelson (1984): Linear Probability, Logit, and Probit Models. Beverly Hills, California: Sage Publications.

Algan, Yann; Dustmann, Christian; Glitz, Albrecht and Alan Manning (2010): The Economic Situation of First and Second-Generation Immigrants in France, Germany and the United Kingdom. In: *The Economic Journal,* 120 (542): F4-F30.

Allasino, Enrico; Reyneri, E.; Venturini, A. and G. Zincone (2004): Labour Market Discrimination against Migrant Workers in Italy. International Labour Organization (ILO). Geneva. International Migration Papers, No. 67.

Altonji, Joseph G. (2005): Employer Learning, Statistical Discrimination and Occupational Attainment. In: *The American Economic Review,* 95 (2): 112–117.

Altonji, Joseph G.; Bharadwaj, Prashant and Fabian Lange (2012): Changes in the Characteristics of American Youth: Implications for Adult Outcomes. In: *Journal of Labor Economics,* 30 (4): 783–828.

Altonji, Joseph G. and Rebecca M. Blank (1999): Race and Gender in the Labor Market. In: Card, David E. and Orley C. Ashenfelter (Eds.): 3143–3259.

Altonji, Joseph G. and Charles R. Pierret (2001): Employer Learning and Statistical Discrimination. In: *The Quarterly Journal of Economics,* 116 (1): 313–350.

Anderson, Deborah and David Shapiro (1996): Racial Differences in Access to High-Paying Jobs and the Wage Gap between Black and White Women. In: *Industrial and Labor Relations Review,* 49 (2): 273–286.

Andriessen, Iris; Nievers, Eline; Dagevos, Jaco and Laila Faulk (2012): Ethnic Discrimination in the Dutch Labor Market: Its Relationship with Job Characteristics and Multiple Group Membership. In: *Work and Occupations,* 39 (3): 237–269.

Angel de Prada, M.; Actis, W.; Pereda, C. and R. Pérez Molina (1996): Labour Market Discrimination against Migrant Workers in Spain. International Labour Organization (ILO). Geneva. International Migration Papers, No. 9.

Arai, Mahmood; Bursell, Moa and Lena Nekby (2011): The Reverse Gender Gap in Ethnic Discrimination: Employer Priors against Men and Women with Arabic Names. Département d'Economie Appliquée, Université Libre de Bruxelles. Brussels. Research Series, No. 11-09.

Arai, Mahmood and Peter Skogman Thoursie (2009): Renouncing Personal Names: An Empirical Examination of Surname Change and Earnings. In: *Journal of Labor Economics,* 27 (1): 127–147.

Arai, Mahmood and Roger Vilhelmsson (2004): Unemployment-Risk Differentials between Immigrant and Native Workers in Sweden. In: *Industrial Relations,* 43 (3): 690–698.

Arrijn, Peter; Feld, Serge and André Nayer (1998): Discrimination in Access to Employment on Grounds of Foreign Origin the Case of Belgium. International Labour Organization (ILO). Geneva. International Migration Papers, No. 23.

Arrow, Kenneth J. (1971): The Theory of Discrimination. Industrial Relations Section, Princeton University. Princeton. Working Paper, No. 30A.

Arulampalam, Wiji; Booth, Alison L. and Mark L. Bryan (2007): Is There a Glass Ceiling over Europe? Exploring the Gender Pay Gap across the Wage Distribution. In: *Industrial and Labor Relations Review,* 60 (2): 163–186.

Arvey, Richard D.; Gordon, Michael E. and Douglas P. Massengill (1975): Differential Dropout Rates of Majority and Minority Job Candidates due to "Time Lags" between Selection Procedures. In: *Personnel Psychology,* 28 (2): 175–180.

Åslund, Olof and Oskar Nordström Skans (2012): Do Anonymous Job Application Procedures Level the Playing Field? In: *Industrial and Labor Relations Review,* 65 (1): 82–107.

Åslund, Olof and Dan-Olof Rooth (2005): Shifts in Attitudes and Labor Market Discrimination: Swedish Experiences after 9-11. In: *Journal of Population Economics,* 18 (4): 603–629.

Aydemir, Abdurrahman and Mikal Skuterud (2008): The Immigrant Wage Differential within and across Establishments. In: *Industrial and Labor Relations Review,* 61 (3): 334–352.

Ayres, Ian (1995): Further Evidence of Discrimination in New Car Negotiations and Estimates of Its Cause. In: *Michigan Law Review,* 94 (1): 109–147.

Ayres, Ian and Peter Siegelman (1995): Race and Gender Discrimination in Bargaining for a New Car. In: *The American Economic Review,* 85 (3): 304–321.

Ayres, Ian; Vars, Fred and Nasser Zakariya (2005): To Insure Prejudice: Racial Disparities in Taxicab Tipping. In: *Yale Law Journal,* 114 (7): 1613–1674.

Backes-Gellner, Uschi; Janssen, Simon and Simone N. Tuor Sartore (2013): Social Norms and Firms Discriminatory Pay-Setting. Department of Business Administration, University of Zurich. Zurich. Working Paper Series, No. 327.

Backes-Gellner, Uschi and Jens Mohrenweiser (2010): Apprenticeship Training: For Investment or Substitution? In: *International Journal of Manpower,* 31 (5): 545–562.

Backes-Gellner, Uschi and Simone N. Tuor Sartore (2010): Avoiding Labor Shortages by Employer Signaling: On the Importance of Good Work Climate and Labor Relations. In: *Industrial and Labor Relations Review,* 63 (2): 271–286.

Backhaus, Klaus; Erichson, Bernd and Rolf Weiber (2011): Multivariate Analysemethoden - Eine anwendungsorientierte Einführung. 13th ed. Berlin: Springer.

Baert, Stijn; Cockx, Bart; Gheyle, Niels and Cora Vandamme (2013): Do Employers Discriminate Less if Vacancies Are Difficult to Fill? Evidence from a Field Experiment. Institute for the Study of Labor (IZA). Bonn. Discussion Paper, No. 7145.

Baert, Stijn and Elsy Verhofstadt (2013): Labour Market Discrimination against Former Juvenile Delinquents: Evidence from a Field Experiment. Institute for the Study of Labor (IZA). Bonn. Discussion Paper, No. 7845.

Banerjee, Abhijit; Bertrand, Marianne; Datta, Saugato and Sendhil Mullainathan (2009): Labor Market Discrimination in Delhi: Evidence from a Field Experiment. In: *Journal of Comparative Economics,* 37 (1): 14–27.

Barth, Erling; Bratsberg, Bernt and Oddbjørn Raaum (2012): Immigrant Wage Profiles within and between Establishments. In: *Labour Economics,* 19 (4): 541–556.

Bayard, Kimberly; Hellerstein, Judith K.; Neumark, David and Kenneth R. Troske (2003): New Evidence on Sex Segregation and Sex Differences in Wages from Matched Employee-Employer Data. In: *Journal of Labor Economics,* 21 (4): 887–922.

Bechara, Peggy (2012): Gender Segregation and Gender Wage Differences During the Early Labour Market Career. Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI). Essen. Ruhr Economic Papers, No. 352.

Becker, Gary S. (1962): Investment in Human Capital: A Theoretical Analysis. In: *Journal of Political Economy,* 70 (5): 9–49.

Becker, Gary S. (1971): The Economics of Discrimination. 2nd ed. Chicago, Illinois: The University of Chicago Press.

Becker, Gary S. (1993): Human Capital - A Theoretical and Empirical Analysis, with Special Reference to Education. 3rd ed. Chicago, Illinois: The University of Chicago Press.

Beller, Andrea H. (1982): Occupational Segregation by Sex: Determinants and Changes. In: *The Journal of Human Resources,* 17 (3): 371–392.

Belley, Philippe; Havet, Nathalie and Guy Lacroix (2012): Wage Growth and Job Mobility in the Early Career: Testing a Statistical Discrimination Model of the Gender Wage Gap. Institute for the Study of Labor (IZA). Bonn. Discussion Paper, No. 6893.

Bellmann, Lutz and Silke Hartung (2010): Übernahmemöglichkeiten im Ausbildungsbetrieb - Eine Analyse mit dem IAB-Betriebspanel. In: *Sozialer Fortschritt,* 59 (6): 160–167.

Bendick, Marc; Brown, Lauren E. and Kennington Wall (1999): No Foot in the Door - An Experimental Study of Employment Discrimination against Older Workers. In: *Journal of Aging & Social Policy,* 10 (4): 5–23.

Bendick, Marc; Jackson, Charles W. and Victor A. Reinoso (1994): Measuring Employment Discrimination through Controlled Experiments. In: *The Review of Black Political Economy,* 23 (1): 25-48.

Bendick, Marc; Jackson, Charles W.; Reinoso, Victor A. and Laura E. Hodges (1991): Discrimination against Latino Job Applicants: A Controlled Experiment. In: *Human Resource Management,* 30 (4): 469–484.

Bertrand, Marianne; Chugh, Dolly and Sendhil Mullainathan (2005): Implicit Discrimination. In: *The American Economic Review,* 95 (2): 94–98.

Bertrand, Marianne and Sendhil Mullainathan (2004): Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. In: *The American Economic Review,* 94 (4): 991–1013.

Biernat, Monica and Diane Kobrynowicz (1997): Gender- and Race-Based Standards of Competence: Lower Minimum Standards but Higher Ability Standards for Devalued Groups. In: *Journal of Personality and Social Psychology,* 72 (3): 544–557.

Bjerk, David (2008): Glass Ceilings or Sticky Floors? Statistical Discrimination in a Dynamic Model of Hiring and Promotion. In: *The Economic Journal,* 118 (530): 961–982.

Black, Dan A. (1995): Discrimination in an Equilibrium Search Model. In: *Journal of Labor Economics,* 13 (2): 309–334.

Blau, Francine D. and Andrea H. Beller (1988): Trends in Earnings Differentials by Gender, 1971-1981. In: *Industrial and Labor Relations Review,* 41 (4): 513–529.

Blau, Francine D. and Lawrence M. Kahn (1992): The Gender Earnings Gap: Learning from International Comparisons. In: *The American Economic Review,* 82 (2): 533–538.

Blau, Francine D. and Lawrence M. Kahn (1997): Swimming Upstream: Trends in the Gender Wage Differential in the 1980s. In: *Journal of Labor Economics,* 15 (1): 1–42.

Blau, Francine D. and Lawrence M. Kahn (1999): Analyzing the Gender Pay Gap. In: *The Quarterly Review of Economics and Finance,* 39 (5): 625–646.

Blau, Francine D. and Lawrence M. Kahn (2000): Gender Differences in Pay. In: *The Journal of Economic Perspectives,* 14 (4): 75–99.

Blau, Francine D. and Lawrence M. Kahn (2003): Understanding International Differences in the Gender Pay Gap. In: *Journal of Labor Economics,* 21 (1): 106–144.

Blau, Francine D. and Lawrence M. Kahn (2006): The U.S. Gender Pay Gap in the 1990s: Slowing Convergence. In: *Industrial and Labor Relations Review,* 60 (1): 45–66.

Blinder, Alan S. (1973): Wage Discrimination: Reduced Form and Structural Estimates. In: *The Journal of Human Resources,* 8 (4): 436–455.

Blommaert, L.; Coenders, M. and F. van Tubergen (2013): Discrimination of Arabic-Named Applicants in the Netherlands: An Internet-Based Field Experiment Examining Different Phases in Online Recruitment Procedures. In: *Social Forces*: mimeo.

Bloom, Nicholas; Genakos, Christos; Sadun, Raffaella and John van Reenen (2012): Management Practices across Firms and Countries. National Bureau of Economic Research (NBER). Cambridge. Working Paper, No. 17850.

Bloom, Nicholas and John van Reenen (2007): Measuring and Explaining Management Practices across Firms and Countries. In: *The Quarterly Journal of Economics,* 122 (4): 1351–1408.

Booth, Alison L. and Andrew Leigh (2010): Do Employers Discriminate by Gender? A Field Experiment in Female-Dominated Occupations. In: *Economics Letters,* 107 (2): 236–238.

Booth, Alison L.; Leigh, Andrew and Elena Varganova (2012): Does Ethnic Discrimination Vary across Minority Groups? Evidence from a Field Experiment. In: *Oxford Bulletin of Economics and Statistics,* 74 (4): 547–573.

Borjas, George J. and Stephen G. Bronara (1989): Consumer Discrimination and Self-Employment. In: *Journal of Political Economy,* 97 (3): 581–605.

Bound, John and Richard B. Freeman (1992): What Went Wrong? The Erosion of Relative Earnings and Employment among Young Black Men in the 1980s. In: *The Quarterly Journal of Economics,* 107 (1): 201–232.

Bovenkerk, Frank; Gras, Mitzi J. I. and D. Ramsoedh (1996): Discrimination against Migrant Workers and Ethnic Minorities in Access to Employment in the Netherlands. International Labour Organization (ILO). Geneva. International Migration Papers, No. 4.

Bowles, Samuel; Gintis, Herbert and Melissa Osborne (2001): Incentive-Enhancing Preferences: Personality, Behavior, and Earnings. In: *The American Economic Review,* 91 (2): 155–158.

Bowlus, Audra J. and Zvi Eckstein (2002): Discrimination and Skill Differences in an Equilibrium Search Model. In: *International Economic Review,* 43 (4): 1309–1345.

Bratsberg, Bernt and Dek Terrell (1998): Experience, Tenure, and Wage Growth of Young Black and White Men. In: *The Journal of Human Resources,* 33 (3): 658–682.

Brown, Charles (1984): Black-White Earnings Ratios since the Civil Rights Act of 1964 - The Importance of Labor Market Dropouts. In: *The Quarterly Journal of Economics,* 99 (1): 31–44.

Brown, Randall S.; Moon, Marilyn and Barbara S. Zoloth (1980): Incorporating Occupational Attainment in Studies of Male-Female Earnings Differentials. In: *The Journal of Human Resources,* 15 (1): 3–28.

Brown, Sarah; Roberts, Jennifer and Karl Taylor (2011): The Gender Reservation Wage Gap: Evidence from British Panel Data. In: *Economics Letters,* 113 (1): 88–91.

Brück-Klingberg, Andrea; Burkert, Carola; Garloff, Alfred; Seibert, Holger and Rüdiger Wapler (2011): Does Higher Education Help Immigrants Find a Job? A Survival Analysis. German Institute for Employment Research (IAB). Nuremberg. Discussion Paper, No. 6/2011.

Bursell, Moa (2007): What's in a Name? A Field Experiment Test for the Existence of Ethnic Discrimination in the Hiring Process. Linnaeus Center for Integration Studies, Stockholm University. Stockholm. Working Paper, No. 2007:7.

Busch, Anne and Elke Holst (2011): Gender-Specific Occupational Segregation, Glass Ceiling Effects, and Earnings in Managerial Positions: Results of a Fixed Effects Model. German Institute for Economic Research (DIW). Berlin. Discussion Papers, No. 1101.

Busch, Anne and Elke Holst (2012): Occupational Sex Segregation and Management-Level Wages in Germany: What Role Does Firm Size Play? Institute for the Study of Labor (IZA). Bonn. Discussion Paper, No. 6568.

Byrne, Donn E. (1971): The Attraction Paradigm. New York, New York: Academic Press.

Caliendo, Marco; Schmidl, Ricarda and Arne Uhlendorff (2011): Social Networks, Job Search Methods and Reservation Wages: Evidence for Germany. In: *International Journal of Manpower,* 32 (7): 796–824.

Camerer, Colin F. and Robin M. Hogarth (1999): The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. In: *Journal of Risk and Uncertainty,* 19 (1-3): 7-42.

Cancio, Silvia A.; Evans, David T. and David J. Jr. Maume (1996): Reconsidering the Declining Significance of Race: Racial Differences in Early Career Wages. In: *American Sociological Review,* 61 (4): 541–556.

Card, David E. and Orley C. Ashenfelter (Eds.) (1999): Handbook of Labor Economics Vol. 3 (Part C). 1st ed. Amsterdam, New York, New York: Elsevier.

Card, David E. and Alan B. Krueger (1993): Trends in Relative Black-White Earnings Revisited. In: *The American Economic Review,* 83 (2): 85–91.

Carlsson, Magnus (2010): Experimental Evidence of Discrimination in the Hiring of First- and Second-Generation Immigrants. In: *Labour,* 24 (3): 263–278.

Carlsson, Magnus (2011): Does Hiring Discrimination Cause Gender Segregation in the Swedish Labor Market? In: *Feminist Economics,* 17 (3): 71–102.

Carlsson, Magnus and Dan-Olof Rooth (2007): Evidence of Ethnic Discrimination in the Swedish Labor Market Using Experimental Data. In: *Labour Economics,* 14 (4): 716–729.

Carlsson, Magnus and Dan-Olof Rooth (2008): Is It Your Foreign Name or Foreign Qualifications? An Experimental Study of Ethnic Discrimination in Hiring. Institute for the Study of Labor (IZA). Bonn. Discussion Paper, No. 3810.

Carlsson, Magnus and Dan-Olof Rooth (2011): Revealing Taste-Based Discrimination in Hiring: A Correspondence Testing Experiment with Geographic Variation. In: *Applied Economics Letters,* 19 (18): 1861–1864.

Carlsson, Magnus and Dan-Olof Rooth (2012): The Power of Media and Changes in Discriminatory Behavior Among Employers. In: *Journal of Media Economics,* 25 (2): 98–108.

Carneiro, Pedro; Heckman, James J. and Dimitri V. Masterov (2005): Labor Market Discrimination and Racial Differences in Premarket Factors. In: *Journal of Law and Economics,* 48 (1): 1–39.

Carrington, William J. and Kenneth R. Troske (1998): Interfirm Segregation and the Black/White Wage Gap. In: *Journal of Labor Economics,* 16 (2): 231–260.

Cediey, E. and F. Foroni (2008): Discrimination in Access to Employment on Grounds of Foreign Origin in France. International Labour Organization (ILO). Geneva. International Migration Papers, No. 85E.

Chandra, Amitabh (2000): Labor-Market Dropouts and the Racial Wage Gap: 1940-1990. In: *The American Economic Review,* 90 (2): 333–338.

Charles, Kerwin K. and Jonathan Guryan (2008): Prejudice and Wages: An Empirical Assessment of Becker's 'The Economics of Discrimination'. In: *Journal of Political Economy,* 116 (5): 773–809.

Charles, Kerwin K. and Jonathan Guryan (2011): Studying Discrimination: Fundamental Challenges and Recent Progress. National Bureau of Economic Research (NBER). Cambridge. Working Paper, No. 17156.

Charles, Kerwin K. and Jonathan Guryan (2013): Taste-Based or Statistical Discrimination: The Economics of Discrimination Returns to its Roots. In: *The Economic Journal,* 123 (572): 417–432.

Chevalier, Arnaud (2007): Education, Occupation and Career Expectations: Determinants of the Gender Pay Gap for UK Graduates. In: *Oxford Bulletin of Economics and Statistics,* 69 (6): 819–842.

Chiswick, Barry R.; Cohen, Yinon and Tzippi Zach (1997): The Labor Market Status of Immigrants: Effects of the Unemployment Rate at Arrival and Duration of Residence. In: *Industrial and Labor Relations Review,* 50 (2): 289–303.

Chzhen, Yekaterina and Karen Mumford (2011): Gender Gaps across the Earnings Distribution for Full-Time Employees in Britain: Allowing for Sample Selection. In: *Labour Economics,* 18 (6): 837–844.

Coate, Stephen and Glenn C. Loury (1993): Will Affirmative-Action Policies Eliminate Negative Stereotypes? In: *The American Economic Review,* 83 (5): 1220–1240.

Coleman, Mary T. and John Pencavel (1993): Trends in Market Work Behavior of Women since 1940. In: *Industrial and Labor Relations Review,* 46 (4): 653–676.

Constant, Amelie F. (1998): The Earnings of Male and Female Guestworkers and Their Assimilation into the German Labor Market: A Panel Study 1984–1993. Vanderbilt University. Nashville, Tennessee.

Constant, Amelie F. (2009): Businesswomen in Germany and Their Performance by Ethnicity: It Pays to Be Self-employed. In: *International Journal of Manpower,* 30 (1/2): 145–162.

Constant, Amelie F. and Douglas S. Massey (2003): Self-selection, Earnings, and Out-migration: A Longitudinal Study of Immigrants to Germany. In: *Journal of Population Economics,* 16 (4): 631–653.

Constant, Amelie F. and Douglas S. Massey (2005): Labor Market Segmentation and the Earnings of German Guestworkers. In: *Population Research and Policy Review,* 24 (5): 489–512.

Constant, Amelie F. and Yochanan Shachmurove (2006): Entrepreneurial Ventures and Wage Differentials between Germans and Immigrants. In: *International Journal of Manpower,* 27 (3): 208–229.

Constant, Amelie F.; Shachmurove, Yochanan and Klaus F. Zimmermann (2007): What Makes an Entrepreneur and Does it Pay? Native Men, Turks, and Other Migrants in Germany. In: *International Migration,* 45 (4): 71–100.

Cornell, Bradford and Ivo Welch (1996): Culture, Information, and Screening Discrimination. In: *Journal of Political Economy,* 104 (3): 542–571.

Correll, Joshua; Park, Bernadette; Judd, Charles M. and Bernd Wittenbrink (2002): The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals. In: *Journal of Personality and Social Psychology,* 83 (6): 1314–1329.

Correll, Shelley J.; Benard, Stephen and In Paik (2007): Getting a Job: Is There a Motherhood Penalty? In: *American Journal of Sociology,* 112 (5): 1297–1339.

Cotton, Jeremiah (1988): On the Decomposition of Wage Differentials. In: *The Review of Economics and Statistics,* 70 (2): 236–243.

Cotton, John L.; O'Neill, Bonnie S. and Andrea Griffin (2008): The "Name Game": Affective and Hiring Reactions to First Names. In: *Journal of Managerial Psychology,* 23 (1): 18–39.

Croson, Rachel and Uri Gneezy (2009): Gender Differences in Preferences. In: *Journal of Economic Literature,* 47 (2): 448–474.

Crossley, Thomas F.; Jones, Stephen R. G. and Peter J. Kuhn (1994): Gender Differences in Displacement Cost: Evidence and Implications. In: *The Journal of Human Resources,* 29 (2): 461–480.

Cunningham, James S. and Nadja Zalokar (1992): The Economic Progress of Black Women, 1940-1980: Occupational Distribution and Relative Wages. In: *Industrial and Labor Relations Review,* 45 (3): 540–555.

D'Amico, Ronald and Nan L. Maxwell (1994): The Impact of Post-School Joblessness on Male Black-White Wage Differentials. In: *Industrial Relations,* 33 (2): 184–205.

Darity, William A.; Guilkey, David K. and William Winfrey (1996): Explaining Differences in Economic Performance among Racial and Ethnic Groups in the USA. In: *American Journal of Economics and Sociology,* 55 (4): 411–425.

Darity, William A. and Patrick L. Mason (1998): Evidence on Discrimination in Employment: Codes of Color, Codes of Gender. In: *The Journal of Economic Perspectives,* 12 (2): 63–90.

Derous, Eva and Ann M. Ryan (2012): Documenting the Adverse Impact of Résumé Screening: Degree of Ethnic Identification Matters. In: *International Journal of Selection and Assessment,* 20 (4): 464–474.

Dickinson, David L. and Ronald L. Oaxaca (2012): Wages Employment and Statistical Discrimination - Evidence from the Laboratory. Department of Economics, Appalachian State University. Boone. Working Paper, No. 12-03.

Dionisius, Regina; Mühlemann, Samuel; Pfeifer, Harald; Walden, Günter; Wenzelmann, Felix and Stefan C. Wolter (2009): Costs and Benefits of Apprenticeship Training - A Comparison of Germany and Switzerland. In: *Applied Economics Quarterly,* 55 (1): 7–37.

Doiron, Denise J. and Craig W. Riddell (1994): The Impact of Unionization on Male-Female Earnings Differences in Canada. In: *The Journal of Human Resources,* 29 (2): 504–534.

Dolton, Peter J. and Michael P. Kidd (1994): Occupational Access and Wage Discrimination. In: *Oxford Bulletin of Economics and Statistics,* 56 (4): 457–474.

Duguet, Emmanuel; Loïc, Du P.; L'Horty, Yannick and Pascale Petit (2012): First Order Stochastic Dominance and the Measurement of Hiring Discrimination: A Ranking

Extension of Correspondence Testings with an Application to Gender and Origin. Université Paris-Est. Créteil, Paris.

Duguet, Emmanuel; Petit, Pascale and Pascal Petit (2005): Hiring Discrimination in the French Financial Sector: An Econometric Analysis on Field Experiment Data. In: *Annals of Economics and Statistics,* Apr.-Jun. (78): 79–102.

Eberharter, Veronika V. (2012): The Intergenerational Transmission of Occupational Preferences, Segregation, and Wage Inequality - Empirical Evidence from Europe and the United States. German Institute for Economic Research (DIW). Berlin. SOEPpapers, No. 506.

Edin, Per-Anders and Jonas Lagerström (2004): Blind Dates: Quasi-Experimental Evidence on Discrimination. Institute for Evaluation of Labour Market and Education Policy (IFAU). Uppsala.

Eliasson, Tove (2013): Decomposing Immigrant Wage Assimilation: The Role of Workplaces and Occupations. Institute for Evaluation of Labour Market and Education Policy (IFAU). Uppsala. Working Paper, No. 2013:7.

England, Paula (1982): The Failure of Human Capital Theory to Explain Occupational Sex Segregation. In: *The Journal of Human Resources,* 17 (3): 358–370.

England, Paula (2005): Gender Inequality in Labor Markets: The Role of Motherhood and Segregation. In: *Social Politics: International Studies in Gender, State and Society,* 12 (2): 264–288.

England, Paula; Farkas, George; Kilbourne, Barbara and Thomas Dou (1988): Explaining Occupational Sex Segregation and Wages: Findings from a Model with Fixed Effects. In: *American Sociological Review,* 53 (4): 544–558.

England, Paula; Hermsen, Joan M. and David A. Cotter (2000): The Devaluation of Women's Work: A Comment on Tam. In: *American Journal of Sociology,* 105 (6): 1741–1751.

Engle, Robert F. (2007): Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics. In: Griliches, Zvi (Ed.): 796–801.

Eriksson, Stefan and Jonas Lagerström (2012): Detecting Discrimination in the Hiring Process: Evidence from an Internet-Based Search Channel. In: *Empirical Economics,* 43 (2): 537–563.

Erosa, Andrés; Fuster, Luisa and Diego Restuccia (2002): Fertility Decisions and Gender Differences in Labor Turnover, Employment, and Wages. In: *Review of Economic Dynamics,* 5 (4): 856–891.

European Commission (2010): Report on Equality between Women and Men 2010. Directorate-General for Employment, Social Affairs and Equal Opportunities, European Commission. Luxembourg.

European Commission (2011): Demography Report 2010: Older, More Numerous and Diverse Europeans. Directorate-General for Employment, Social Affairs and Equal Opportunities, European Commission. Luxembourg.

Even, William E. and David A. Macpherson (1993): The Decline of Private-Sector Unionism and the Gender Wage Gap. In: *The Journal of Human Resources,* 28 (2): 279–296.

Ewens, Michael; Tomlin, Bryan and Liang C. Wang (2012): Statistical Discrimination or Prejudice? A Large Sample Field Experiment. In: *Review of Economics and Statistics,* mimeo.

Fairlie, Robert W. and William A. Sundstrom (1997): The Racial Unemployment Gap in Long-Run Perspective. In: *The American Economic Review,* 87 (2): 306–310.

Falk, Armin and Ernst Fehr (2003): Why Labour Market Experiments? In: *Labour Economics,* 10 (4): 399–406.

Falk, Armin; Lalive, Rafael and Josef Zweimüller (2005): The Success of Job Applications: A New Approach to Program Evaluation. In: *Labour Economics,* 12 (6): 739–748.

Fama, Eugene F. (1980): Agency Problems and the Theory of the Firm. In: *Journal of Political Economy,* 88 (2): 288–307.

Farkas, George and Keven Vicknair (1996): Appropriate Tests of Racial Wage Discrimination Require Controls for Cognitive Skill: Comment on Cancio, Evans, and Maume. In: *American Sociological Review,* 61 (4): 557–560.

Fearon, Gervan and Steven Wald (2011): The Earnings Gap between Black and White Workers in Canada: Evidence from the 2006 Census. In: *Industrial Relations,* 66 (3): 324–348.

Fernandez, Roberto M. and Colette Friedrich (2011): Gender Sorting at the Application Interface. In: *Industrial Relations,* 50 (4): 591–609.

Fertig, Michael and Stefanie Schurer (2007): Labour Market Outcomes of Immigrants in Germany: The Importance of Heterogeneity and Attrition Bias. Institute for the Study of Labor (IZA). Bonn. Discussion Paper, No. 2915.

Fibbi, Rosita; Lerch, Mathias and Philippe Wanner (2006): Unemployment and Discrimination against Youth of Immigrant Origin in Switzerland: When the Name Makes the Difference. In: *Journal of International Migration and Integration,* 7 (3): 351-366.

Fidell, L. S. (1970): Empirical Verification of Sex Discrimination in Hiring Practices in Psychology. In: *American Psychologist,* 25 (12): 1094–1098.

Fields, Judith and Edward N. Wolff (1995): Interindustry Wage Differentials and the Gender Wage Gap. In: *Industrial and Labor Relations Review,* 49 (1): 105–120.

Filer, Randall K. (1985): Male-Female Wage Differences: The Importance of Compensating Differentials. In: *Industrial and Labor Relations Review,* 38 (3): 426–437.

Filippin, Antonio and Andrea Ichino (2005): Gender Wage Gap in Expectations and Realizations. In: *Labour Economics,* 12 (1): 125–145.

Finke, Claudia (2011): Verdienstunterschiede zwischen Männern und Frauen - Eine Ursachenanalyse auf Grundlage der Verdienststrukturerhebung 2006. German Federal Statistical Office (Destatis). Wiesbaden.

Firth, Michael (1981): Racial Discrimination in the British Labor Market. In: *Industrial and Labor Relations Review,* 34 (2): 265–272.

Firth, Michael (1982): Sex Discrimination in Job Opportunities for Women. In: *Sex Roles,* 8 (8): 891-901.

Fitzenberger, Bernd; Schnabel, Reinhold and Gaby Wunderlich (2004): The Gender Gap in Labor Market Participation and Employment: A Cohort Analysis for West Germany. In: *Journal of Population Economics,* 17 (1): 83–116.

Fitzenberger, Bernd and Gaby Wunderlich (2002): Gender Wage Differences in West Germany: A Cohort Analysis. In: *German Economic Review,* 3 (4): 379–414.

Fix, Michael and Raymond Struyk (Eds.) (1993): Clear and Convincing Evidence: Measurement of Discrimination in America. Washington D.C.: Urban Institute Press.

Forstenlechner, Ingo and Mohammed A. Al-Waqfi (2010): "A Job Interview for Mo, but None for Mohammed": Religious Discrimination against Immigrants in Austria and Germany. In: *Personnel Review,* 39 (6): 767–784.

Fortin, Nicole M. (2005): Gender Role Attitudes and the Labour-Market Outcomes of Women across OECD Countries. In: *Oxford Review of Economic Policy,* 21 (3): 416–438.

Fransen, Eva; Plantenga, Janneke and Jan D. Vlasblom (2012): Why Do Women Still Earn Less Than Men? Decomposing the Dutch Gender Pay Gap, 1996–2006. In: *Applied Economics,* 44 (33): 4343–4354.

Frick, Bernd and Michael Maihaus (2013): The Structure and Determinants of Expected and Actual Starting Salaries of Higher Education Students in Germany: Identical or Different? Department of Management, University of Paderborn. Paderborn.

Fryer, Ronald G. and Steven D. Levitt (2004): The Causes and Consequences of Distinctively Black Names. In: *The Quarterly Journal of Economics,* 119 (3): 767–805.

German Chambers of Commerce and Industry (DIHK) (2011): Ausbildung 2011 - Ergebnisse einer IHK-Online-Unternehmensbefragung. German Chambers of Commerce and Industry (DIHK). Berlin, Brussels.

German Federal Employment Agency (BA) (1988): Klassifizierung der Berufe - Systematisches und alphabetisches Verzeichnis der Berufsbenennungen von 1988. German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2005): Nationaler Pakt für Ausbildung und Fachkräftenachwuchs in Deutschland vom 16. Juni 2004 - Berichte und Dokumente zu den Ergebnissen des Paktjahres 2004 und Ausblick auf 2005. German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2007): Nationaler Pakt für Ausbildung und Fachkräftenachwuchs in Deutschland 2007-2010. German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2010a): Arbeitsmarkt in Zahlen - Ausbildungsstellenmarkt - Bewerber und Berufsausbildungsstellen in Deutschland, September 2010. German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2010b): Klassifikation der Berufe 2010 – Band 1: Systematischer und alphabetischer Teil mit Erläuterungen. German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2010c): Nationaler Pakt für Ausbildung und Fachkräftenachwuchs in Deutschland 2010-2014. German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2011): Arbeitsmarkt in Zahlen - Ausbildungsstellenmarkt - Bewerber und Berufsausbildungsstellen in Deutschland, September 2011. German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2012a): Analyse des Arbeitsmarktes für Frauen und Männer. German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2012b): Arbeitsmarkt in Zahlen - Ausbildungsstellenmarkt - Bewerber und Berufsausbildungsstellen in Deutschland, September 2012. German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2012c): Arbeitsmarkt in Zahlen, Sozialversicherungspflichtig Beschäftigte nach Wirtschaftszweigen (WZ 2008). German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2012d): Berufsstatistik Altenpfleger. German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2012e): Berufsstatistik Elektroniker-Betriebstechnik. German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2012f): Berufsstatistik Fachkraft-Lagerlogistik. German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2012g): Berufsstatistik Industriekaufmann und Bürokommunikation. German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2012h): Berufsstatistik Industriemechaniker. German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2012i): Berufsstatistik Mechatroniker. German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2012j): Berufsstatistik Verfahrensmechaniker. German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2012k): Berufsstatistik Zerspanungsmechaniker. German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2012l): Kurzinformationen zur Ausbildungsstellenmarktstatistik. German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2012m): Migrationshintergrund nach § 281 Abs. 2 SGB III - Grundlagen der Erhebung. German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2013a): Arbeitslosigkeit im Zeitverlauf. German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2013b): Arbeitslosigkeit im Zeitverlauf 2012. German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2013c): Beschäftigungsquoten der sozialversicherungspflichtig Beschäftigten. German Federal Employment Agency (BA). Nuremberg.

German Federal Employment Agency (BA) (2013d): Sozialversicherungspflichtige Beschäftigte nach ausgewählten Merkmalen - Zeitreihen. German Federal Employment Agency (BA). Nuremberg.

German Federal Institute for Vocational Education and Training (BIBB) (2006): Werden ausländische Jugendliche aus dem dualen System der Berufsausbildung verdrängt? German Federal Institute for Vocational Education and Training (BIBB). Bonn. Berufsbildung in Wissenschaft und Praxis, No. 3/2006.

German Federal Institute for Vocational Education and Training (BIBB) (2009a): Betriebliche Berufsausbildung: Eine lohnende Investition für die Betriebe - Ergebnisse der BIBB-Kosten- und Nutzenerhebung 2007. German Federal Institute for Vocational Education and Training (BIBB). Bonn. BIBB Report, No. 8/09.

German Federal Institute for Vocational Education and Training (BIBB) (2009b): Unbesetzte Ausbildungsplätze – Warum Betriebe erfolglos bleiben - Ergebnisse des BIBB-Ausbildungsmonitors. German Federal Institute for Vocational Education and Training (BIBB). Bonn. BIBB Report, No. 10/09.

German Federal Institute for Vocational Education and Training (BIBB) (2010a): Datenreport zum Berufsbildungsbericht 2009 - Informationen und Analysen zur Entwicklung der beruflichen Bildung. German Federal Institute for Vocational Education and Training (BIBB). Bonn.

German Federal Institute for Vocational Education and Training (BIBB) (2010b): Rangliste 2010 der Ausbildungsberufe nach Anzahl der Neuabschlüsse. German Federal Institute for Vocational Education and Training (BIBB). Bonn.

German Federal Institute for Vocational Education and Training (BIBB) (2011a): Datenreport zum Berufsbildungsbericht 2010 - Informationen und Analysen zur Entwicklung der beruflichen Bildung. German Federal Institute for Vocational Education and Training (BIBB). Bonn.

German Federal Institute for Vocational Education and Training (BIBB) (2011b): Rekrutierung von Auszubildenden – Betriebliches Rekrutierungsverhalten im Kontext des demografischen Wandels. German Federal Institute for Vocational Education and Training (BIBB). Bonn.

German Federal Institute for Vocational Education and Training (BIBB) (2012a): BIBB Qualifizierungspanel - Gründe für unbesetzte Ausbildungsstellen aus Sicht von Betrieben. German Federal Institute for Vocational Education and Training (BIBB). Bonn. Kurzinformationen, No. 3.

German Federal Institute for Vocational Education and Training (BIBB) (2012b): Datenreport zum Berufsbildungsbericht 2011 - Informationen und Analysen zur Entwicklung der beruflichen Bildung. German Federal Institute for Vocational Education and Training (BIBB). Bonn.

German Federal Statistical Office (Destatis) (2012a): Bevölkerung und Erwerbstätigkeit - Bevölkerung mit Migrationshintergrund - Ergebnisse des Mikrozensus 2011. German Federal Statistical Office (Destatis). Wiesbaden. Fachserie 1, Reihe 2.2.

German Federal Statistical Office (Destatis) (2012b): Mikrozensus 2011 - Bevölkerung und Erwerbstätigkeit - Beruf, Ausbildung und Arbeitsbedingungen der Erwerbstätigen in Deutschland. German Federal Statistical Office (Destatis). Wiesbaden. Fachserie 1, Reihe 4.1.2.

German Federal Statistical Office (Destatis) (2013): Verdienste und Arbeitskosten im Jahr 2012 - Arbeitnehmerverdienste. German Federal Statistical Office (Destatis). Wiesbaden. Fachserie 16, Reihe 2.3.

German Socio-Economic Panel (GSOEP) (2012): Data for Years 1984-2011, Version 28. Berlin. German Institute for Economic Research (DIW).

Gittleman, Maury B. and David R. Howell (1995): Changes in the Structure and Quality of Jobs in the United States: Effects by Race and Gender, 1973-1990. In: *Industrial and Labor Relations Review,* 48 (3): 420–440.

Giuliano, Laura and David I. Levine (2009): Manager Race and the Race of New Hires. In: *Journal of Labor Economics,* 27 (4): 589–631.

Giuliano, Laura and Michael R. Ransom (2011): Manager Ethnicity and Employment Segregation. Institute for the Study of Labor (IZA). Bonn. Discussion Paper, No. 5437.

Glass, Jennifer (1990): The Impact of Occupational Segregation on Working Conditions. In: *Social Forces,* 68 (3): 779–796.

Glass, Jennifer and Valerie Camarigg (1992): Gender, Parenthood, and Job-Family Compatibility. In: *American Journal of Sociology,* 98 (1): 131–151.

Glauber, Rebecca (2012): Women's Work and Working Conditions: Are Mothers Compensated for Lost Wages? In: *Work and Occupations,* 39 (2): 115–138.

Glick, Peter; Zion, Cari and Cynthia Nelson (1988): What Mediates Sex Discrimination in Hiring Decisions? In: *Journal of Personality and Social Psychology,* 55 (2): 178–186.

Gneezy, Uri; List, John and Michael K. Price (2012): Toward an Understanding of Why People Discriminate: Evidence from a Series of Natural Field Experiments. National Bureau of Economic Research (NBER). Cambridge. Working Paper, No. 17855.

Gneezy, Uri and John A. List (2004): Are the Disabled Discriminated against in Product Markets? Evidence from Field Experiments. University of Chicago; University of Maryland; National Bureau of Economic Research (NBER). Chicago, Illinois.

Gneezy, Uri; Niederle, Muriel and Aldo Rustichini (2003): Performance in Competitive Environments: Gender Differences. In: *The Quarterly Journal of Economics,* 118 (3): 1049–1074.

Gobillon, Laurent; Meurs, Dominique and Sébastien Roux (2012): Estimating Gender Differences in Access to Jobs. Institute for the Study of Labor (IZA). Bonn. Discussion Paper, No. 6928.

Goldberg, Andreas; Mourinho, Dora and Ursula Kulke (1996): Labour Market Discrimination against Foreign Workers in Germany. International Labour Organization (ILO). Geneva. International Migration Papers, No. 7.

Goldberg, Pinelopi K. (1996): Dealer Price Discrimination in New Car Purchases: Evidence from the Consumer Expenditure Survey. In: *Journal of Political Economy,* 104 (3): 622–654.

Goldin, Claudia and Cecilia Rouse (2000): Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians. In: *The American Economic Review,* 90 (4): 715–741.

Goldsmith, Arthur H.; Hamilton, Darrick and William A. Darity (2007): From Dark to Light: Skin Color and Wages among African-Americans. In: *The Journal of Human Resources,* 42 (4): 701–738.

Gottschalk, Peter (1997): Inequality, Income Growth, and Mobility: The Basic Facts. In: *The Journal of Economic Perspectives,* 11 (2): 21–40.

Graddy, Kathryn (1997): Do Fast-Food Chains Price Discriminate on the Race and Income Characteristics of an Area? In: *Journal of Business & Economic Statistics,* 15 (4): 391–401.

Greenwald, Anthony G. and Mahzarin R. Banaji (1995): Implicit Social Cognition: Attitudes, Self-esteem, and Stereotypes. In: *Psychological Review,* 102 (1): 4–27.

Greenwald, Anthony G.; McGhee, Debbie E. and Jordan L. Schwartz (1998): Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. In: *Journal of Personality and Social Psychology,* 74 (6): 1464–1480.

Gringart, Eyal and Edward Helmes (2001): Age Discrimination in Hiring Practices against Older Adults in Western Australia: the Case of Accounting Assistants. In: *Australasian Journal on Ageing,* 20 (1): 23–28.

Groshen, Erica L. (1991): The Structure of the Female/Male Wage Differential - Is It Who You Are, What You Do, or Where You Work. In: *The Journal of Human Resources,* 26 (3): 457–472.

Grossman, Sanford J. and Oliver D. Hart (1983): An Analysis of the Principal-Agent Problem. In: *Econometrica,* 51 (1): 7–45.

Grove, Wayne A.; Hussey, Andrew and Michael Jetter (2011): The Gender Pay Gap beyond Human Capital - Heterogeneity in Noncognitive Skills and in Labor Market Tastes. In: *The Journal of Human Resources,* 46 (4): 827–874.

Gujarati, Damodar N. and Dawn C. Porter (2009): Basic Econometrics. 5th ed. Boston, Massachusetts: McGraw-Hill Irwin.

Gupta, Nabanita D. and Donna S. Rothstein (2005): The Impact of Worker and Establishment-Level Characteristics on Male-Female Wage Differentials: Evidence from Danish Matched Employee-Employer Data. In: *Labour,* 19 (1): 1–34.

Gwartney, James and Charles Haworth (1974): Employer Costs and Discrimination: The Case of Baseball. In: *Journal of Political Economy,* 82 (4): 873–881.

Haile, Getinet A. (2009): Workplace Disability Diversity and Job-Related Well-Being in Britain: A WERS2004 Based Analysis. Institute for the Study of Labor (IZA). Bonn. Discussion Paper, No. 3993.

Haile, Getinet A. (2012): Unhappy Working with Men? Workplace Gender Diversity and Job-Related Well-Being in Britain. In: *Labour Economics,* 19 (3): 329–350.

Haile, Getinet A. (2013): Are You Unhappy Having Minority Co-Workers? Institute for the Study of Labor (IZA). Bonn. Discussion Paper, No. 7423.

Hamermesh, Daniel S. and Jeff E. Biddle (1994): Beauty and the Labor Market. In: *The American Economic Review,* 84 (5): 1174–1194.

Harless, David W. and George E. Hoffer (2002): Do Women Pay More for New Vehicles? Evidence from Transaction Price Data. In: *The American Economic Review,* 92 (1): 270–279.

Harris, Anne-Marie G.; Henderson, Geraldine R. and Jerome D. Williams (2005): Courting Customers: Assessing Consumer Racial Profiling and Other Marketplace Discrimination. In: *Journal of Public Policy and Marketing,* 24 (1): 163–171.

Harrison, Glenn W. and John A. List (2004): Field Experiments. In: *Journal of Economic Literature,* 42 (4): 1009–1055.

Heath, Anthony F.; Rothon, Catherine and Elina Kilpi (2008): The Second Generation in Western Europe: Education, Unemployment, and Occupational Attainment. In: *Annual Review of Sociology,* 34 (1): 211–235.

Heckman, James J. (1998): Detecting Discrimination. In: *The Journal of Economic Perspectives,* 12 (2): 101–116.

Heckman, James J. and Brook S. Payner (1989): Determining the Impact of Federal Antidiscrimination Policy on the Economic Status of Blacks: A Study of South Carolina. In: *The American Economic Review,* 79 (1): 138–177.

Heckman, James J. and Peter Siegelman (1993): The Urban Institute Audit Studies: Their Methods and Findings. In: Fix, Michael and Raymond Struyk (Eds.): 187–258.

Heckman, James J.; Stixrud, Jora and Sergio S. Urzúa (2006): The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. In: *Journal of Labor Economics,* 24 (3): 411–482.

Heilman, Madeline E. (1984): Information as a Deterrent against Sex Discrimination: The Effects of Applicant Sex and Information Type on Preliminary Employment Decisions. In: *Organizational Behavior and Human Performance,* 33 (2): 174–186.

Heilman, Madeline E.; Martell, Richard F. and Michael C. Simon (1988): The Vagaries of Sex Bias: Conditions Regulating the Undervaluation, Equivaluation, and Overvaluation of Female Job Applicants. In: *Organizational Behavior and Human Decision Processes,* 41 (1): 98–110.

Hellerstein, Judith K. and David Neumark (2006): Using Matched Employer-Employee Data to Study Labor Market Discrimination. In: Rodgers, William M. (Ed.): 29–60.

Hellerstein, Judith K.; Neumark, David and Kenneth R. Troske (2002): Market Forces and Sex Discrimination. In: *The Journal of Human Resources,* 37 (2): 353–380.

Hirsch, Boris and Elke J. Jahn (2012): Is There Monopsonistic Discrimination against Immigrants? First Evidence from Linked Employer-Employee Data. Institute for the Study of Labor (IZA). Bonn. Discussion Paper, No. 6472.

Hirsch, Boris; König, Marion and Joachim Möller (2009): Is There a Gap in the Gap? Regional Differences in the Gender Pay Gap. Institute for the Study of Labor (IZA). Bonn. Discussion Paper, No. 4231.

Hirsch, Boris; Schank, Thorsten and Claus Schnabel (2010): Differences in Labor Supply to Monopsonistic Firms and the Gender Pay Gap: An Empirical Analysis Using Linked Employer-Employee Data from Germany. In: *Journal of Labor Economics,* 28 (2): 291–330.

Hitt, Michael A. and William G. Zikmund (1985): Forewarned is Forearmed: Potential between and within Sex Discrimination. In: *Sex Roles,* 12 (7-8): 807-812.

Holzer, Harry J. (1987): Informal Job Search and Black Youth Unemployment. In: *The American Economic Review,* 77 (3): 446–452.

Holzer, Harry J. and Keith R. Ihlanfeldt (1998): Customer Discrimination and Employment Outcomes for Minority Workers. In: *The Quarterly Journal of Economics,* 113 (3): 835–867.

Holzer, Harry J. and David Neumark (2000a): Assessing Affirmative Action. In: *Journal of Economic Literature,* 38 (3): 483–568.

Holzer, Harry J. and David Neumark (2000b): What Does Affirmative Action Do? In: *Industrial and Labor Relations Review,* 53 (2): 240–271.

Huffman, Matt L. and Philip N. Cohen (2004): Racial Wage Inequality: Job Segregation and Devaluation across U.S. Labor Markets. In: *American Journal of Sociology,* 109 (4): 902–936.

Hunt, Priscillia (2012): From the Bottom to the Top: A More Complete Picture of the Immigrant-Native Wage Gap in Britain. In: *IZA Journal of Migration,* 1 (1): 1-18.

Ioannides, Yannis M. and Linda D. Loury (2004): Job Information Networks, Neighborhood Effects, and Inequality. In: *Journal of Economic Literature,* 42 (4): 1056–1093.

Jacquemet, Nicolas and Constantine Yannelis (2012): Indiscriminate Discrimination: A Correspondence Test for Ethnic Homophily in the Chicago Labor Market. In: *Labour Economics,* 19 (6): 824–832.

Jarrell, Stephen B. and Tom D. Stanley (1998): Gender Wage Discrimination Bias? A Meta-Regression Analysis. In: *The Journal of Human Resources,* 33 (4): 947–973.

Jarrell, Stephen B. and Tom D. Stanley (2004): Declining Bias and Gender Wage Discrimination? A Meta-Regression Analysis. In: *The Journal of Human Resources,* 39 (3): 828–838.

Javdani, Mohsen (2013): Glass Ceilings or Glass Doors? The Role of Firms in Male-Female Wage Disparities. Department of Economics, University of British Columbia Okanagan. Kelowna.

Jensen, Michael C. and William H. Meckling (1976): Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. In: *Journal of Financial Economics,* 3 (4): 305–360.

Jirjahn, Uwe (2011): Gender, Worker Representation and the Profitability of Firms in Germany. In: *European Journal of Comparative Economics,* 8 (2): 281–298.

Johnston, David W. and Wang-Sheng Lee (2012): Climbing the Job Ladder: New Evidence of Gender Inequity. In: *Industrial Relations,* 51 (1): 129–151.

Jowell, Roger and Patricia Prescott-Clarke (1970): Racial Discrimination and White-Collar Workers in Britain. In: *Race and Class,* 11 (4): 397–417.

Juhn, Chinhui (2003): Labor Market Dropouts and Trends in the Wages of Black and White Men. In: *Industrial and Labor Relations Review,* 56 (4): 643–662.

Juhn, Chinhui; Murphy, Kevin M. and Brooks Pierce (1991): Accounting for the Slowdown in Black-White Wage Convergence. In: Kosters, Marvin H. (Ed.): 107–199.

Jurajda, Štěpán and Daniel Münich (2011): Gender Gap in Performance under Competitive Pressure: Admissions to Czech Universities. In: *The American Economic Review,* 101 (3): 514–518.

Kaas, Leo and Christian Manger (2012): Ethnic Discrimination in Germany's Labour Market: A Field Experiment. In: *German Economic Review,* 13 (1): 1–20.

Kahn, Lawrence M. (1991): Discrimination in Professional Sports: A Survey of the Literature. In: *Industrial and Labor Relations Review,* 44 (3): 395–418.

Kahn, Lawrence M. (2009): The Economics of Discrimination: Evidence from Basketball. Institute for the Study of Labor (IZA). Bonn. Discussion Paper, No. 3987.

Kahn, Lawrence M. and Peter D. Sherer (1988): Racial Differences in Professional Basketball Players' Compensation. In: *Journal of Labor Economics,* 6 (1): 40–61.

Kalev, Alexandra; Dobbin, Frank and Erin Kelly (2006): Best Practices or Best Guesses? Assessing the Efficacy of Corporate Affirmative Action and Diversity Policies. In: *American Sociological Review,* 71 (4): 589–617.

Kalter, Frank (1999): Ethnische Kundenpräferenz im professionellen Sport? Der Fall der Fußballbundesliga. In: *Zeitschrift für Soziologie,* 28 (3): 219–234.

Kalter, Frank (2002): Demographic Change, Educational Expansion, and Structural Assimilation of Immigrants: The Case of Germany. In: *European Sociological Review,* 18 (2): 199–216.

Kalter, Frank (2008): Ethnische Ungleichheit auf dem Arbeitsmarkt. In: Abraham, Martin and Thomas Hinz (Eds.): 303-332.

Kalter, Frank and Nadia Granato (2002): Ethnic Minorities' Education and Occupational Attainment: The Case of Germany. Mannheimer Zentrum für Europäische Sozialforschung (MZES). Mannheim. Working Papers, No. 58, 2002.

Kenney, Genevieve M. and Douglas A. Wissoker (1994): An Analysis of the Correlates of Discrimination Facing Young Hispanic Job-Seekers. In: *The American Economic Review,* 84 (3): 674–683.

Kilbourne, Barbara; England, Paula and Kurt Beron (1994): Effects of Individual, Occupational, and Industrial Characteristics on Earnings: Intersections of Race and Gender. In: *Social Forces,* 72 (4): 1149–1176.

Kim, Moon-Kak and Solomon W. Polachek (1994): Panel Estimates of Male-Female Earnings Functions. In: *The Journal of Human Resources,* 29 (2): 406–428.

Kim, Seik (2012): Statistical Discrimination, Employer Learning, and Employment Differentials by Race, Gender, and Education. Department of Economics, University of Washington. Washington D.C.

King, Eden B. and Afra S. Ahmad (2010): Experimental Field Study of Interpersonal Discrimination toward Muslim Job Applicants. In: *Personnel Psychology,* 63 (4): 881–906.

King, Mary C. (1992): Occupational Segregation by Race and Sex, 1940-88. In: *Monthly Labor Review,* 115 (4): 30–39.

Knowles, John; Persico, Nicola and Petra Todd (2001): Racial Bias in Motor Vehicle Searches: Theory and Evidence. In: *Journal of Political Economy,* 109 (1): 203–229.

Koedel, Cory and Eric Tyhurst (2012): Math Skills and Labor-Market Outcomes: Evidence from a Resume-Based Field Experiment. In: *Economics of Education Review,* 31 (1): 131–140.

Kogan, Irena (2004): Last Hired, First Fired? The Unemployment Dynamics of Male Immigrants in Germany. In: *European Sociological Review,* 20 (5): 445–461.

Kogan, Irena (2007): A Study of Immigrants' Employment Careers in West Germany Using the Sequence Analysis Technique. In: *Social Science Research,* 36 (2): 491–511.

Korenman, Sanders and David Neumark (1992): Marriage, Motherhood, and Wages. In: *The Journal of Human Resources,* 27 (2): 233–255.

Krause, Annabelle; Rinne, Ulf and Klaus F. Zimmermann (2010): Anonymisierte Bewerbungsverfahren. Institute for the Study of Labor (IZA). Bonn. Research Report, No. 27.

Krause, Annabelle; Rinne, Ulf and Klaus F. Zimmermann (2012a): Anonymous Job Applications of Fresh Ph.D. Economists. In: *Economics Letters,* 117 (2): 441–444.

Krause, Annabelle; Rinne, Ulf; Zimmermann, Klaus F.; Böschen, Ines and Ramona Alt (2012b): Pilotprojekt Anonymisierte Bewerbungsverfahren. Institute for the Study of Labor (IZA). Bonn. Research Report, No. 44.

Kroh, Martin (2012): Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (SOEP) (1984 until 2011). German Institute for Economic Research (DIW). Berlin. Research Report, No. 66.

Kuhn, Peter J. and Kailing Shen (2013): Gender Discrimination in Job Ads: Evidence from China. In: *The Quarterly Journal of Economics,* 128 (1): 287–336.

Kunze, Astrid and Kenneth R. Troske (2009): Life-Cycle Patterns in Male/Female Differences in Job Search. Institute for the Study of Labor (IZA). Bonn. Discussion Paper, No. 4656.

Ladd, Helen F. (1998): Evidence on Discrimination in Mortgage Lending. In: *The Journal of Economic Perspectives,* 12 (2): 41–62.

Lahey, Joanna N. (2008): Age, Women, and Hiring. In: *The Journal of Human Resources,* 43 (1): 30–56.

Lang, Kevin (1986): A Language Theory of Discrimination. In: *The Quarterly Journal of Economics,* 101 (2): 363–382.

Lang, Kevin and Jee-Yeon K. Lehmann (2012): Racial Discrimination in the Labor Market: Theory and Empirics. In: *Journal of Economic Literature,* 50 (4): 959–1006.

Lang, Kevin and Michael Manove (2011): Education and Labor Market Discrimination. In: *The American Economic Review,* 101 (4): 1467–1496.

Lang, Kevin; Manove, Michael and William T. Dickens (2005): Racial Discrimination in Markets with Announced Wages. In: *The American Economic Review,* 95 (4): 1327–1340.

Lazear, Edward P. and Sherwin Rosen (1990): Male-Female Wage Differentials in Job Ladders. In: *Journal of Labor Economics,* 8 (1): S106-S123.

Lee, Jungmin and Sokbae Lee (2012): Does It Matter Who Responded to the Survey - Trends in the U.S. Gender Earnings Gap Revisited. In: *Industrial and Labor Relations Review,* 65 (1): 148–160.

Lehmer, Florian and Johannes Ludsteck (2011): The Immigrant Wage Gap in Germany: Are East Europeans Worse Off? In: *International Migration Review,* 45 (4): 872–906.

Lehmer, Florian and Johannes Ludsteck (2012): Wage Assimilation of Immigrants: Which Factors Close the Gap? – Evidence from Germany. German Institute for Employment Research (IAB). Nuremberg.

Levinson, Richard M. (1975): Sex Discrimination and Employment Practices: An Experiment with Unconventional Job Inquiries. In: *Social Problems,* 22 (4): 533–543.

Liao, Tim F. (1994): Interpreting Probability Models - Logit, Probit, and Other Generalized Linear Models. Thousand Oaks, California: Sage Publications.

Lips, Hilary M. (2013): The Gender Pay Gap: Challenging the Rationalizations. Perceived Equity, Discrimination, and the Limits of Human Capital Models. In: *Sex Roles,* 68 (3-4): 169–185.

List, John A. (2004): The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field. In: *The Quarterly Journal of Economics,* 119 (1): 49–89.

López Bóo, Florencia; Rossi, Martín A. and Sergio S. Urzúa (2013): The Labor Market Return to an Attractive Face: Evidence from a Field Experiment. In: *Economics Letters,* 118 (1): 170–172.

Lundberg, Shelly J. and Richard Startz (1983): Private Discrimination and Social Intervention in Competitive Labor Market. In: *The American Economic Review,* 73 (3): 340–347.

Luthra, Renee R. (2013): Explaining Ethnic Inequality in the German Labor Market: Labor Market Institutions, Context of Reception, and Boundaries. In: *European Sociological Review*: 1–13.

Macpherson, David A. and Barry T. Hirsch (1995): Wages and Gender Composition: Why Do Women's Jobs Pay Less? In: *Journal of Labor Economics,* 13 (3): 426–471.

Madden, Janice F. (1987): Gender Differences in the Cost of Displacement: An Empirical Test of Discrimination in the Labor Market. In: *The American Economic Review,* 77 (2): 246–251.

Maurer-Fazio, Margaret (2012): Ethnic Discrimination in China's Internet Job Board Labor Market. In: *IZA Journal of Migration,* 1 (1): 1–24.

Maxwell, Nan L. (1993): The Effect on Black-White Wage Differences of Differences in the Quantity and Quality of Education. In: *Industrial and Labor Relations Review,* 47 (2): 249–264.

McFadden, Daniel (1974): Conditional Logit Analysis of Qualitative Choice Behavior. In: Zarembka, Paul (Ed.): 105–142.

McGinnity, Frances and Peter D. Lunn (2011): Measuring Discrimination Facing Ethnic Minority Job Applicants: An Irish Experiment. In: *Work, Employment and Society,* 25 (4): 693–708.

McIntosh, Neil and David J. Smith (1974): The Extent of Racial Discrimination. London: P.E.P.

Melly, Blaise (2005): Public-Private Sector Wage Differentials in Germany: Evidence from Quantile Regression. In: *Empirical Economics,* 30 (2): 505–520.

Miller, Paul W. (1987): The Wage Effect of the Occupational Segregation of Women in Britain. In: *The Economic Journal,* 97 (388): 885–896.

Mincer, Jacob A. (1974): Schooling, Experience, and Earnings. New York, New York: National Bureau of Economic Research.

Mincer, Jacob A. and Solomon W. Polachek (1974): Family Investments in Human Capital: Earnings of Women. In: *Journal of Political Economy,* 82 (2): 76–110.

Mobius, Markus M. and Tanya S. Rosenblat (2006): Why Beauty Matters. In: *The American Economic Review,* 96 (1): 222–235.

Mohrenweiser, Jens and Thomas Zwick (2009): Why Do Firms Train Apprentices? The Net Cost Puzzle Reconsidered. In: *Labour Economics,* 16 (6): 631–637.

Müller, Gerrit and Erik Plug (2006): Estimating the Effect of Personality on Male and Female Earnings. In: *Industrial and Labor Relations Review,* 60 (1): 3–22.

Mulligan, Casey B. and Yona Rubinstein (2008): Selection, Investment, and Women's Relative Wages Over Time. In: *The Quarterly Journal of Economics,* 123 (3): 1061–1110.

Munnell, Alicia H.; Tootell, Geoffrey M. B.; Browne, Lynn E. and James McEneaney (1996): Mortgage Lending in Boston: Interpreting HMDA Data. In: *The American Economic Review,* 86 (1): 25–53.

Neal, Derek A. and William R. Johnson (1996): The Role of Premarket Factors in Black-White Wage Differences. In: *Journal of Political Economy,* 104 (5): 869–895.

Neumark, David (1988): Employers' Discriminatory Behavior and the Estimation of Wage Discrimination. In: *The Journal of Human Resources,* 23 (3): 279–295.

Neumark, David (1996): Sex Discrimination in Restaurant Hiring: An Audit Study. In: *The Quarterly Journal of Economics,* 111 (3): 915–941.

Neumark, David (1999): Wage Differentials by Race and Sex: The Roles of Taste Discrimination and Labor Market Information. In: *Industrial Relations,* 38 (3): 414–445.

Neumark, David (2012): Detecting Discrimination in Audit and Correspondence Studies. In: *The Journal of Human Resources,* 47 (4): 1128–1157.

Newman, Jerry M. (1978): Discrimination in Recruitment: An Empirical Analysis. In: *Industrial and Labor Relations Review,* 32 (1): 15–23.

Niederalt, Michael (2005): Bestimmungsgründe des betrieblichen Ausbildungsverhalten in Deutschland. Lehrstuhl für Arbeitsmarkt- und Regionalpolitik, University of Erlangen-Nuremberg. Nuremberg. Discussion Paper, No. 36.

Nunes, Ana P. and Ben Seligman (2000): A Study of the Treatment of Female and Male Applicants by San Francisco Bay Area Auto Service Shops. Discrimination Research Center of The Impact Fund. Berkeley, California.

Oaxaca, Ronald L. (1973): Male-Female Wage Differentials in Urban Labor Markets. In: *International Economic Review,* 14 (3): 693–709.

Oaxaca, Ronald L. and Michael R. Ransom (1994): On Discrimination and the Decomposition of Wage Differentials. In: *Journal of Econometrics,* 61 (1): 5–21.

Oberholzer-Gee, Felix (2008): Nonemployment Stigma as Rational Herding: A Field Experiment. In: *Journal of Economic Behavior & Organization,* 65 (1): 30–40.

O'Neill, June (1990): The Role of Human Capital in Earnings Differences between Black and White Men. In: *The Journal of Economic Perspectives,* 4 (4): 25–45.

O'Neill, June and Solomon W. Polachek (1993): Why the Gender Gap in Wages Narrowed in the 1980s. In: *Journal of Labor Economics,* 11 (1): 205–228.

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

Pager, Devah (2003): The Mark of a Criminal Record. In: *American Journal of Sociology,* 108 (5): 937–975.

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

Pager, Devah; Bonikowski, Bart and Bruce Western (2009): Discrimination in a Low-Wage Labor Market: A Field Experiment. In: *American Sociological Review,* 74 (5): 777–799.

Pager, Devah and Lincoln Quillian (2005): Walking the Talk? What Employers Say versus What They Do. In: *American Sociological Review,* 70 (3): 355–380.

Pager, Devah and Hana Shepherd (2008): The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets. In: *Annual Review of Sociology,* 34 (1): 181–209.

Pendakur, Krishna and Simon Woodcock (2010): Glass Ceilings or Glass Doors? Wage Disparity within and between Firms. In: *Journal of Business & Economic Statistics,* 28 (1): 181–189.

Petersen, Trond and Ishak Saporta (2004): The Opportunity Structure for Discrimination. In: *American Journal of Sociology,* 109 (4): 852–901.

Petersen, Trond; Saporta, Ishak and Marc-David L. Seidel (2000): Offering a Job: Meritocracy and Social Networks. In: *American Journal of Sociology,* 106 (3): 763–816.

Petit, Pascale (2007): The Effects of Age and Family Constraints on Gender Hiring Discrimination: A Field Experiment in the French Financial Sector. In: *Labour Economics,* 14 (3): 371–391.

Pfeifer, Christian and Tatjana Sohr (2009): Analysing the Gender Wage Gap (GWG) Using Personnel Records. In: *Labour,* 23 (2): 257–282.

Phelps, Edmund S. (1972): The Statistical Theory of Racism and Sexism. In: *The American Economic Review,* 62 (4): 659–661.

Pinkston, Joshua C. (2003): Screening Discrimination and the Determinants of Wages. In: *Labour Economics,* 10 (6): 643–658.

Pinkston, Joshua C. (2006): A Test of Screening Discrimination with Employer Learning. In: *Industrial and Labor Relations Review,* 59 (2): 267–284.

Piore, Michael J. (1979): Birds of Passage - Migrant Labor and Industrial Societies. Cambridge, Massachusetts: Cambridge University Press.

Polachek, Solomon W. (1981): Occupational Self-Selection: A Human Capital Approach to Sex Differences in Occupational Structure. In: *The Review of Economics and Statistics,* 63 (1): 60–69.

Pope, Devin G.; Price, Joseph and Justin Wolfers (2011): Awareness Reduces Racial Bias. The Booth School, University of Chicago; Brigham Young University; The Wharton School, University of Pennsylvania. Chicago, Illinois.

Pope, Devin G. and Justin R. Sydnor (2011): What's in a Picture? In: *The Journal of Human Resources,* 46 (1): 53–92.

Ransom, Michael R. and Ronald L. Oaxaca (2010): New Market Power Models and Sex Differences in Pay. In: *Journal of Labor Economics,* 28 (2): 267–289.

Reimers, Cordelia W. (1983): Labor Market Discrimination against Hispanic and Black Men. In: *The Review of Economics and Statistics,* 65 (4): 570–579.

Riach, Peter A. and Judith Rich (1987): Testing for Sexual Discrimination in the Labour Market. In: *Australian Economic Papers,* 26 (49): 165–178.

Riach, Peter A. and Judith Rich (1991): Testing for Racial Discrimination in the Labour Market. In: *Cambridge Journal of Economics* (15): 239–256.

Riach, Peter A. and Judith Rich (2002): Field Experiments of Discrimination in the Market Place. In: *The Economic Journal,* 112 (483): 480–518.

Riach, Peter A. and Judith Rich (2004): Deceptive Field Experiments of Discrimination: Are They Ethical? In: *Kyklos,* 57 (3): 457–470.

Riach, Peter A. and Judith Rich (2006a): An Experimental Investigation of Age Discrimination in the French Labour Market. Institute for the Study of Labor (IZA). Bonn. Discussion Paper, No. 2522.

Riach, Peter A. and Judith Rich (2006b): An Experimental Investigation of Sexual Discrimination in Hiring in the English Labor Market. In: *Advances in Economic Analysis and Policy,* 6 (2): 1–20.

Riach, Peter A. and Judith Rich (2007a): An Experimental Investigation of Age Discrimination in the English Labor Market. Institute for the Study of Labor (IZA). Bonn. Discussion Paper, No. 3029.

Riach, Peter A. and Judith Rich (2007b): An Experimental Investigation of Age Discrimination in the Spanish Labour Market. Institute for the Study of Labor (IZA). Bonn. Discussion Paper, No. 2654.

Riphahn, Regina T. (2003): Cohort Effects in the Educational Attainment of Second Generation Immigrants in Germany: An Analysis of Census Data. In: *Journal of Population Economics,* 16 (4): 711–737.

Rodgers, William M. and William E. Spriggs (1996): What Does the AFQT Really Measure: Race, Wages, Schooling. In: *The Review of Black Political Economy,* 24 (4): 13–46.

Rooth, Dan-Olof (2002): Adopted Children in the Labour Market - Discrimination or Unobserved Characteristics? In: *International Migration,* 40 (1): 71–98.

Rooth, Dan-Olof (2009): Obesity, Attractiveness and Differential Treatment in Hiring - A Field Experiment. In: *The Journal of Human Resources,* 44 (3): 710–735.

Rooth, Dan-Olof (2010): Automatic Associations and Discrimination in Hiring: Real World Evidence. In: *Labour Economics,* 17 (3): 523–534.

Rooth, Dan-Olof (2011): Work Out or Out of Work — The Labor Market Return to Physical Fitness and Leisure Sports Activities. In: *Labour Economics,* 18 (3): 399–409.

Rosen, Asa (1997): An Equilibrium Search-Matching Model of Discrimination. In: *European Economic Review,* 41 (8): 1589–1613.

Ross, Stephen A. (1973): The Economic Theory of Agency: The Principal's Problem. In: *The American Economic Review,* 63 (2): 134–139.

Ross, Stephen L. and Margery A. Turner (2005): Housing Discrimination in Metropolitan America: Explaining Changes between 1989 and 2000. In: *Social Problems,* 52 (2): 152–180.

Rudolph, Udo; Böhm, Robert and Michaela Lummer (2007): Ein Vorname sagt mehr als 1000 Worte - Zur sozialen Wahrnehmung von Vornamen. In: *Zeitschrift für Sozialpsychologie,* 38 (1): 17–31.

Rudolph, Udo and Matthias Spörrle (1999): Alter, Attraktivität und Intelligenz von Vornamen: Wortnormen für Vornamen im Deutschen. In: *Experimental Psychology,* 46 (2): 115–128.

Ruffle, Bradley J. and Ze'ev Shtudiner (2013): Are Good-Looking People More Employable? Department of Economics, Ben-Gurion University. Beer Sheva.

Sasaki, Masaru (1999): An Equilibrium Search Model with Coworker Discrimination. In: *Journal of Labor Economics,* 17 (2): 377–407.

Schmidt, Christoph M. (1997): Immigrant Performance in Germany: Labor Earnings of Ethnic German Migrants and Foreign Guest-workers. In: *The Quarterly Review of Economics and Finance,* 37 (Supplement 1): 379–397.

Schweitzer, Linda; Ng, Eddy; Lyons, Sean and Lisa Kuron (2011): Exploring the Career Pipeline: Gender Differences in Pre-Career Expectations. In: *Industrial Relations,* 66 (3): 422–444.

Scott Morton, Fiona; Zettelmeyer, Florian and Jorge Silva-Risso (2003): Consumer Information and Discrimination: Does the Internet Affect the Pricing of New Cars to Women and Minorities? In: *Quantitative Marketing and Economics,* 1 (1): 65–92.

Segendorf, Åsa O. and Dan-Olof Rooth (2006): Wage Effects of Search Methods for Immigrants and Natives: The Case of Sweden. European Society for Population Economics (ESPE). Verona.

Siddique, Zahra (2011): Evidence on Caste Based Discrimination. In: *Labour Economics,* 18 (Supplement 1): 146–159.

Silber, Jacques and Michal Weber (1999): Labour Market Discrimination: Are There Significant Differences between the Various Decomposition Procedures? In: *Applied Economics,* 31 (3): 359–365.

Siniver, Erez (2011): Testing for Statistical Discrimination: The Case of Immigrant Physicians in Israel. In: *Labour,* 25 (2): 155–166.

Smith, James P. and Finis R. Welch (1989): Black Economic Progress after Myrdal. In: *Journal of Economic Literature,* 27 (2): 519–564.

Sorensen, Elaine (1990): The Crowding Hypothesis and Comparable Worth. In: *The Journal of Human Resources,* 25 (1): 55–89.

Spence, Michael (1973): Job Market Signaling. In: *The Quarterly Journal of Economics,* 87 (3): 355–374.

Stoll, Michael A.; Raphael, Steven and Harry J. Holzer (2004): Black Job Applicants and the Hiring Officer's Race. In: *Industrial and Labor Relations Review,* 57 (2): 267–287.

Szymanski, Stefan (2000): A Market Test for Discrimination in the English Professional Soccer Leagues. In: *Journal of Political Economy,* 108 (3): 590–603.

Tajfel, Henri (1982): Social Identity and Intergroup Relations. New York, New York: Cambridge University Press.

Tam, Tony (1997): Sex Segregations and Occupational Gender Inequality in the United States: Devaluation or Specialized Training? In: *American Journal of Sociology,* 102 (6): 1652–1692.

Tam, Tony (2000): Occupational Wage Inequality and Devaluation: A Cautionary Tale of Measurement Error. In: *American Journal of Sociology,* 105 (6): 1752–1760.

Telles, Edward E. and Edward Murguia (1990): Phenotypic Discrimination and Income Differences among Mexican Americans. In: *Social Science Quarterly,* 71 (4): 682–696.

The Bundestag (2002): Schlussbericht der Enquête-Kommission „Demographischer Wandel – Herausforderungen unserer älter werdenden Gesellschaft an den Einzelnen und die Politik". Berlin.

The Bundestag (2005): Vocational Training Act (BBiG).

The Bundestag (2006): General Act on Equal Treatment (AGG).

Theunissen, Gert; Verbruggen, Marijke; Forrier, Anneleen and Luc Sels (2011): Career Sidestep, Wage Setback? The Impact of Different Types of Employment Interruptions on Wages. In: *Gender, Work and Organization,* 18 (1): 110–131.

Uhlendorff, Arne and Klaus F. Zimmermann (2006): Unemployment Dynamics among Migrants and Natives. Institute for the Study of Labor (IZA). Bonn. Discussion Paper, No. 2299.

Urban, Dieter (1993): Logit-Analyse - Statistische Verfahren zur Analyse von Modellen mit qualitativen Response-Variablen. Stuttgart: Fischer.

Velling, Johannes (1995): Wage Discrimination and Occupational Segregation of Foreign Male Workers in Germany. Zentrum für Europäische Wirtschaftsforschung (ZEW). Mannheim. Discussion Papers, No. 95-04.

Waldfogel, Jane (1997): The Effect of Children on Women's Wages. In: *American Sociological Review,* 62 (2): 209–217.

Waldfogel, Jane (1998): Understanding the "Family Gap" in Pay for Women with Children. In: *The Journal of Economic Perspectives,* 12 (1): 137–156.

Watson, Stevie; Appiah, Osei and Corliss G. Thornton (2011): The Effect of Name on Pre-Interview Impressions and Occupational Stereotypes: The Case of Black Sales Job Applicants. In: *Journal of Applied Social Psychology,* 41 (10): 2405–2420.

Weber, Andrea and Christine Zulehner (2009): Competition and Gender Prejudice: Are Discriminatory Employers Doomed to Fail? Center for Economic Studies and Institute for Economic Research (CESifo). Munich. Working Paper, No. 2842.

Weichselbaumer, Doris (2004): Is It Sex or Personality - The Impact of Sex Stereotypes on Discrimination in Applicant Selection. In: *Eastern Economic Journal,* 30 (2): 159–186.

Weichselbaumer, Doris (2013): Testing for Discrimination against Lesbians of Different Marital Status - A Field Experiment. Department of Economics, University of Linz. Linz. Working Paper, No. 1308.

Weichselbaumer, Doris and Rudolf Winter-Ebmer (2005): A Meta-Analysis of the International Gender Wage Gap. In: *Journal of Economic Surveys,* 19 (3): 479–511.

Weinberger, Catherine J. (2011): In Search of the Glass Ceiling: Gender and Earnings Growth among U.S. College Graduates in the 1990s. In: *Industrial and Labor Relations Review,* 64 (5): 949–980.

Weitzel, Tim; Eckhardt, Andreas; Laumer, Sven and Alexander von Stetten (2011a): Recruiting Trends 2011 - Eine empirische Untersuchung mit den Top-1.000-Unternehmen aus Deutschland sowie den Top-300-Unternehmen aus den Branchen Finanzdienstleistung, IT und Öffentlicher Dienst. Centre of Human Resources Information Systems (CHRIS); University of Bamberg; University of Frankfurt on the Main; Monster Worldwide Deutschland GmbH. Bamberg/ Frankfurt am Main.

Weitzel, Tim; Eckhardt, Andreas; Laumer, Sven and Alexander von Stetten (2011b): Recruiting Trends im Mittelstand 2011 - Eine empirische Untersuchung mit 1.000 Unternehmen aus dem deutschen Mittelstand. Centre of Human Resources Information Systems (CHRIS); University of Bamberg; University of Frankfurt on the Main; Monster Worldwide Deutschland GmbH. Bamberg/ Frankfurt am Main.

Wenzelmann, Felix (2012): Ausbildungsmotive und die Zeitaufteilung der Auszubildenden im Betrieb. In: *Journal for Labour Market Research,* 45 (2): 125–145.

Western, Bruce and Becky Pettit (2005): Black-White Wage Inequality, Employment Rates, and Incarceration. In: *American Journal of Sociology,* 111 (2): 553–578.

Wood, Martin; Hales, Jon; Purdon, Susan; Sejersen, Tanja and Oliver Hayllar (2009): A Test for Racial Discrimination in Recruitment Practice in British Cities. Department for Work and Pensions. Norwich. Research Report, No. 607.

Wooldridge, Jeffrey M. (2009): Introductory Econometrics - A Modern Approach. 4th ed. Mason, Ohio, London: South-Western.

Wooldridge, Jeffrey M. (2010): Econometric Analysis of Cross Section and Panel Data. 2nd ed. Cambridge, Massachusetts: MIT Press.

Wozniak, Abigail (2012): Discrimination and the Effects of Drug Testing on Black Employment. Institute for the Study of Labor (IZA). Bonn. Discussion Paper, No. 6605.

Yinger, John (1986): Measuring Racial Discrimination with Fair Housing Audits: Caught in the Act. In: *The American Economic Review,* 76 (5): 881–893.

Zibrowius, Michael (2012): Convergence or Divergence? Immigrant Wage Assimilation Patterns in Germany. German Institute for Economic Research (DIW). Berlin. SOEPpapers, No. 479-2012.

APPENDIX

A. OVERVIEW OF EMPIRICAL FINDINGS FROM CORRESPONDENCE STUDIES

Table A-1: A Partial List of Correspondence Studies Investigating Gender Discrimination

Author(s) and year of publication	Location	Time period	Occupation	No. of job offers addressed	Callback rate		
					Men	**Women**	**Difference**
Carlsson (2011)	Sweden (Stockholm, Gothenburg)	05/2005-02/2006	Computer professional	106	0.22	0.23	-0.01
			Motor-vehicle driver	78	0.24	0.21	0.03
			Construction worker	64	0.30	0.20	0.10
			Business sales assistant	278	0.35	0.41	-0.06**
			Lower secondary school teacher (language)	60	0.47	0.47	0.00
			Upper secondary school teacher	64	0.33	0.3	0.03
			Restaurant worker	140	0.08	0.19	-0.11***
			Accountant	186	0.13	0.21	-0.08***
			Cleaner	62	0.08	0.11	-0.03
			Preschool teacher	184	0.61	0.67	-0.06
			Shop sales assistant	200	0.15	0.15	0.00
			Lower secondary school teacher (math and science)	42	0.57	0.55	0.02
			Nurse	150	0.33	0.29	0.04
Albert et al. (2011)[1]	Spain (Madrid)	10/05-11/05 & 01/06-06/06	Sales representative	1,130	0.17	0.16	0.01
			Marketing technician	1,080	0.02	0.02	0.00
			Accountant assistant	990	0.08	0.11	-0.03***
			Accountant	830	0.06	0.07	-0.01
			Administrative assistant/receptionist	880	0.03	0.10	-0.07***
			Executive secretary	400	0.05	0.16	-0.11***
Booth and Leigh (2010)[1]	Australia (Brisbane, Melbourne, Sydney)	04/07-10/07	Waitstaff	863	0.30	0.40	-0.10***
			Data-entry	851	0.19	0.33	-0.14***
			Customer service	832	0.26	0.29	-0.03
			Sales	819	0.25	0.26	-0.01
Riach and Rich (2006b)	U.K. (London)	N/A	Chartered accountant	339	0.10	0.13	-0.03*
			Computer analyst programmer	130	0.14	0.23	-0.09***
			Engineer	173	0.17	0.12	0.05*
			Secretary	231	0.09	0.19	-0.10***
Weichselbaumer (2004)	Austria (Vienna)	Early 1998 – fall 1999	Network technician	117	0.73	0.58	0.15***
			Computer programmer	88	0.82	0.81	0.01
			Accountant	149	0.40	0.43	-0.03
			Secretary	123	0.20	0.44	-0.24***
Neumark (1996)	U.S. (Philadelphia)	N/A	High-priced restaurants	23	0.61	0.26	0.35**
			Medium-priced restaurants	21	0.62	0.43	0.19
			Low-priced restaurants	21	0.19	0.38	-0.19
Riach and Rich (1987)	Australia (State of Victoria)	11/1983–11/1986	Computer analyst	152	0.57	0.50	0.07**
			Computer operator	99	0.43	0.41	0.02
			Computer programmer	115	0.52	0.56	-0.03
			Gardener	148	0.39	0.32	0.07**
			Industrial relations officer	94	0.33	0.35	-0.02
			Management accountant	211	0.46	0.43	0.04
			Payroll clerk	172	0.41	0.42	-0.01
Levinson (1975)	U.S. (Atlanta)	Spring 1974	Female-dominated job	110	N/A	N/A	-0.44***
			Male-dominated job	146	N/A	N/A	0.28***

Notes: [1] As no information on the number of matched-pairs is available, number of single applications is reported. * denotes 10% significance level, ** denotes 5% significance level and *** denotes 1% significance level of a chi-squared test that the male and female candidates are equally likely to receive a callback at any matched-pair application.

Table A-2: A Partial List of Correspondence Studies Investigating Ethnic Discrimination

Author(s) and year of publication	Location	Time period	Occupation(s)	Minority group(s)	No. of job offers addressed	Callback rate: Natives	Callback rate: Ethnic minorities	Callback rate: Difference
Baert et al. (2013)	Belgium (Flanders)	11/2011-03/2012	Bottleneck occupations	Turks	181	0.17	0.17	0.00
			Non-bottleneck occupations		195	0.21	0.10	0.11***
Andriessen et al. (2012)	The Netherlands	05/2008-12/2008	62 high- and low-skilled professions in 5 sectors	Moroccans	323	0.51	0.46	0.05**
				Turks	338	0.49	0.42	0.07**
				Surinamese	356	0.42	0.34	0.08***
				Antilleans	323	0.42	0.36	0.06**
Maurer-Fazio (2012)	China (6 different regions)	Summer 2011	Accountants, administrative assistants, sales representatives	Mongolians	3,594	0.08	0.06	0.02***
				Tibetans	3,548	0.08	0.04	0.04***
				Uighurs	3,654	0.08	0.04	0.04***
Arai et al. (2011)	Sweden (Stockholm)	03/2006-07/2007	Computer specialists, drivers, accountants, senior high school teachers, assistant nurses	Arabs (Men)	374	0.42	0.23	0.19***
				Arabs (Women)	192	0.37	0.15	0.22***
Jacquemet and Yannelis (2012)	U.S. (Chicago)	08/2009-02/2010	Healthcare, accounting, IT	Black name	330	0.23	0.16	0.07***
				Foreign name	330	0.23	0.16	0.07***
McGinnity and Lunn (2011)	Ireland (Dublin)	03/2008-10/2008	Accountancy, lower administration, retail sales	Africans	81	0.27	0.11	0.16***
				Asians	80	0.34	0.19	0.15**
				Germans	79	0.37	0.18	0.19***
Booth et al. (2012)	Australia (Brisbane, Melbourne, Sydney)	04/2007-10/2007	Waitstaff, data entry, customer service, sales jobs	Middle Easterners	845	0.35	0.22	0.13***
				Native Australians	848	0.35	0.26	0.09***
				Italians	835	0.35	0.32	0.03*
				Chinese	845	0.35	0.21	0.14***
Carlsson (2010)	Sweden (Stockholm, Gothenburg)	08/2006-04/2007	Shop sales assistants, construction workers, restaurant workers, motor vehicle drivers, accountants, 4 types of teachers, business sales assistants, computer professionals, nurses	Middle Easterners (1st gen.)	1,314	0.41	0.20	0.21***
				Middle Easterners (2nd gen.)	1,314	0.41	0.24	0.17***
Kaas and Manger (2012)	Germany	12/2007-01/2008, 12/2008	Management and economics student internships	Turks (2nd gen.)	528	0.40	0.35	0.05*
Oreopoulos (2011)	Canada (Toronto)	04/2008-11/2008	Administrative, finance, marketing, sales, programmer, retail	Indians	328	0.16	0.05	0.11***[1]
				Chinese	302	0.16	0.05	0.11***[1]
				Pakistanis	187	0.16	0.05	0.11***[1]
				Brits	299	0.16	0.14	0.02[1]
Wood et al. (2009)	U.K. (Bradford, Bristol, Glasgow, Leeds, London, Manchester)	11/2008-05/2009	IT technicians, accountants, HR managers, teaching assistants, IT support, account clerks, office assistants, care assistants	Black Africans	400	0.13	0.08	0.05[2]
				Black Caribbeans	399	0.10	0.05	0.05[2]
				Chinese	393	0.10	0.06	0.04[2]
				Indians	393	0.11	0.06	0.04[2]
				Pakistani/ Bangladeshi	389	0.10	0.06	0.04[2]
Cediey and Foroni (2008)	France (Lille, Lyon, Nantes, Paris, Strasbourg)	End 2005-mid 2006	21 occupations in 10 sectors (e.g. hotel and restaurants, commerce, personal and community services, tourism and transport, management and administration)	North and Sub-Saharan Africans	694	0.27	0.10	0.17***

Carlsson and Rooth (2007)	Sweden (Stockholm, Gothenburg)	05/2005-02/2006	See Carlsson (2010)	Middle-Easterners	1,552	0.29	0.20	0.09***
Bursell (2007)	Sweden (Stockholm)	03/2006-09/2007	15 different occupations	Arabs and Africans	1,776	0.37	0.20	0.17***
Bertrand and Mullainathan (2004)	U.S. (Chicago, Boston)	07/2001-01/2002 (Boston), 07/2001-05/2002 (Chicago)	Sales, administrative support, clerical services, customer services	African-Americans	2,435	0.10	0.06	0.04***
Goldberg et al. (1996)	Germany (Berlin, Rhine-Ruhr region)	02/1994-N/A	11 occupations in 3 sectors (e.g. caring professions, commercial professions, technical professions	Turks (1st gen.)	2,633	0.10	0.09	0.01[2]
Bovenkerk et al. (1996)	The Netherlands (Randstad area)	10/1993-06/1994	Teachers, lab assistants, admin/ finance managers, personnel managers	Surinamese	290	0.46	0.36	0.10**
Bendick et al. (1991)	U.S. (Washington D.C.)	02/1992-03/1992	Sales, service and office jobs	Latinos	741	0.19	0.22	-0.03
Riach and Rich (1991)	Australia (State of Victoria)	11/1983-11/1988	Sales representatives, clerks, secretaries	Greeks	462	0.35	0.31	0.04[2]
				Vietnamese	519	0.29	0.20	0.09[2]
Firth (1981)	U.K.	10/1977-03/1978	Accounting and financial management jobs	Australians	282	0.85	0.75	0.10[2]
				Frenchmen	282	0.85	0.68	0.17[2]
				Africans	282	0.85	0.53	0.32[2]
				Indians	282	0.85	0.44	0.41[2]
				Pakistani	282	0.85	0.44	0.41[2]
				West Indians	282	0.85	0.48	0.37[2]
Jowell and Prescott-Clarke (1970)	U.K. (4 different regions)	Spring till summer 1969	Sales and marketing, accountancy and office management, electrical engineering, secretarial jobs	Australians	32	0.78	0.78	0.00
				West Indians	32	0.78	0.69	0.09[2]
				Cypriots	32	0.78	0.69	0.09[2]
				Asians	32	0.78	0.35	0.43[2]

Notes: [1] Results reported for immigrants with foreign education and work experience. [2] Level of significance not indicated. If not explicitly stated, callback rates are based on own calculations with information provided in the studies. Ethnic affiliation is generally signaled by names. * denotes 10% significance level, ** denotes 5% significance level and *** denotes 1% significance level of a chi-squared test that the native and ethnic minority candidates are equally likely to receive a callback at any matched-pair application.

B. SELECTED SAMPLE OF APPLICATIONS USED IN THE FIELD EXPERIMENTS

B.1 GERMAN-NAMED MALE APPLICANT

Cover Letter

Jan Lange
XXX
XXX

Employer's address
XXX
XXX

XXX, 25 May 2011

Application for an industrial mechanics apprenticeship

Dear Mr./Mrs. XXX,

I am writing to you in response to your advertisement, which appeared on the job platform of the Federal Employment Agency and directly caught my attention. Having collected further information on your firm as well as on the expertise required, I would like to apply for the offered apprenticeship since I will be shortly moving to your region.

I am currently in 10th grade of Secondary School from which I will graduate this summer. At school as well as in my free-time I pursue my passion for technology leading to excellent grades especially in the natural science subjects. To make use of my interests and abilities, I would like to put the focus of my professional career on this specific area. Therefore, I decided to apply for an apprenticeship in your company.

According to my friends and teachers, I am an attentive and ambitious person. Furthermore, I like facing new challenges and possess the ability to easily get in touch with other people. Due to my experiences from playing handball, I am aware of the significance of relying on other group members and reaching goals in a team.

I would be happy to be invited for an interview to personally convince you of my qualifications. I am looking forward to hearing from you.

Yours sincerely,
Jan Lange

Curriculum vitae

Jan Lange

XXX

XXX

Mobile: 0176-74684211

Email: janlange94@gmx.de

Personal Details

Date of Birth:	18 September 1994
Nationality:	German
Family Status:	Single

School Education

08/2005 - present	Secondary School Carl Theodor Ottmer , XXX
08/2001 – 07/2005	Primary School Humboldtstraße, XXX

Additional Skills

Languages:
- German as native language
- Good command of English

Computer Skills:
- Good knowledge in MS Word
- Basic skills in MS Excel and MS Powerpoint

Driving license:
- Category M

Leisure Time Activities

- Handball, running
- Building and extending railway models

XXX, 25 May 2011

B.2 FEMALE APPLICANT

Cover Letter

Anna Schneider
XXX
XXX

Employer's address
XXX
XXX

XXX, September 2011

Application for an apprenticeship as an industrial mechanics

Dear Mr./Mrs. XXX,

Your job offer posted on the job website of the Federal Employment Agency has called my attention and aroused my interest for your business and the apprenticeship as an industrial mechanic. After in-depth internet research on the professional requirements and on your company, I decided to send you this application.

Graduating this summer with the secondary education certificate, I intend to do a dual apprenticeship in a technical occupation. As my grades and the participation in the voluntary fire brigade show, my strengths and interests definitively cover this field. Additionally, first practical experiences have confirmed that doing technical work fascinates me and requires the skills and the understanding I possess.

According to my leisure time activities, I am a team player who knows that relying on each other is essential. Furthermore, I am a curious person and always open to minded. In addition to that, my work constantly shows great thoroughness.

Since I am planning to move to your region shortly after having completed school, I will be resident to and hence in direct reach of your company. With regard to the training program, I am sure that my willingness and commitment to acquire new skills will convince you. Therefore, I would be happy to presenting myself in a personal interview. I look forward to hearing from you.

Yours faithfully,
Anna Schneider

Curriculum Vitae

Personal Data

ANNA SCHNEIDER

XXX
XXX
Mobile: 0176-63009012
Mail: annaschneider95@gmx.net

Date of Birth:	September 3, 1995
Family Status:	Single
Nationality:	German

Schooling

Since 08/2006	Middle School, XXX
08/2002 – 07/2006	Primary School, XXX

Internships

02/2011	School internship at a machine tools producer

Other Qualifications and Extracurricular Activities

Languages	German: First language English: Good skills
Computer Skills	Good knowledge in Word Basic skills in Excel and Powerpoint
Driving License	Mopeds (Category M)
Leisure Time Activities	Voluntary fire brigade Table tennis

XXX, September 2011

B.3 TURKISH-NAMED MALE APPLICANT

Cover Letter

Kenan Yilmaz
XXX
XXX

Employer's address
XXX
XXX

XXX, September 2011

Application for an industrial mechanics apprenticeship

Dear Mr./Mrs. XXX,

The website of the Federal Employment Agency has drawn my attention to the training program for industrial mechanics offered by your company. The job profile and the tasks described sound very interesting to me and have convinced me to apply for an apprenticeship.

I will be shortly graduating from secondary school. As I have been interested in technical issues since my early childhood and especially like doing handicrafts and tinkering, I intend working in this specific field. At school I particularly enjoy following scientific courses. This pleasure has led to excellent grades and was also quite helpful when doing a school internship.

I am a very committed person that has a great willingness to learn new things and likes being challenged. Additionally, I am a reliable as well as aim-oriented person and like working in teams. Furthermore, friends and teachers appreciate my readiness to speak up for others and to always give a helping hand.

I look forward to attending a job interview in order to get further information on your company and to persuade you of my personal strengths. Although spatial distance to your company currently exists, I will soon be moving to your region with my family.

With kind regards,
Kenan Yilmaz

Curriculum Vitae

PERSONAL DATA

Kenan Yilmaz
XXX
XXX
0176-74688046
Kenanyilmaz95@gmx.de
September 10, 1995
Single

SCHOOL EDUCATION

Since 8/2006	Secondary School, XXX
8/2002 - 7/2006	Primary School, XXX

PRACTICAL EXPERIENCE

02/2011	School internship, XXX

ADDITIONAL SKILLS

Computer Skills:

Word	Excellent skills
Excel, Powerpoint	Good knowledge

Languages:

German	Native language
Turkish	Native language
English	Good command

Driving licence:
Category M (mopeds)

LEISURE ACTIVITIES

Playing tennis and bicycling
Tinkering with motor scooters

XXX, September 2011

C. Supplemental Descriptive Statistics and Regression Tables

C.1 Study on Gender Discrimination

Table C-1: Firms' Responses by Gender in Male-Dominated Jobs

	Male (N=540)	Female (N=540)	Total (N=1,080)	Difference
No response	20.00%	17.59%	18.80%	
	(108)	(95)	(203)	
Rejection	39.07%	47.96%	43.52%	
	(211)	(259)	(470)	
Callback	40.93%	34.44%	37.69%	6.49 pps**
	(221)	(186)	(407)	(35)

Notes: The table reports detailed responses by gender in male-dominated jobs as a fraction of overall applications in percent. Absolute numbers are in parentheses. ** denotes 5% significance level of a chi-squared test (H0: The male and female candidates are equally likely to receive a callback at any matched-pair application).

Table C-2: Marginal Effects from Probit Regressions on Response Dummy (Gender Study)

Response	(I)	(II)	(III)	(IV)
Female	0.019	0.019	0.019	0.019
	(0.014)	(0.014)	(0.014)	(0.014)
Medium		0.086***	0.088***	0.085***
		(0.033)	(0.033)	(0.033)
Large		0.121***	0.122***	0.120***
		(0.032)	(0.032)	(0.032)
South		0.008	0.003	0.006
		(0.042)	(0.042)	(0.042)
East		-0.105**	-0.109**	-0.085*
		(0.049)	(0.050)	(0.049)
Industry		0.001	0.000	0.003
		(0.038)	(0.038)	(0.038)
Late recruiter		0.054	0.031	0.016
		(0.044)	(0.067)	(0.065)
Female responsible		0.044	0.044	0.042
		(0.027)	(0.027)	(0.028)
Share of females t-1			0.010	-0.142
			(0.025)	(0.086)
Vacancies/total jobs t-1			0.004	0.006
			(0.015)	(0.015)
Certificate				0.015
				(0.024)
Female-dominated job				0.246***
				(0.075)
Controls	Yes	Yes	Yes	Yes
No. of obs.	1,312	1,312	1,312	1,312
Pseudo R^2	0.013	0.046	0.046	0.050
Log likelihood	-619.372	-598.300	-598.076	-595.741
Wald chi-squared	13.291	41.443	42.526	45.217
P-value	0.065	0.000	0.000	0.000

Notes: Each model reports average marginal effects of a probit regression on the response dummy (Y=1: employer gives the applicant either a rejection or a callback). Marginal effects are calculated at the means of all independent variables and denote an infinitesimal change in case of continuous variables and a discrete change in case of dummy variables. Standard errors clustered on firm level are in parentheses. Regressions consider the entire sample. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

Table C-3: Marginal Effects from Probit Regressions on Callback Dummy for Male Applicants

Callback	**(Ia)**	**(Ib)**	**(IIa)**	**(IIb)**	**(IIIa)**	**(IIIb)**
Certificate	0.068*	0.074	0.063	0.068	0.089	0.061
	(0.040)	(0.052)	(0.045)	(0.061)	(0.093)	(0.100)
Lukas Schmidt	0.063	0.041	0.019	-0.023	0.127	0.101
	(0.072)	(0.085)	(0.132)	(0.169)	(0.094)	(0.107)
Male photo B	-0.066	-0.025	0.160	0.115	-0.103	-0.085
	(0.068)	(0.083)	(0.155)	(0.181)	(0.092)	(0.102)
Distance	-0.000*	-0.000	-0.000***	-0.000	0.001**	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
Design B	0.014	0.018	0.034	0.040	-0.000	-0.008
	(0.044)	(0.045)	(0.051)	(0.052)	(0.094)	(0.102)
Design C	0.075	0.075	0.081	0.082	-/-	-/-
	(0.053)	(0.055)	(0.055)	(0.057)		
Rank 2	0.030	0.032	0.035	0.037	0.015	0.022
	(0.039)	(0.040)	(0.044)	(0.045)	(0.094)	(0.107)
Controls	No	Yes	No	Yes	No	Yes
No. of obs.	656	656	540	540	116	116
Pseudo R^2	0.012	0.032	0.020	0.040	0.054	0.141
Log likelihood	-437.018	-428.124	-358.205	-350.778	-72.315	-65.671
LR chi-squared	10.011	26.074	13.607	24.982	7.792	18.593
P-value	0.188	0.128	0.059	0.125	0.254	0.233

Notes: Each model reports average marginal effects of a probit regression on the callback dummy (Y=1: employer calls back the job applicant). Marginal effects are calculated at the means of all independent variables and denote an infinitesimal change in case of continuous variables and a discrete change in case of dummy variables. Robust standard errors are in parentheses. Regressions restrict the sample to male applicants. The models in (I) report the effects of all applications by the male candidate while models (II) and (III) show the results for male- and female-dominated jobs, respectively. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

Table C-4: Marginal Effects from Probit Regressions on Callback Dummy for Female Applicants

Callback	**(Ia)**	**(Ib)**	**(IIa)**	**(IIb)**	**(IIIa)**	**(IIIb)**
Certificate	0.033	-0.035	0.053	-0.022	-0.036	-0.103
	(0.040)	(0.049)	(0.044)	(0.058)	(0.097)	(0.102)
Laura Müller	-0.008	-0.018	-0.065	-0.078	0.044	-0.052
	(0.069)	(0.079)	(0.153)	(0.168)	(0.094)	(0.104)
Female photo B	0.059	0.079	-0.021	-0.036	0.085	0.150
	(0.072)	(0.083)	(0.168)	(0.180)	(0.094)	(0.100)
Distance	-0.000	0.000	-0.000	-0.000	0.001	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Design B	-0.055	-0.052	-0.071	-0.069	0.015	0.083
	(0.044)	(0.045)	(0.050)	(0.052)	(0.094)	(0.100)
Design C	-0.071	-0.055	-0.073	-0.059	-/-	-/-
	(0.050)	(0.052)	(0.054)	(0.055)		
Rank 2	-0.071*	-0.058	-0.059	-0.047	-0.166*	-0.148
	(0.042)	(0.042)	(0.047)	(0.047)	(0.096)	(0.104)
Controls	No	Yes	No	Yes	No	Yes
No. of obs.	656	656	540	540	116	116
Pseudo R^2	0.008	0.030	0.010	0.026	0.040	0.166
Log likelihood	-423.210	-413.929	-344.150	-338.566	-75.135	-65.315
LR chi-squared	7.225	24.575	6.899	17.198	5.864	24.561
P-value	0.406	0.175	0.439	0.510	0.439	0.056

Notes: Each model reports average marginal effects of a probit regression on the callback dummy (Y=1: employer calls back the job applicant). Marginal effects are calculated at the means of all independent variables and denote an infinitesimal change in case of continuous variables and a discrete change in case of dummy variables. Robust standard errors are in parentheses. Regressions restrict the sample to female applicants. The models in (I) report the effects of all applications by the female candidate while models (II) and (III) show the results for male- and female-dominated jobs, respectively. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

Table C-5: Marginal Effects from Probit Regressions on Callback Dummy for a Standard Applicant at a Standard Employer (Gender Study)

Callback	**(I)**	**(II)**	**(III)**	**(IV)**	**(V)**	**X**
Female	-0.051***	-0.051***	-0.052***	-0.051***	-0.071***	0
	(0.018)	(0.018)	(0.019)	(0.018)	(0.021)	
Medium		0.107***	0.108***	0.105**	0.107**	1
		(0.040)	(0.041)	(0.041)	(0.042)	
Large		0.081	0.081	0.079	0.080	0
		(0.065)	(0.065)	(0.065)	(0.065)	
South		-0.054	-0.045	-0.044	-0.042	1
		(0.056)	(0.059)	(0.059)	(0.059)	
East		0.061	0.067	0.067	0.068	0
		(0.056)	(0.057)	(0.058)	(0.059)	
Industry		-0.069	-0.070	-0.071	-0.071	1
		(0.053)	(0.054)	(0.054)	(0.054)	
Late recruiter		-0.014	-0.001	-0.001	-0.001	1
		(0.059)	(0.086)	(0.086)	(0.087)	
Female responsible		0.019	0.019	0.018	0.019	1
		(0.036)	(0.036)	(0.036)	(0.036)	
Share of females t-1			-0.004	-0.015	-0.016	0
			(0.032)	(0.120)	(0.122)	
Vacancies/total jobs t-1			-0.011	-0.011	-0.011	0
			(0.021)	(0.021)	(0.021)	
Certificate				0.026	0.025	0
				(0.033)	(0.033)	
Female-dominated job				0.032	-0.019	0
				(0.321)	(0.321)	
Female x Female-dominated job					0.107**	0
					(0.051)	
Controls	Yes	Yes	Yes	Yes	Yes	
No. of obs.	1,312	1,312	1,312	1,312	1,312	
Pseudo R^2	0.010	0.021	0.021	0.021	0.022	
Log likelihood	-861.957	-852.607	-852.331	-852.064	-851.026	
Wald chi-squared	17.315	29.007	29.341	30.279	35.429	
P-value	0.015	0.010	0.022	0.035	0.012	

Notes: Each model reports average marginal effects of a probit regression on the callback dummy (Y=1: employer calls back the job applicant). Marginal effects are calculated at the mean in case of continuous and at the modus in case of discrete independent variables (see last column for value of independent variables). Standard errors clustered on firm level are in parentheses. Regressions consider the entire sample. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

Table C-6: Marginal Effects from Probit Regressions on Callback Dummy (Including Models without Control Variables) and Hypotheses Testing (Gender Study)

Callback	(Ia)	(Ib)	(IIa)	(IIb)	(IIc)	(IId)	(IIe)	(IIf)	(IIg)	(IIh)	(IIIa)	(IIIb)
Female	-0.065***	-0.067***	-0.057**	-0.062**	-0.065***	-0.067***	-0.029	-0.029	-0.065***	-0.067***	0.047	0.043
	(0.019)	(0.020)	(0.028)	(0.028)	(0.019)	(0.019)	(0.025)	(0.027)	(0.019)	(0.020)	(0.062)	(0.062)
Certificate	0.039	0.025	0.049	0.033	0.040	0.026	0.039	0.025	0.038	0.024	0.095*	0.078
	(0.036)	(0.036)	(0.046)	(0.046)	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)	(0.057)	(0.056)
Female x Certificate			-0.021	-0.016							-0.107	-0.100
			(0.057)	(0.057)							(0.079)	(0.078)
Share of females t-1	-0.011	-0.021	-0.011	-0.021	-0.035*	-0.047**	-0.011	-0.021	-0.011	-0.021	-0.037*	-0.049**
	(0.019)	(0.020)	(0.019)	(0.020)	(0.021)	(0.023)	(0.019)	(0.020)	(0.019)	(0.020)	(0.021)	(0.023)
Female x					0.049**	0.052**					0.052**	0.055**
Share of females t-1					(0.021)	(0.022)					(0.021)	(0.022)
Late recruiter	-0.035	-0.021	-0.035	-0.021	-0.035	-0.021	-0.001	0.017	-0.036	-0.021	0.035	0.052
	(0.044)	(0.088)	(0.044)	(0.088)	(0.044)	(0.088)	(0.048)	(0.091)	(0.044)	(0.088)	(0.055)	(0.095)
Female x							-0.069*	-0.072*			-0.136**	-0.134**
Late recruiter							(0.036)	(0.038)			(0.059)	(0.059)
Vacancies/total jobs t-1	-0.013	0.002	-0.013	0.002	-0.013	0.002	-0.013	0.002	-0.030	-0.016	-0.026	-0.012
	(0.020)	(0.022)	(0.020)	(0.022)	(0.020)	(0.022)	(0.020)	(0.022)	(0.022)	(0.023)	(0.022)	(0.023)
Female x									0.033*	0.037**	0.026	0.030
Vacancies/total jobs t-1									(0.018)	(0.018)	(0.018)	(0.018)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
No. of obs.	1,080	1,080	1,080	1,080	1,080	1,080	1,080	1,080	1,080	1,080	1,080	1,080
Pseudo R^2	0.008	0.026	0.008	0.026	0.010	0.028	0.009	0.027	0.009	0.027	0.013	0.031
Log likelihood	-710.026	-696.980	-709.969	-696.948	-708.660	-695.456	-709.327	-696.244	-709.385	-696.198	-706.341	-693.120
Wald chi-squared	16.472	31.831	16.747	32.142	26.185	42.605	18.103	32.828	20.247	35.728	34.227	49.631
P-value	0.006	0.016	0.010	0.021	0.000	0.001	0.006	0.018	0.003	0.008	0.000	0.000

Notes: Each model reports average marginal effects of a probit regression on the callback dummy (Y=1: employer calls back the job applicant). Marginal effects are calculated at the means of all independent variables and denote an infinitesimal change in case of continuous variables and a discrete change in case of dummy variables. Standard errors clustered on firm level are in parentheses. Regressions consider only male-dominated jobs. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

Figure C-1: Interaction Effect between Female and Certificate Dummy

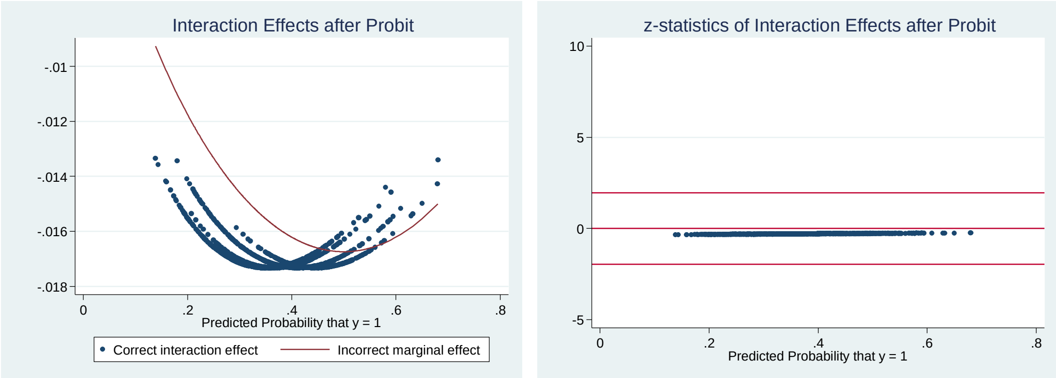


Figure C-2: Interaction Effect between Female Dummy and Share of Females t-1

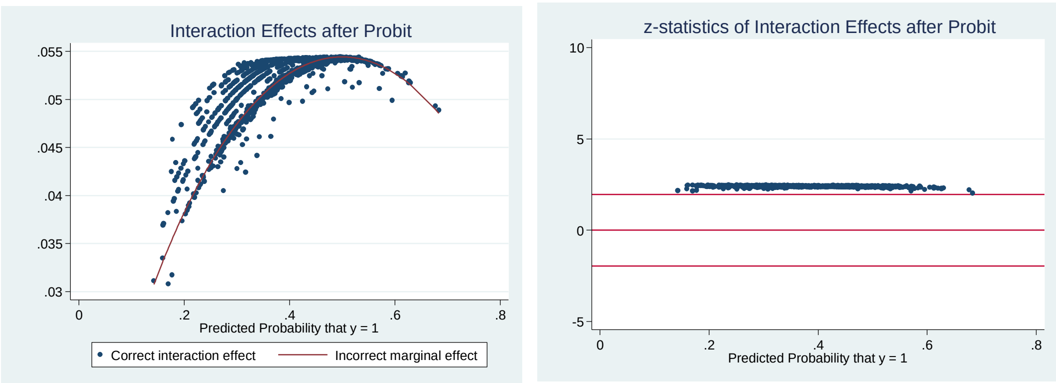


Figure C-3: Interaction Effect between Female and Late Recruiter Dummy

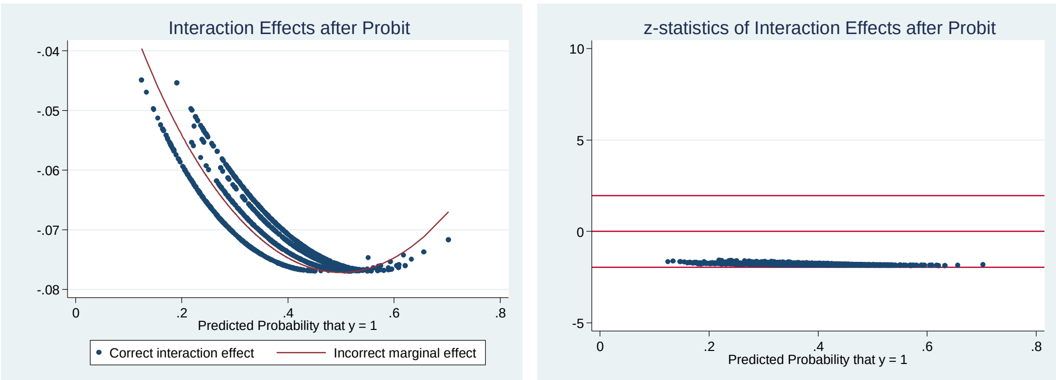


Figure C-4: Interaction Effect between Female Dummy and Vacancies/Total Jobs t-1

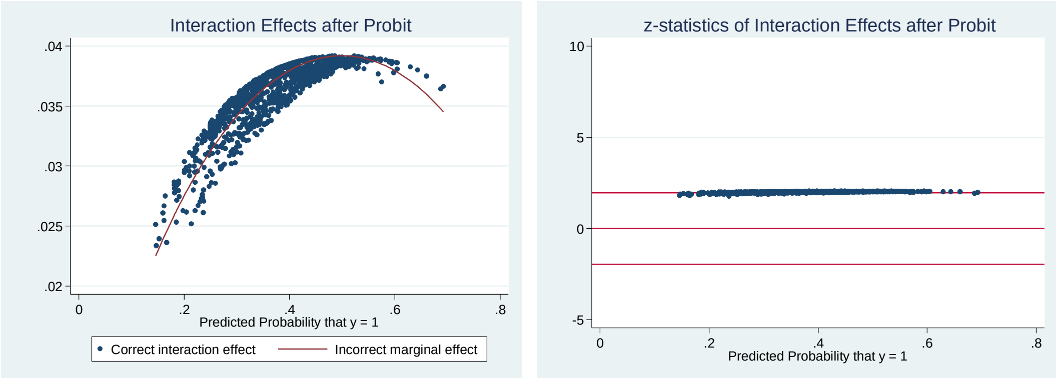


Table C-7: Firms' Responses of Correspondence Testing by Gender and Apprenticeship Program

	Firms' responses					Callback rates		
(1) No. of paired applications	(2) Rejection/ no response	(3) At least one callback	(4) Both	(5) Only male	(6) Only female	(7) Male (4+5)/(1)	(8) Female (4+6)/(1)	(9) Difference (7)-(8)
Industrial mechanic (173)	52.02 (90)	47.98 (83)	56.63 (47)	28.92 (24)	14.46 (12)	0.410	0.341	0.069 (p=0.183)
Electronics technician (69)	47.83 (33)	52.17 (36)	80.56 (29)	8.33 (3)	11.11 (4)	0.464	0.478	-0.014 (p=0.865)
Milling machine operator (69)	65.22 (45)	34.78 (24)	58.33 (14)	29.17 (7)	12.50 (3)	0.304	0.246	0.058 (p=0.446)
Mechatronics fitter (102)	51.96 (53)	48.04 (49)	61.22 (30)	26.53 (13)	12.24 (6)	0.422	0.353	0.069 (p=0.314)
Warehouse logistics operator (33)	39.39 (13)	60.61 (20)	40.00 (8)	45.00 (9)	15.00 (3)	0.515	0.333	0.182 (p=0.135)
Mechanic in plastics and rubber processing (94)	54.26 (51)	45.74 (43)	55.81 (24)	30.23 (13)	13.95 (6)	0.394	0.319	0.074 (p=0.286)
Geriatric nurse (24)	25.00 (6)	75.00 (18)	72.22 (13)	11.11 (2)	16.67 (3)	0.625	0.667	-0.042 (p=0.763)
Industrial clerk (43)	51.16 (22)	48.84 (21)	52.38 (11)	19.05 (4)	28.57 (6)	0.349	0.395	-0.047 (p=0.655)
Management assistant for office communication (49)	61.22 (30)	38.78 (19)	42.11 (8)	26.32 (5)	31.58 (6)	0.265	0.286	-0.020 (p=0.821)

Notes: This table shows the distribution of firms' responses. Absolute numbers are in parentheses. Column (1) displays the number of employers in each stratum. Column (2) reports the fraction of firms that gave none of the candidates a callback, so the remainder in column (3) called back at least one applicant. Firms that gave both candidates a positive answer, column (4), are considered as equal treatment, while the rest preferred either the male or the female candidate (columns (5) and (6)). Columns (7) and (8) contain the callback rate for the male and female applicant, respectively, while column (9) computes the difference in callback rates between the two candidate groups. In column (9), p-values of a chi-squared test that the male and female candidates are equally likely to receive a callback at any matched-pair application are in parentheses.

C.2 STUDY ON ETHNIC DISCRIMINATION

Table C-8: Marginal Effects from Probit Regressions on Response Dummy (Ethnicity Study)

Response	**(I)**	**(II)**	**(III)**	**(IV)**
Turkish name	-0.029*	-0.028*	-0.028*	-0.028*
	(0.015)	(0.015)	(0.015)	(0.015)
Medium		0.098***	0.099***	0.099***
		(0.035)	(0.035)	(0.035)
Large		0.171***	0.172***	0.172***
		(0.032)	(0.032)	(0.032)
South		-0.043	-0.043	-0.043
		(0.045)	(0.046)	(0.046)
East		-0.109**	-0.123**	-0.123**
		(0.055)	(0.061)	(0.061)
Industry		-0.088**	-0.089**	-0.089**
		(0.040)	(0.040)	(0.040)
Late recruiter		0.046	0.045	0.046
		(0.045)	(0.045)	(0.045)
Female responsible		0.056*	0.056*	0.056*
		(0.030)	(0.030)	(0.030)
Share of foreigners t-1			-0.009	-0.009
			(0.019)	(0.019)
Vacancies/total jobs t-1			0.004	0.005
			(0.015)	(0.015)
Certificate				0.007
				(0.029)
Controls	Yes	Yes	Yes	Yes
No. of obs.	1,216	1,216	1,216	1,216
Pseudo R^2	0.013	0.058	0.059	0.059
Log likelihood	-615.354	-586.867	-586.638	-586.610
Wald chi-squared	15.244	49.486	49.729	49.777
P-value	0.033	0.000	0.000	0.000

Notes: Each model reports average marginal effects of a probit regression on the response dummy (Y=1: employer gives the applicant either a rejection or a callback). Marginal effects are calculated at the means of all independent variables and denote an infinitesimal change in case of continuous variables and a discrete change in case of dummy variables. Standard errors clustered on firm level are in parentheses. Regressions consider the entire sample. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

Table C-9: Marginal Effects from Probit Regressions on Callback Dummy for a Standard Applicant at a Standard Employer (Ethnicity Study)

Callback	(I)	(II)	(III)	(IV)	X
Turkish name	-0.108***	-0.116***	-0.116***	-0.113***	0
	(0.016)	(0.017)	(0.017)	(0.017)	
Medium		0.082*	0.080*	0.076	1
		(0.047)	(0.047)	(0.047)	
Large		0.089	0.086	0.082	0
		(0.066)	(0.066)	(0.068)	
South		-0.047	-0.034	-0.033	1
		(0.060)	(0.062)	(0.062)	
East		0.020	0.038	0.034	0
		(0.062)	(0.067)	(0.068)	
Industry		-0.161***	-0.165***	-0.174***	1
		(0.059)	(0.059)	(0.061)	
Late recruiter		0.083	0.089	0.096*	1
		(0.058)	(0.058)	(0.057)	
Female responsible		0.086**	0.086**	0.087**	1
		(0.040)	(0.040)	(0.039)	
Share of foreigners t-1			0.002	0.001	0
			(0.023)	(0.023)	
Vacancies/total jobs t-1			-0.027	-0.026	0
			(0.023)	(0.023)	
Certificate				0.081**	0
				(0.035)	
Controls	Yes	Yes	Yes	Yes	
No. of obs.	1,216	1,216	1,216	1,216	
Pseudo R^2	0.023	0.044	0.045	0.048	
Log likelihood	-783.842	-767.369	-766.136	-764.143	
Wald chi-squared	58.024	76.345	78.194	81.306	
P-value	0.000	0.000	0.000	0.000	

Notes: Each model reports average marginal effects of a probit regression on the callback dummy (Y=1: employer calls back the job applicant). Marginal effects are calculated at the mean in case of continuous and at the modus in case of discrete independent variables (see last column for value of independent variables). Standard errors clustered on firm level are in parentheses. Regressions consider the entire sample. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

Table C-10: Marginal Effects from Probit Regressions on Callback Dummy for German-Named Applicants

Callback	**(Ia)**	**(Ib)**
Certificate	0.080*	0.076
	(0.042)	(0.055)
Lukas Schmidt	-0.002	-0.105
	(0.080)	(0.094)
Photo B	0.071	-0.035
	(0.083)	(0.101)
Distance	-0.001***	-0.000
	(0.000)	(0.000)
Design B	0.064	0.073
	(0.048)	(0.049)
Design C	0.081	0.098*
	(0.054)	(0.056)
Rank 2	-0.009	0.009
	(0.041)	(0.043)
Controls	No	Yes
No. of obs.	608	608
Pseudo R^2	0.020	0.040
Log likelihood	-405.779	-397.440
LR chi-squared	15.963	31.891
P-value	0.025	0.023

Notes: Each model reports average marginal effects of a probit regression on the callback dummy (Y=1: employer calls back the job applicant). Marginal effects are calculated at the means of all independent variables and denote an infinitesimal change in case of continuous variables and a discrete change in case of dummy variables. Robust standard errors are in parentheses. The sample is restricted to German-named applicants. Controls include firm characteristics and labor market data. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

Table C-11: Marginal Effects from Probit Regressions on Callback Dummy for Turkish-Named Applicants

Callback	**(Ia)**	**(Ib)**
Certificate	0.082**	0.080
	(0.041)	(0.053)
Onur Öztürk	-0.004	-0.066
	(0.078)	(0.099)
Photo B	-0.067	-0.083
	(0.082)	(0.102)
Distance	-0.000**	-0.000
	(0.000)	(0.000)
Design B	0.010	0.003
	(0.047)	(0.048)
Design C	-0.002	0.009
	(0.050)	(0.052)
Rank 2	0.025	-0.005
	(0.041)	(0.042)
Controls	No	Yes
No. of obs.	608	608
Pseudo R^2	0.015	0.055
Log likelihood	-375.707	-360.491
LR chi-squared	11.484	41.405
P-value	0.119	0.001

Notes: Each model reports average marginal effects of a probit regression on the callback dummy (Y=1: employer calls back the job applicant). Marginal effects are calculated at the means of all independent variables and denote an infinitesimal change in case of continuous variables and a discrete change in case of dummy variables. Robust standard errors are in parentheses. The sample is restricted to Turkish-named applicants. Controls include firm characteristics and labor market data. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

Table C-12: Marginal Effects from Probit Regressions on Callback Dummy (Including Models without Control Variables) and Hypotheses Testing (Ethnicity Study)

Callback	(Ia)	(Ib)	(IIa)	(IIb)	(IIc)	(IId)	(IIe)	(IIf)	(IIg)	(IIh)	(IIIa)	(IIIb)
Turkish name	-0.103***	-0.109***	-0.107***	-0.117***	-0.103***	-0.110***	-0.072***	-0.079***	-0.103***	-0.109***	-0.052	-0.070
	(0.015)	(0.016)	(0.025)	(0.025)	(0.015)	(0.016)	(0.022)	(0.023)	(0.015)	(0.016)	(0.053)	(0.053)
Certificate	0.101***	0.077**	0.096**	0.067	0.101***	0.077**	0.101***	0.077**	0.101***	0.076**	0.115**	0.083*
	(0.032)	(0.034)	(0.041)	(0.043)	(0.032)	(0.034)	(0.032)	(0.033)	(0.032)	(0.034)	(0.048)	(0.050)
Turkish name x			0.010	0.021							-0.029	-0.013
Certificate			(0.053)	(0.053)							(0.069)	(0.070)
Share of foreigners t-1	-0.018	0.001	-0.018	0.001	-0.007	0.014	-0.018	0.001	-0.018	0.001	-0.005	0.016
	(0.018)	(0.022)	(0.018)	(0.022)	(0.020)	(0.024)	(0.018)	(0.022)	(0.018)	(0.022)	(0.020)	(0.024)
Turkish name x					-0.023	-0.027					-0.027	-0.031*
Share of foreigners t-1					(0.017)	(0.018)					(0.017)	(0.018)
Late recruiter	0.021	0.091*	0.021	0.091*	0.021	0.091*	0.047	0.117**	0.021	0.091*	0.055	0.121**
	(0.040)	(0.054)	(0.040)	(0.054)	(0.040)	(0.054)	(0.042)	(0.056)	(0.040)	(0.054)	(0.046)	(0.060)
Turkish name x							-0.054*	-0.052*			-0.070	-0.060
Late recruiter							(0.030)	(0.031)			(0.048)	(0.049)
Vacancies/total jobs t-1	-0.029	-0.025	-0.029	-0.025	-0.029	-0.025	-0.029	-0.025	-0.036*	-0.033	-0.037*	-0.034
	(0.020)	(0.022)	(0.020)	(0.022)	(0.020)	(0.022)	(0.020)	(0.022)	(0.021)	(0.023)	(0.021)	(0.023)
Turkish name x									0.013	0.017	0.016	0.020
Vacancies/total jobs t-1									(0.015)	(0.015)	(0.015)	(0.015)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
No. of obs.	1,216	1,216	1,216	1,216	1,216	1,216	1,216	1,216	1,216	1,216	1,216	1,216
Pseudo R^2	0.018	0.048	0.018	0.048	0.019	0.048	0.019	0.048	0.018	0.048	0.020	0.049
Log likelihood	-787.804	-764.143	-787.787	-764.080	-787.478	-763.698	-787.334	-763.729	-787.686	-763.960	-786.710	-762.972
Wald chi-squared	53.483	81.306	53.500	81.789	53.171	80.762	54.574	81.164	55.370	83.031	56.698	82.739
P-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: Each model reports average marginal effects of a probit regression on the callback dummy (Y=1: employer calls back the job applicant). Marginal effects are calculated at the means of all independent variables and denote an infinitesimal change in case of continuous variables and a discrete change in case of dummy variables. Standard errors clustered on firm level are in parentheses. Regressions consider the entire sample. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

Figure C-5: Interaction Effect between Turkish Name and Certificate Dummy

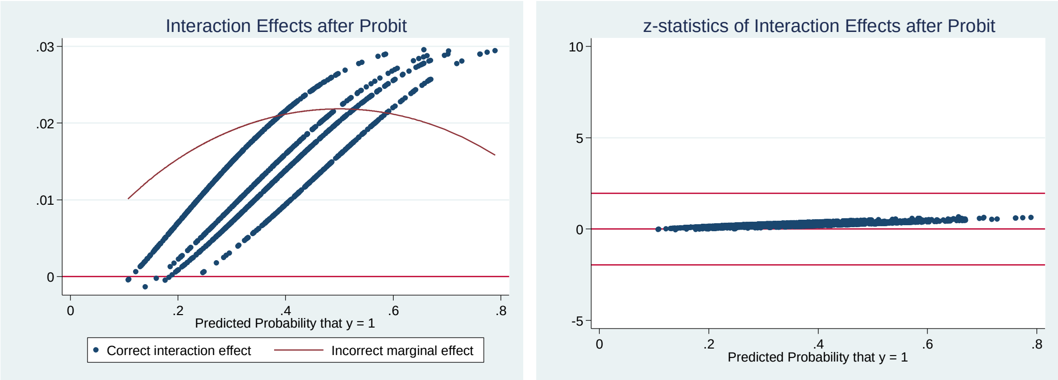


Figure C-6: Interaction Effect between Turkish Name Dummy and Share of Foreigners t-1

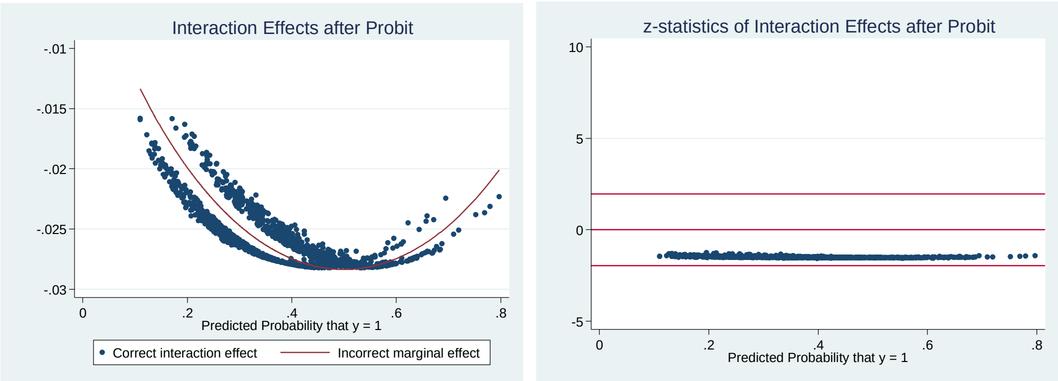


Figure C-7: Interaction Effect between Turkish Name and Late Recruiter Dummy

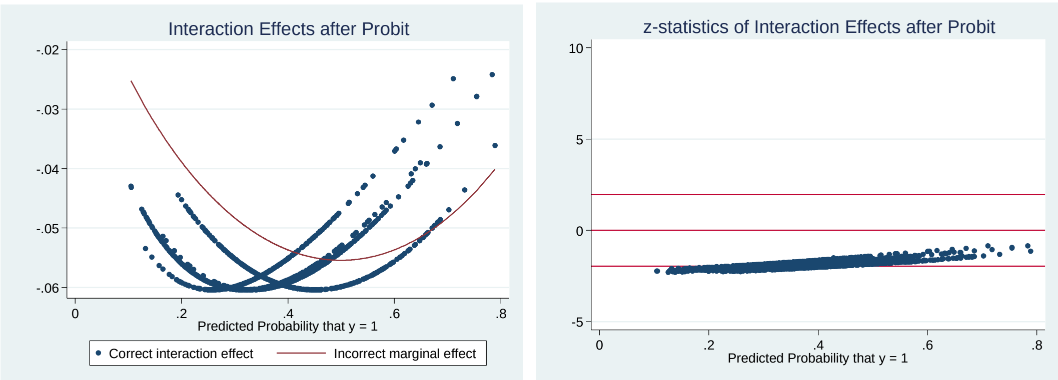


Figure C-8: Interaction Effect between Turkish Name Dummy and Vacancies/Total Jobs t-1

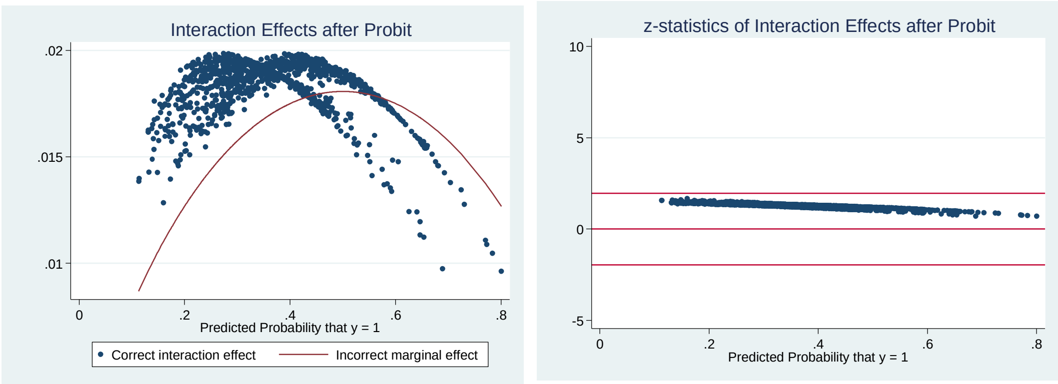


Table C-13: Marginal Effects from Probit Regression on Late Recruiter Dummy

Late recruiter	**(I)**
Medium	-0.26***
	(0.05)
Large	-0.35***
	(0.07)
South	0.11**
	(0.05)
East	0.37***
	(0.05)
Industry	-0.07
	(0.07)
Female responsible	-0.09**
	(0.04)
Share of foreigners t-1	0.03
	(0.03)
Vacancies/total jobs t-1	-0.07***
	(0.02)
Open positions	-0.03*
	(0.02)
No. of obs.	1,216
Pseudo R^2	0.129
Log likelihood	-723.861
Wald chi-squared	90.631
P-value	0.000

Notes: Table reports average marginal effects of a probit regression on the late recruiter dummy (Y=1: firm offers vacancy in May) for the entire sample. Standard errors clustered on firm level are in parentheses. * denotes 10% significance level. ** denotes 5% significance level. *** denotes 1% significance level.

C.3 STUDY ON METHODOLOGICAL VARIATIONS

Table C-14: Descriptive Statistics of the Method Comparison in the Study on Gender Discrimination

Variable	Operationalization	# of Obs.	Mean	SD	Min	Max
DEPENDENT VARIABLES						
Response	Dummy: Equals 1 if the applicant receives a response (either invitation or rejection) by the employer, 0 otherwise	444	0.806	-	0	1
Callback	Dummy: Equals 1 if the applicant receives a callback (e.g. invitation) by the employer, 0 otherwise	444	0.394	-	0	1
INDEPENDENT VARIABLES						
Method						
Correspondence	Dummy: Equals 1 if pairwise applications are sent out, 0 otherwise	444	0.671	-	0	1
Applicant information						
Female	Dummy: Equals 1 if the applicant is female, 0 otherwise	444	0.500	-	0	1
Design						
Design A	Dummy: Equals 1 if the application has design A, 0 otherwise	444	0.502	-	0	1
Design B	Dummy: Equals 1 if the application has design B, 0 otherwise	444	0.498	-	0	1
Rank						
Rank 1	Dummy: Equals 1 if the application was sent out first, 0 otherwise	444	0.665	-	0	1
Rank 2	Dummy: Equals 1 if the application was sent out second, 0 otherwise	444	0.336	-	0	1
Certificate	Dummy: Equals 1 if the applicant provides an additional certificate, 0 otherwise	444	0.541	-	0	1
Distance	Linear distance between applicant's home and location of employer (in km)	444	243.38	110.15	0	533
Information on jobs						
Female-dominated job	Dummy: Equals 1 if the majority in the respective apprenticeship is female, 0 otherwise (i.e., the majority is male)	444	0.777	-	0	1
Firm characteristics						
Size						
Small	Dummy: Equals 1 if the employer has less than 50 employees, 0 otherwise	444	0.570	-	0	1
Medium	Dummy: Equals 1 if the employer has between 50 and 500 employees, 0 otherwise	444	0.405	-	0	1
Large	Dummy: Equals 1 if the employer has more than 500 employees, 0 otherwise	444	0.025	-	0	1
Location						
Other	Dummy: Equals 1 if the employer is not located in the South or East of Germany, 0 otherwise	444	0.405	-	0	1
South	Dummy: Equals 1 if the employer is located in the South of Germany, 0 otherwise	444	0.383	-	0	1
East	Dummy: Equals 1 if the employer is located in Eastern Germany, 0 otherwise	444	0.212	-	0	1
Industry	Dummy: Equals 1 if the employer operates in the industry sector, 0 otherwise (i.e., service sector)	444	0.293	-	0	1
Female responsible	Dummy: Equals 1 if the person responsible for recruiting as mentioned in the job offer is female, 0 otherwise	444	0.570	-	0	1
Open positions	Number of open positions for an apprenticeship as indicated by the employer's job offer	444	1.28	0.928	1	10
Labor market data						
Vacancies/total jobs t-1	Ratio of vacancies and total apprenticeships in the corresponding Employment Agency region of the employer in 2010/2011	444	0.057	0.035	0.009	0.163
Share of females t-1	Share of female applicants in the corresponding Employment Agency region of the employer in 2010/2011	444	0.520	0.201	0.120	0.740

Table C-15: Descriptive Statistics of the Method Comparison in the Study on Ethnic Discrimination

Variable	Operationalization	# of Obs.	Mean	SD	Min	Max
DEPENDENT VARIABLES						
Response	Dummy: Equals 1 if the applicant receives a response (either invitation or rejection) by the employer, 0 otherwise	302	0.801	-	0	1
Callback	Dummy: Equals 1 if the applicant receives a callback (e.g. invitation) by the employer, 0 otherwise	302	0.454	-	0	1
INDEPENDENT VARIABLES						
Method						
Correspondence	Dummy: Equals 1 if pairwise applications are sent out, 0 otherwise	302	0.669	-	0	1
Applicant information						
Turkish name	Dummy: Equals 1 if the applicant has a Turkish-sounding name, 0 otherwise	302	0.501	-	0	1
Design						
Design A	Dummy: Equals 1 if the application has design A, 0 otherwise	302	0.510	-	0	1
Design B	Dummy: Equals 1 if the application has design B, 0 otherwise	302	0.490	-	0	1
Rank						
Rank 1	Dummy: Equals 1 if the application was sent out first, 0 otherwise	302	0.666	-	0	1
Rank 2	Dummy: Equals 1 if the application was sent out second, 0 otherwise	302	0.334	-	0	1
Certificate	Dummy: Equals 1 if the applicant provides an additional certificate, 0 otherwise	302	0.520	-	0	1
Distance	Linear distance between applicant's home and location of employer (in km)	302	254.00	98.78	39	494
Firm characteristics						
Size						
Small	Dummy: Equals 1 if the employer has less than 50 employees, 0 otherwise	302	0.460	-	0	1
Medium	Dummy: Equals 1 if the employer has between 50 and 500 employees, 0 otherwise	302	0.487	-	0	1
Large	Dummy: Equals 1 if the employer has more than 500 employees, 0 otherwise	302	0.053	-	0	1
Location						
Other	Dummy: Equals 1 if the employer is not located in the South or East of Germany, 0 otherwise	302	0.262	-	0	1
South	Dummy: Equals 1 if the employer is located in the South of Germany, 0 otherwise	302	0.523	-	0	1
East	Dummy: Equals 1 if the employer is located in Eastern Germany, 0 otherwise	302	0.215	-	0	1
Industry	Dummy: Equals 1 if the employer operates in the industry sector, 0 otherwise (i.e., service sector)	302	0.871	-	0	1
Female responsible	Dummy: Equals 1 if the person responsible for recruiting as mentioned in the job offer is female, 0 otherwise	302	0.424	-	0	1
Open positions	Number of open positions for an apprenticeship as indicated by the employer's job offer	302	1.25	0.683	1	8
Labor market data						
Vacancies/total jobs t-1	Ratio of vacancies and total apprenticeships in the corresponding Employment Agency region of the employer in 2010/2011	302	0.053	0.027	0.004	0.130
Share of foreigners t-1	Share of foreign applicants in the corresponding Employment Agency region of the employer in 2010/2011	302	0.103	0.081	0.000	0.340

EIDESSTATTLICHE ERKLÄRUNG

Hiermit versichere ich, Andre Kolle, die vorliegende Arbeit selbstständig und unter ausschließlicher Verwendung der angegebenen Literatur und Hilfsmittel erstellt zu haben. Alle Stellen, die wörtlich oder sinngemäß veröffentlichtem oder unveröffentlichtem Schrifttum entnommen sind, habe ich als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht.

Andre Kolle Paderborn, 30. März 2014