AARHUS UNIVERSITY

Coversheet

This is the accepted manuscript (post-print version) of the article.

Contentwise, the post-print version is identical to the final published version, but there may be differences in typography and layout.

How to cite this publication

Please cite the final published version:

Villadsen, A. R., & Wulff, J. (2016). Field experimental evidence on sector differences in ethnic discrimination: A Bayesian approach. *Academy of Management Proceedings*.

DOI: [10.5465/AMBPP.2016.14416abstract](https://doi.org/10.5465/AMBPP.2016.14416abstract)

Publication metadata

Title:	Field experimental evidence on sector differences in ethnic discrimination: A Bayesian approach
Author(s):	Villadsen, A. R., & Wulff, J.
Journal:	Academy of Management Proceedings
DOI/Link:	10.5465/AMBPP.2016.14416abstract
Document version:	Accepted manuscript (post-print)

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Field Experimental Evidence on Sector Differences in Ethnic Discrimination: A Bayesian Approach

Abstract

Research has found considerable ethnic discrimination in employment decisions of private firms. Yet, little is known about sector differences in employment discrimination and whether public organizations are more or less likely to discriminate against prospective employees based on their ethnicity. Theoretical arguments offer opposing predictions and few empirical studies exist. We present evidence of private and public sector differences in ethnic discrimination in the recruitment process of Danish organizations. The data have been collected through a field experiment. Danish and non-Western immigrant names are randomly assigned to job applications with identical skills and sent to real job openings in four different occupational categories in both the private and public sector. A total of 888 job applications were sent out. To get more reasonable inferences, provide more intuitive interpretations of results and to mitigate problems with assumptions of repeated sampling, we take a Bayesian Modelling approach. Danish-sounding applications have a 35% higher call-back rate among private sector firms compared to Middle-Eastern sounding names. In the public sector this difference is considerably higher and applications with Danish names are 59% more likely to receive a call-back for an interview. While consistent across occupational categories, the discrimination effect is by far the largest in the school teacher category. In total, the evidence suggests ethnic discrimination in hiring decisions to be substantial and to be more pronounced in the public sector than in the private sector. Implications for research are discussed.

INTRODUCTION

Employee diversity is an increasingly important challenge for workplaces in both the public and private sector. In many countries, workforces are becoming more diverse with respect to dimensions such as ethnicity, education, and age. It is important that organizations be ready for this challenge in order to effectively identify, attract, retain, and utilize the best possible human talents. Most diversity research has focused on the latter phases and the question of how organizations cope with the potentials and challenges of a diverse group of employees. Much fewer studies are interested in former phases involving the process in which job candidates are screened and hired into organizations. This is an important process where organizations make explicit and implicit choices about whom to hire. This is also a process that can be distorted by employment discrimination that decreases employment opportunities for certain groups of job candidates. This type of discrimination is problematic not only because it in most places is illegal. From an economic viewpoint discrimination introduces irrelevant factors into the evaluation process of job candidates and makes it less likely that organizations end up hiring the best possible talents in the pool of candidates.

Very few studies of employment discrimination in the public sector exist (Leasher & Miller, 2012). While a best case interpretation of this fact would be that the problem is limited, more realistically it likely reflects an important gap in our understanding of how public sector organizations work with diversity. In this study we focus on ethnic discrimination in the employment process of public and private organizations in Denmark. Discrimination based on ethnicity is a very salient type of discrimination as migration has made workforces in many countries increasingly diverse with respect to this dimension. For public organizations this is an important dimension as research on representative bureaucracy has shown that positive outcomes may follow from ethnic representation (Pitts, 2005).

Arguments suggest that public sector organizations may be both more and less prone to discrimination compared to (otherwise similar) private organizations. This is also reflected in mixed findings in a recent comparative study of discrimination (Leasher & Miller, 2012). Public organizations with political leadership and constrained by rules may be less likely to discriminate. On the other hand, private organizations may respond more quickly to a changing environment and adapt hiring practices to allow for an unbiased evaluation of all candidates. Because they typically have more clear and tangible goals the costs associated with not hiring the best employees may be more immediately visible for private organizations.

To investigate the potential discriminatory difference between public and private organizations we conduct a field experiment. Field experiments are especially useful to evaluate the existence and level of discrimination amongst employers (Riach & Rich, 2002). This type of research design combines the benefits of a controlled experiment with actual situations in realistic surroundings (Harrison & List, 2004). By manipulating the ethnicity of names in otherwise equal applications, we hold all other effects constant and isolate the actual discriminatory effect of having a name that signals membership of the group of non-Western immigrants (Grohs, Adam, & Knill, 2015).

Contrary to past research, we adopt a Bayesian approach when analyzing the data we collect on discriminatory differences. As explained in more detail below, compared to traditional statistical approaches the Bayesian method has several desirable properties advantageous to our research purpose. First, Bayesian analysis makes us able to answer the questions we are interested in: How likely is it that there is a discriminatory difference between the public and private sector and how large is it? Second, because we apply for all the positions we can find in our data collection period, the assumption of repeated sampling upon which traditional statistics

rely is questionable. Third, a Bayesian approach is very well suited for hierarchical analysis and allows for intuitive interpretations through e.g. credible intervals and plots of posterior distributions (Gelman & Hill, 2007; Jackman, 2009). Finally, there are strong substantive reasons why research on public administration including the literature on sector differences should adopt a Bayesian approach (Gill & Witko, 2013). This point was first made more than 15 years ago (Gill & Meier, 2000) but still very few studies in public administration embrace a Bayesian approach. Decision-making processes in public organizations can be thought of as Bayesian in the sense that employees and executives update their past beliefs based on acquired information (Boyne, 2005). Thus, our research objectives already fit the assumptions of Bayesian statistics well.

Based on 888 applications to 222 job openings in four different occupational categories, we find that applications attached with a Middle-Eastern sounding name on average have a 52.7% lower callback rate than applications with a Danish sounding names. The probability that this difference is larger than zero is 0.999. Interestingly, the average difference in callback rates favoring Danish applicants is 11%-points higher amongst organizations within the public sector and is larger than zero with a probability of 0.996. This difference in discriminatory effect was notably stronger among the school teacher positions than the three remaining job categories.

With this study we make two substantial contributions to existing research. First, we show that ethnic discrimination is present in employment decisions in Denmark and seems to be so most notably in the public sector. Contrary to the few existing studies, we use a well-tested field experimental method that enhances the validity of the results. This is an important finding that clearly points to the importance of focusing on the earlier stages of hiring processes in studies of diversity. The case of Denmark, often heralded for being among the world's least corrupt

countries and with a well-functioning public administration, makes this finding even more noteworthy. Second, the study is among the first in the public administration literature to use Bayesian statistics for the analyses as called for by Gill and Meier (2000). Compared to traditional approaches Bayesian statistics does not rely on null hypotheses significance testing and allow for a more direct assessment of the research question. The method is becoming more popular in the fields of management and organization (Zyphur, Oswald, & Rupp, 2015) and we believe the time has come to more frequently using Bayesian approaches in public administration research.

THEORY

Employment discrimination

Employment discrimination is an irrational practice by which an organization let irrelevant factors such as gender, race, or ethnicity factor into employment decisions (Becker, 1971). As such an a priori expectation would be that organizations do not discriminate. Ample evidence suggests this not to be true. Within public administration we have relatively few studies of employment discrimination. The studies we find fall within two types. The first type of studies uses archival data to observe numerical differences between groups of employees or to analyze claims of discrimination. An example of this is Llorens, Wenger, and Kellough (2008) who investigate the representation of women and racial minorities in the public sector in the U.S.. Related, Alkadry and Tower (Alkadry & Tower, 2006, 2011) look at gender wage differences as an indication of discrimination. Another example is Leasher and Miller (2012) who use discrimination claims filed in Ohio to explore difference between public and private sector workplaces. The second type of studies is based on questionnaire data and employees' reported

perception of discrimination. An example of this is Hopkins' (1980) study pointing to relatively widespread discrimination perceptions among respondents in five states. These types of data are routinely collected by governments (Naff, 1995).

The two types of studies mentioned above both suffer from serious limitations. First, the validity of the measure of discrimination is questionable. Indeed, in the first type of studies discrimination is inferred, not measured. That a certain group is over/under-represented may be a sign of discrimination but may also result from idiosyncratic labor market features or the fact that one group on average exhibit higher levels of a desired skill or trait. The second type of studies measure discrimination more directly but with the caveats following from self-reported measurement (Meier & O'Toole Jr, 2013). Second, these types of studies are solely focusing on employees already working in the organizations under investigation. This means that these methods are not able to capture discrimination taking place in the hiring process where knowledge about all applicants would be necessary.

To assess the extent of discrimination in hiring processes researchers have turned to field experimental methods (Pager, 2007). One particularly useful type of field experiment in this area is the audit study (Pager, Western, & Bonikowski, 2009). Such studies send out identical applications for real job openings only varying the dimension of interest (in this case ethnicity) and use differences in call back rates as an indication of discrimination. As this type of data collection is quite resource intensive, only a relatively modest number of studies has been published. An early example of this type of study is Bertrand and Mullainathan (Bertrand & Mullainathan, 2003), who using job opening in Chicago and New York found applications with white sounding names being about 50 % more likely to receive a callback than applications with black sounding names.

In many countries ethnic diversity is becoming increasingly pronounced with intensifying migration. The multi-ethnic societies are faced with the challenge of securing equal access to the labor market for all. While focusing on race, the study described above may indicate that ethnic discrimination can be a real threat to this process. So far, very few of the studies of ethnic employment discrimination have paid particular attention to the public sector. One exception is a Swedish study of ethnic discrimination using the audit method. In this study, Carlsson and Rooth (2007) find that applicants with traditional Swedish sounding names are about twice as likely to receive a call back as applicants with a traditional middle-eastern sounding name. Some of their applications seem to be public sector positions with a similar effect. Overall, however, there is a substantial gap in our knowledge of ethnic employment discrimination in public sector workplaces.

Differences between the public and private sector

Empirical studies of the public and private sector both support the presence of employment discrimination in workplaces, yet little proper comparative evidence exists. This is a problem as discrimination is a topic that cuts across many of the sector demarcation lines brought up in the literature. As such, comparative studies in this area may help us learn more about the core characteristics of both sectors as well as important similarities and differences. We argue below that it is not clear whether discrimination is likely to be more prevalent in the public or private sector as arguments offer opposing predictions.

Drawing on the four dimensions proposed by Boyne (2002), one dimension on which public and private organizations may differ is in their environments. Because they are politically led and mostly financed by tax payer money, public organizations are typically more closely scrutinized by environmental audiences than are private organizations. This relates to media

attention as well as requirements for openness and transparency. In such an environment illegal discrimination may be a risky practice that is likely to be exposed. As such we might expect the simpler and less invasive environments of private sector organizations (Boyne, 2002) to make discrimination more likely.

The second dimension forwarded by Boyne relates to organizational goals. Private organizations are argued to focus mainly on financial goals where public organizations have more varied and often ambiguous goals (Chun & Rainey, 2005). A focus on financial goals would lead organizations to hire the candidate that is most appropriate for the job and discrimination would lead to lower profits as suboptimal candidates are hired. Ambiguous and hard-to-measure goals in the public sector make it less clear if and how discrimination affects performance. Therefore, it is easier for decision makers to include irrelevant factors such as ethnicity in hiring processes.

A third dimension that Boyne (2002) highlights is organizational structures. Public organizations are traditionally thought to be more bureaucratic, have more red tape, and less managerial autonomy than private organizations. A greater prevalence of rules, regulation, and standards may be related to a smaller likelihood of discrimination in hiring processes. If the hiring process as well as job descriptions are clearly formalized and specified, it is more difficult to enter irrelevant factors into the decisions. It will, for instance, be difficult to disregard an applicant based on ethnicity if the person is clearly the strongest based on objective criteria that are specified by the organization as desirable. In private organizations where processes are less formalized and structures less rigid it may be easier for decision makers to discriminate.

Finally, Boyne points to sector differences in the managerial values. In the public sector managers are argued to be more focused on the public interest and serving public needs than in the

private sector (Perry, 2000). It is not entirely clear how such values relate to discriminatory practices. On the one hand, such values could be related to values such as equality and equity indicating that everybody should be treated fairly as well as to the idea that a representative bureaucracy will better serve the needs of the public. On the other hand, managers may perceive that it is easier to serve the public if public organizations are staffed with many similar people that work well together.

While a fairly substantial amount of research has investigated differences between public and private organizations (Rainey, 2009), very little evidence exists about dubious practices such as employment discrimination. As suggested above, theory does not offer precise predictions on this question.

EXPERIMENTAL DESIGN

To examine ethnic discrimination, we carried out a field experiment. Following previous research (Agerström & Rooth, 2011; Bertrand & Mullainathan, 2004; Carlsson & Rooth, 2007) we used the audit method and send out artificial applications to a large number of real job openings with either an ethnic Danish or an ethnic Middle-Eastern name randomly assigned. Our measure of discrimination is the difference in callback rate between the two types of applications. Because names are randomly assigned to applications we would expect similar call-back rates if no discrimination exists. The experiment was conducted between February 2015 and July 2015. During this period, we applied for all relevant job openings in four selected job categories through the homepage of the Danish job database jobindeks.dk. To each job opening we sent four applications as described below. In total, we sent 888 applications to 222 employers. Three graduate students were employed with sending out applications and receiving the callbacks via

telephone or e-mail. When an application received an invitation to an interview, the invitation was declined within a day to cause as little inconvenience as possible for the employer.

Job Categories

We chose our occupational categories based on a range of criteria. First, for our research purpose it was important that we picked jobs that exist both in the public and private sector. Second, it was necessary that labour demand was high in the selected job categories to secure a sufficient number of job openings. Third, to insure some variation between the job types, we both applied for front office jobs with client and / or customer contact on a regular basis as well as back office jobs mainly involving desk work. Finally, we restricted the job search to the two major regions in Denmark: Copenhagen and Central Jutland. The following job categories were selected: school teacher, physiotherapist, office assistants and financial controller¹.

Applications

We created a bank of applications. The applications consisted of a general biography on the first page and a detailed resume on the second page. We used an online job application builder to create templates for a pool of six applications for each job category². After drafts had been constructed these were compared to applications from actual job searchers and adjusted accordingly to seem as real as possible. Finally, we adjusted the layout, structure and language to make each application appear distinct.

Each application was given fictitious addresses based on real streets in the same area where the job was applied for. This means that if an application was sent to a job opening in Central Jutland,

¹ Applications were also crafted and sent to two other job categories: IT-professionals and secretaries. With regard to the former, we received only three callbacks out of the first 36 applications. This callback rate was so low that we decided to terminate the category and focus on the other categories instead. Similarly, we only received one callback out of 72 applications sent to secretary positions, which is why we also chose to shut down that category as well.

² We used the resumé builder on www.jobindex.dk

the application was attached with an address in Central Jutland. In this way, we avoid potential influences from employers preferring applications who lived closer. Besides this adjustment, all applications had 3-4 years of relevant work experience and a similar relevant educational level at the same type of school but at different locations.

Following Bertrand and Mullainathan (2004), we also separated each occupational category into a high and a low quality group of applicants, respectively. The idea behind this was to explore whether discrimination might depend on the skill level of the applicants. To signal additional quality among the high quality applications, we added a subset of the following characteristics: Volunteering experience, certification degrees, relevant courses, IT-skills and foreign language skills. The challenge was to make the quality manipulation strong enough to make the low and high categories distinct, while avoiding making the applications appear over- or under-qualified. Thus, we only add a few of the listed features to each application instead of all at once.

Using data from Statistics Denmark, we chose the three most commonly used names for citizens with Danish and a non-western origin, respectively. Henceforth, the names for non-western immigrants are labelled as ethnic names. The names are available in Appendix A.1. The names in the two groups are very distinct from each other and provide an unmistakable stimulus. For each job advertisement, four applications were drawn at random from the relevant resume bank. Next, two Danish and two ethnic names (one female and one male each) were drawn at random from each list and attached to the applications.

If an applicant is called to an interview, we record it as a callback. We log the name and contact information for each employer and any relevant additional information, including whether the company belongs to the public or private sector.

BAYESIAN ANALYSIS

The traditional method of assessing the statistical support for a hypothesis advocated by classical statistics is p -value null hypothesis significance testing (Andraszewicz et al., 2015). While this method is dominant in management research, it regrettably suffers from several severe drawbacks both statistically and logically (Morey, Hoekstra, Rouder, Lee, & Wagenmakers, 2015; Wagenmakers, 2007). As p -values indicate the extremity of the data under the null hypothesis, researchers often fall for the interpretational fallacy of believing that the smaller the p -value, the greater the probability that the null hypothesis is false (Gill, 1999). This confusion is made worse by the facts that p -values overstate the evidence against the null hypothesis (Gallistel, 2009), they do not allow researchers to quantitatively state evidence for the null hypothesis (Berger & Delampady, 1987), and they depend on the intention with which the data were collected (Andraszewicz et al., 2015; Lindley, 1993).

Bayesian methods provide an attractive alternative that overcomes the shortcomings of traditional methods (Jeffreys, 1961). When performing a Bayesian analysis the outcome is a posterior probability statement, i.e. a statement about our beliefs after having looked at the data (Jackman, 2009). This makes us able to make straightforward statements about the probability that public and private firms discriminate differently against applicants with an Ethnic name. Instead of asking “What is the probability of observing a given difference (or more extreme) given that that there is no difference?”, we post the question, “What is the probability that the difference is larger than zero?”. In the light of previous research, we deem it unreasonable to assume that there will be exactly zero difference in the callback probability between attaching a Danish or an Ethnic name to an application. Without conditioning our results on improbable hypotheses, the Bayesian

approach allows us to answer questions of more substantive interest while keeping the answers simple and clear (Mackey, Barney, & Dotson, 2015; Orlitzky, 2011).

Bayesian analysis does not express statements about model parameters as single point estimates, but rather as probabilistic assertions through distributions (Gill, 2015). Consequently, we can avoid traditional confidence intervals and instead use the more intuitive credible interval to describe our posterior distribution (Morey et al., 2015). In contrast to the traditional and often confusing confidence interval, a 90% credibility interval (CI) means exactly what we are interested in knowing: The interval that covers the true value with a 90% probability (Andraszewicz et al., 2015).

Furthermore, a Bayesian approach fits our research design in another important way, namely the problem of characterizing our data as a sample in the traditional, frequentist sense. As described above, we applied for all the jobs we could find within the four job categories in the specified period. One could argue that repeating the data collection process would not yield a new sample from the population, but rather the exact same data. In contrast to traditional statistics, Bayesian methods do not rely on the assumption of repeated sampling from the same population (Jackman, 2009). Instead, uncertainty in the Bayesian paradigm is attached to our state of knowledge about the effect of ethnicity on the probability of callback and how this effect differs between public and private firms, which is very desirable in our research setting.

We note that to our knowledge no prior audit studies of discrimination rely on Bayesian analysis. Given the arguments above, the results of the present study may present more precise estimates of discrimination effects.

All data management and analysis was done using the statistical software package R (R Core Team, 2015). The probabilities of differences in proportions were calculated using the BayesianFirstAid package (Bååth, 2013), while the Bayes factors testing for independence are computed using Gunel and Dickey's (1974) approach implemented in the BayesFactor package (Morey & Rouder, 2015). The hierarchical models were estimated using the JAGS program accessed and processed in R through rjags (Plummer, 2015).

EMPIRICAL RESULTS

Table 1 tabulates average callback rates by names and sector type. The bottom row presents the results for the full data. The 'Total' column shows the total number of "pairs of applications" sent out. A pair consists of a Danish and a Middle-Eastern named applicant. Thus, the 444 in the bottom row shows that 888 (444×2) applications were sent in total. The 'n' column shows the number of positive responses, i.e. the number of applicants who were called to an interview. In total, we received 205 positive callbacks inviting our applicant to a job interview equivalent to an overall success rate of 23%. Compared to similar field experiments, this is a very favorable callback rate.

Ethnic Origin and Sector

Continuing with Table 1, our data show that applicants with a Danish name have 27.9% probability of receiving a positive callback, while it is 18.2% for the applicants with an Ethnic name. This result implies that Danish-named applicants are called to about three interviews for every ten applications they send out, while the applicants with an Ethnic name must send out about 15 to get the same number of callbacks. This represents a difference of 9.6 %-points or 52% that is solely due to the name difference. The credibility interval ranges from 4.1% to 15.1% signifying that given our data there is a 95% probability that the true difference in callback falls within this region. Not

surprisingly, it is very likely that the callback rate for applicants with Danish names is larger than the one for applicants with ethnic names. The calculated Bayes factor is 25.149 indicating that the hypothesis of dependence is more than 25 times more likely to be true compared to the null-hypothesis of independence. This is decisive evidence of a discrimination effect.

Insert Table 1 around here

Table 1 further compares the differences in callback rates across private and public sectors. Both sectors tend to discriminate but to a very different extend. While applicants to private firms have 3.6%-points higher probability of a positive callback, the difference in probability is 14.2%-points when applied to public firms. The probability that the callback rate is higher for applications attached with a Danish name is 85.35%. When consulting the Bayes factor, however, the evidence points positively in favour of independence. With regard to public firms, the probability of a higher callback rate among Danish applicants is 99.98%. The Bayes factor indicates decisive evidence that it is more than 40 times more likely that the callback depends on the attached name than the alternative. In sum, we find decisive evidence of higher name-based discrimination for public compared to private firms. This strength of the evidence in favour of this difference is presented through the hierarchical modelling below in Figure 2.

Insert Table 2 around here

Ethic Origin, Sector and Gender

Table 2 further tabulates the average callback rates by gender. Within both private and public firms, male applicants experience a higher callback rate than female applicants. With respect to private firms, this difference between female Danish and ethnic applications is 5.1%-points while

it is 2%-points for male applicants. The reverse pattern is evident for applications sent to public firms. While applications attached with a female name experience a Danish-ethnic callback difference of 13.3%-points, the male equivalents experience a 14.8%-points difference. However, these differences are marginal and, as explained below, might not be constant across the four different job types in the sample. Furthermore, when looking at the credibility intervals for the differences in proportions, there is a substantial overlap. Thus, our data present evidence barely worth mentioning in favour of gender differences between public and private firms.

Ethnic Origin, Sector and Quality

Table 3 tabulates the average callback rates first by sector and then by quality. Because we drew the applications at random from our applications pool for each job opening, it had the consequence that the total number of low quality applications differed from the total number of high quality applications. This difference was largest for private firms, while the total number came close for public firms.

For public firms, our quality differentiation seems to have worked as intended. Both Danish and ethnic applicants have a higher callback rate when their name was applied to a high quality compare to a low quality application. For instance, the callback rate for ethnic applicants grows from 20.5% to 28% when comparing a low to a high quality application. The difference in callback rate between Danish and ethnic applications drops from 15.3%-points to 12.8%-points when comparing low and high quality applications. This indicates that the discriminating effect of attaching an ethnic name to an application is lower if the application is of higher quality. However, as it is evident from the credibility intervals, there is little evidence that the discrimination effect is quality dependent. The Bayes factors show that there is stronger evidence for discrimination against applicants sending low quality application than for high quality applications.

Insert Table 3 around here

Insert Table 4 around here

For private firms, our quality measure did not work as intended. On the contrary, a low quality application actually has a higher callback rate than one of high quality. This is cause for concern as to whether the quality signals in the applications is working correctly. Still, we observe the same quality dependence as for public firms, namely that the discriminating name-effect is smaller for high quality applications (4.8%-points vs. 1.8%-points).

In sum, we observe some marginal effect of quality on name-based discrimination both within private and public firms such that higher quality applicants are discriminated against less. However, due to the uncertainty about our parameter estimates and the potential problems with the quality differentiation, we should be cautious concluding that callback discrimination depends on application quality.

Ethnic Origin, Sector and Job Type

Table 4 tabulates the average callback rates first by job type and next by sector. As we do not control which job openings are from which sectors, public and private applications are unevenly distributed across job types. Controller and office jobs opening are mostly from the private sector while teacher and physiotherapist jobs are mostly from the public sector. Applications sent to high school teacher job openings were by far the most successful ones with relatively high callback rates. This job type is also the one where we have the strongest evidence of name-based discrimination. Job applications attached with an ethnic name were 21.6%-points less likely to receive a positive callback compared to applications attached with a Danish name when sent to teaching openings from the private sector. Albeit substantial, this difference in callback is very

uncertain due to the small number of total private teaching jobs. Thus, the Bayes factor indicates that it is only 1.3 times more likely that ethnicity and callback are not independent than the alternative, which is a result barely worth mentioning. For the public jobs, however, we are much more certain about the discrimination. The Bayes factor shows that it is almost 77 times more likely that the data are dependent compared to the alternative. For public firms, we observe a 20.5%-points difference in callback favoring the applications attached with Danish names and there is 90% probability that this difference in callback lies within the credibility interval 8.9%-points and 31.1%-points.

From the table, it is clear that the cells quickly become sparse with regard to the other three job types. This is due to a much lower callback rate within the other three job categories. As the Bayes factors indicate, stratifying the sample into categories in this fashion leaves too small a sample size to provide convincing evidence for the hypothesis that the callback and ethnicity are not independent. However, where we have more than a couple of positive callbacks, we see the same pattern, namely that public firms tend to discriminate more based on the attached name than private firms do.

In sum, the discrimination difference between public and private firms seems to be largely driven by a relatively larger sample and callback rate within the teacher category. However, the tendency of larger name-based discrimination by public relative to private firms can still be seen across job types. To achieve more precise estimates of the discrimination effect between public and private firms and in order to more fully exploit our experimental setting while controlling for difference between job types, we turn to hierarchical regression techniques in the next section.

Hierarchical Bayesian Analysis

One caveat to the results presented above is that data is nested in structure as multiple applications are sent to the same organization. As such the applications are not independent but connected in groups of four by the organization to which they are sent. To take this feature of the data into account and dig deeper into our results, we specify a series of hierarchical logistic regression models. These models allow us to model regression coefficients for particular groups and provide us with reasonable estimates even when the sample size gets small in some of our job categories (Gelman & Hill, 2007). Further, our Bayesian approach is extremely well-suited for modeling how the causal structure that operates on the individual level varies across sectors (Jackman, 2009) while making the estimation very clean and direct with distributional statements (Gill, 2015).

First, a regular logit model is fit treating individual application-level and firm-level data the same. Second, the model is expanded to a hierarchical model including random intercepts for each firm in the sample. Third, the hierarchical model is modified to let the effect of ethnic origin signaled by the attached name vary by sector. Finally, controls for the different job types are included. We use solely weakly informative priors with a mean of zero and a precision of 10^{-4} . We estimate the models using Markov Chain Monte Carlo methods using 15,000 draws with an 11,000-draw burn-in from the JAGS program. Visual inspection of traceplots and the Geweke Time-Series Diagnostic indicate very satisfactory convergence using three chains (Appendices A.2-A.6). Letting $H_1 = P(\text{Difference} > 0)$ and $H_2 = P(\text{Difference} \leq 0)$, the Bayes factors are calculated assuming that the hypotheses H_1 and H_2 are equally likely a priori, thus reducing the calculation to the posterior odds in favor of H_1 (Gill, 2015; Kass & Raftery, 1995).

Discrimination by Ethnic Origin

Figure 1 shows the posterior distributions for the logit and hierarchical logit, respectively. The effect is substantial with a change in log odds with a mean of -0.556 and a CI spanning firmly in the negative region. A negative sign on the logit coefficient indicates a lower probability of callback when we had attached an ethnic name to the application. Precisely how much lower is evident from the next distribution that shows the predicted probabilities for Danish and ethnic applications, respectively. These results are almost identical to the results presented in the bottom row of Tables 1-4. The same is the case for the difference in probabilities. The conclusion from above is confirmed, namely that there exists a substantial discrimination based on the ethnic origin of the attached name.

Insert Figure 1 around here

Next, we exploit that each job opening is in itself its own experiment with four similar applications and a randomly attached name. We perform a hierarchical logit with a random intercept for each firm in the sample. The posterior distribution of the fixed effects logit coefficient is available in the bottom left corner of Figure 1. That the procedure more effectively uses the information from the data is clear from the posterior distributions of the predicted probabilities for an average firm. The mean estimates are identical, but our CIs are narrower making us more certain about the estimated difference in callback rate. With this model, the true difference in callback lies within the interval -13.3%-points and -6.1%-points.

Discrimination by Ethnic Origin and Sector

Hierarchically modelling the callback probability, we can now obtain a precise estimate of the difference in discrimination between public and private companies while controlling for the

different job types in our sample. In Figure 2, we have let the effect vary across sector type. To the left, we have the posterior distributions of the name-based differences within public and private firms for an average firm, respectively. Note that these probabilities are identical to the ones presented in Table 1.

Besides a graphical illustration of the difference in callback, the hierarchical model provides us with the possibility of directly assessing the difference between public and private firms without the number of observations in the cells gets too small. This difference-in-difference is presented to the right in Figure 2. We observe a mean difference of 10.7%-points, which is a substantial discrimination difference. With a probability of 90%, the true difference lies within the interval between 4.3%-points and 17.8%-points. The probability that public firms discriminate more than private firms is, on average, 0.996. Assuming equally likely hypotheses a priori, the Bayes factor indicates that it is 252 times more likely that there is a positive difference than the difference is equal to or below zero. This is decisive evidence that public institutions in our sample, on average, discriminate more than private firms do.

Discrimination by Ethnic Origin and Sector Controlling for Job Types

Table 4 gave reason to suspect that the average difference in discrimination we observe between public and private firms might be driven by the high school teacher category alone. To investigate if this is the case, we control for job types in our hierarchical model and predict difference in callback probabilities between public and private firms for each job category. The hierarchical model allows us to get reasonable predictions for particular groups even if the available sample is very sparse (Gelman & Hill, 2007).

 Insert Figure 2 around here

The posterior distributions of these differences are presented in Figure 3 with an extra vertical dashed line indicating no difference. Common for all four job types is that an overwhelming part of the mass of the distributions lies above zero indicating a positive discrimination difference. As expected, the largest mean difference is found among the high school teachers who discriminate close to 11.3%-points on average. This substantial difference is also the most uncertain one indicated by the wide CI. Nevertheless, there is a 93.5% probability that the difference is indeed above zero and the Bayes factor indicates positive evidence of a difference between the public and private sector.

The mean difference for the other three categories is much smaller ranging from 4.5%-points for the physiotherapists to 5.9%-points for the controller. All three have a larger than 96% probability that there is a higher name-based discrimination within public firms compared to private and all Bayes factors indicate strong evidence of sector discriminatory differences.

In sum, these results suggest strong evidence of difference in discrimination within the non-teacher categories, and positive evidence for one within the teacher category. The mean effect difference within the Teacher category is by far the most substantial.

Insert Figure 3 around here

DISCUSSION

The results of our field experiment show a substantial discrimination of applicants with typical names of non-western immigrants in Denmark. In addition, we do not find very convincing evidence that gender or applicant quality moderates the discriminating name-effect. Because we randomized the names, Danish and ethnic applicants should have received similar callback rates

on average, if the review process were neutral. Thus, our results show that employers use the name of an applicant as a selection factor in the review process. Moreover, these results are in line with previous non-experimental studies investigating experienced discrimination done in Denmark (Mogensen & Matthiesen, 2002; Møller & Togeby, 1999) and field experimental studies done in Sweden (Carlsson & Rooth, 2007) and the US (Bertrand & Mullainathan, 2004).

We find very strong evidence of a difference in how much private and public discriminate against applicants with an ethnic name. Although this difference is mostly prevalent within the school teacher category, our Bayesian analysis provides more precise inferences for the difference in effect between sectors within each occupational category. These results show positive evidence that within our four job categories public sector employers discriminate more based on the applicant's name than private sector employers do. We find it very interesting that the effect is considerably stronger within the school teacher category. These results highlight a little discussed sector difference. While public organizations are sometimes referred to as a preferred employer because of better salaries and general working conditions apparently it is difficult for ethnic minorities to get in. This is an interesting finding that should be followed up on. Future studies will have to include other job categories (and other countries) to investigate whether this difference in effect can be replicated in other occupational categories (and countries). If the public sector is more likely to discriminate against ethnic minorities in their hiring decisions it calls into question some of the values (such as equality) often thought to characterize public employment. Also it brings into question whether the public sector is equally capable of servicing all parts of society. The results of this study indicate that there may be something to learn from otherwise similar private sector organizations.

It is somewhat surprising that we find substantial discrimination within schools and a little more pronounced within the public sector. Previous research has pointed out that ethnic representation may be positive for school outcomes (Pitts, 2005) yet this may either not be the case in Denmark or employers prefer ethnic Danish employees for reasons unrelated to their stated skills and experience. As schools are fundamental in installing values in their students it appears problematic that some members of society are disadvantaged in participating in the process solely because of their name. Future research should explore this find more carefully to try to identify the causes of the discriminatory practices.

The two dominant theoretical branches used to describe unequal treatment effects triggered by a name are statistical (Arrow, 1973) and taste-based discrimination (Becker, 1957). As discussed in the experimental studies, it is unlikely that statistical discrimination theories can explain this difference as education and previous work experience is easily observable from the applications. Still, employers might use ethnicity as an indicator for unobservable worker productivity. Taste-based theories do not fare any better, as we cannot find larger racial bias among occupational categories that require higher customer contact (e.g. physiotherapist).

Bertrand and Mullainathan (2004) suggest that one perhaps more plausible explanation is so-called lexicographic search where the employers use heuristics such as “if an application has attached to it an Ethnic name, discard it”. Because employers would never reach the part of the application where the applicant’s skills are listed, this would explain why we find little evidence of a moderating effect of quality amongst private employers. It does not explain, however, why we find that public sector employers discriminate less when reviewing ethnic applications of higher quality. More importantly for this study, it fails to explain why we see a difference in discrimination between the public and private sector at all. Thus, more studies are needed to

unravel the underlying mechanisms driving differences in discrimination across the public and private sector.

This study illustrates the applicability of the audit methodology to research topics in public administration. Another recent example is Grohs and colleagues' study (2015) of responses to citizen requests in Germany. That study, unlike the present one, found little evidence of ethnic discrimination. By creatively thinking about the use of field experimental methods like these, it might be possible to move attention to new questions surrounding organizational diversity and representative bureaucracy. Unlike laboratory experiments or the widely used survey experiments, the advantage of the audit method is that it is possible to obtain data about the actual practice that the research is interested in.

Like any other study there are some limitations to be noted. First, our experiment focuses only on whether an applicant is called to an interview or not, and not whether the person in questions actually gets the job. However, we expect that a lower interview rate would result in a lower job success rate. Second, we only use personal names to represent ethnicity. One concern is that the names might signal social status and therefore the observed effect might be mixing racial and social biases. Third, we are not covering other channels of being called to an interview such as social networks. Our results are sensitive to if employers who more often hire non-western immigrants more often use their networks in the hiring process or non-western immigrants primarily get jobs through their social networks (Pager, 2007). Finally, we are only able to assess the presence of discrimination not the underlying reasons. Employers may discriminate because of racism but also because of prior experiences with a certain type of employees which unfairly colors subsequent hiring decisions. Future research should investigate this important question.

REFERENCES

- Agerström, J., & Rooth, D.-O. 2011. The role of automatic obesity stereotypes in real hiring discrimination. *Journal of applied psychology*, 96(4): 790.
- Alkadry, M. G., & Tower, L. E. 2006. Unequal pay: The role of gender. *Public Administration Review*, 66(6): 888-898.
- Alkadry, M. G., & Tower, L. E. 2011. Covert pay discrimination: How authority predicts pay differences between women and men. *Public Administration Review*, 71(5): 740-750.
- Andrzejewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., & Wagenmakers, E.-J. 2015. An introduction to Bayesian hypothesis testing for management research. *Journal of Management*, 41: 521-543.
- Arrow, K. J. 1973. The theory of discrimination. In O. Ashenfelter, & A. Rees (Eds.), *Discrimination in Labor Markets*: 3-33. Princeton, NJ: Princeton University Press.
- Becker, G. S. 1957. The economics of discrimination.
- Becker, G. S. 1971. The Economics of Discrimination. *University of Chicago Press Economics Books*.
- Berger, J. O., & Delampady, M. 1987. Testing precise hypotheses. *Statistical Science*, 2: 317-352.
- Bertrand, M., & Mullainathan, S. 2003. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination: National Bureau of Economic Research.
- Bertrand, M., & Mullainathan, S. 2004. Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review*, 94: 991.

- Boyne, G. 2002. Public and private management: what's the difference? *Journal of management studies*, 39(Journal Article): 97-122.
- Boyne, G. A. 2005. Where Next? Research Directions on Performance in Public Organizations. *Journal of Public Administration Research and Theory*, 15: 633-639.
- Bååth, R. 2013. *Bayesian First Aid*.
- Carlsson, M., & Rooth, D.-O. 2007. Evidence of ethnic discrimination in the Swedish labor market using experimental data. *Labour Economics*, 14(4): 716-729.
- Chun, Y. H., & Rainey, H. G. 2005. Goal ambiguity and organizational performance in US federal agencies. *Journal of Public Administration Research and Theory*, 15(4): 529-557.
- Gallistel, C. R. 2009. The importance of proving the null. *Psychological Review*, 116: 439-453.
- Gelman, A., & Hill, J. 2007. *Data analysis using regression and multilevel/hierarchical models*: Cambridge University Press New York.
- Gill, J. 1999. The insignificance of null hypothesis significance testing. *Political Research Quarterly*, 52: 647-674.
- Gill, J. 2015. Bayesian Methods: A Social and Behavioral Sciences Approach.
- Gill, J., & Meier, K. J. 2000. Public administration research and practice: A methodological manifesto. *Journal of Public Administration Research and Theory*, 10(1): 157-199.
- Gill, J., & Witko, C. 2013. Bayesian Analytical Methods: A Methodological Prescription for Public Administration. *Journal of Public Administration Research and Theory*, 23: 457-494.
- Grohs, S., Adam, C., & Knill, C. 2015. Are Some Citizens More Equal than Others? Evidence from a Field Experiment. *Public Administration Review*.

- Gunel, E., & Dickey, J. 1974. Bayes Factors for Independence in Contingency Tables. *Biometrika*, 61: 545-557.
- Hopkins, A. H. 1980. Perceptions of employment discrimination in the public sector. *Public Administration Review*: 131-137.
- Jackman, S. 2009. Bayesian Analysis for the Social Sciences.
- Jeffreys, H. 1961. Theory of probability.
- Kass, R. E., & Raftery, A. E. 1995. Bayes factors. *Journal of American Statistical Association*, 90: 773-795.
- Leasher, M. K., & Miller, C. E. 2012. Discrimination across the sectors: A comparison of discrimination trends in private and public organizations. *Public Personnel Management*, 41(2): 281-326.
- Lindley, D. V. 1993. The analysis of experimental data: The appreciation of tea and wine. *Teaching Statistics*, 15: 22-25.
- Llorens, J. J., Wenger, J. B., & Kellough, J. E. 2008. Choosing public sector employment: The impact of wages on the representation of women and minorities in state bureaucracies. *Journal of Public Administration Research and Theory*, 18(3): 397-413.
- Mackey, T. B., Barney, J. B., & Dotson, J. 2015. Corporate diversification and the value of individual firms: A bayesian approach. *Strategic Management Journal*.
- Meier, K. J., & O'Toole Jr, L. J. 2013. I Think (I Am Doing Well), Therefore I Am: Assessing the Validity of Administrators' Self-Assessments of Performance. *International Public Management Journal*, 16(1): 1-27.
- Mogensen, G. V., & Matthiesen, P. H. 2002. Indvanderne og arbejdsmarkedet – Mødet med det danske velfærdssamfund.

- Morey, R. D., Hoekstra, R., Rouder, J., Lee, M., & Wagenmakers, E.-J. 2015. The Fallacy of Placing Confidence in Confidence Intervals. *Psychonomic Bulletin & Review*: 1-21.
- Morey, R. D., & Rouder, J. N. 2015. *BayesFactor: Computation of Bayes Factors for Common Designs*.
- Møller, B., & Togeby, L. 1999. Oplevet diskrimination. En undersøgelse blandt etniske minoriteter.
- Naff, K. C. 1995. Perceptions of Discrimination. *Policy Studies Journal*, 23(3): 483-498.
- Orlitzky, M. 2011. How Can Significance Tests Be Deinstitutionalized? *Organizational Research Methods*, 15: 199-228.
- Pager, D. 2007. The use of field experiments for studies of employment discrimination: Contributions, critiques, and directions for the future. *The Annals of the American Academy of Political and Social Science*, 609(1): 104-133.
- Pager, D., Western, B., & Bonikowski, B. 2009. Discrimination in a low-wage labor market a field experiment. *American Sociological Review*, 74(5): 777-799.
- Perry, J. L. 2000. Bringing Society in: Toward a Theory of Public-Service Motivation. *J-PART*, 10(2): 471-488.
- Pitts, D. W. 2005. Diversity, representation, and performance: Evidence about race and ethnicity in public organizations. *Journal of Public Administration Research and Theory*, 15(4): 615-631.
- Plummer, M. 2015. *rjags: Bayesian Graphical Models using MCMC*.
- R Core Team, R. 2015. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rainey, H. G. 2009. *Understanding and managing public organizations*: Jossey-Bass.

Wagenmakers, E.-J. 2007. A practical solution to the pervasive problems of p-values.

Psychonomic Bulletin & Review, 17: 779-804.

Zyphur, M. J., Oswald, F. L., & Rupp, D. E. 2015. Rendezvous Overdue Bayes Analysis Meets

Organizational Research. *JOURNAL OF MANAGEMENT*, 41(2): 387-389.

TABLE 1. Callback Averages: Sector Difference

		Callback				Difference in proportions					
Sector	Total	Danish (D)		Ethnic (E)		D - E	95% CI		P(D>E)	Bayes factor (Bf)	
		n	Rate	n	Rate		Low	High		Bf	1 / Bf
Private	192	27	0.141	20	0.104	0.036	-0.031	0.101	0.8535	0.150	6.667
Public	252	97	0.385	61	0.242	0.142	0.064	0.223	0.9995	40.869	0.024
	444	124	0.279	81	0.182	0.096	0.041	0.151	0.9998	25.149	0.04

TABLE 2. Callback Averages: Sector Differences and Gender

		Callback					Difference in proportions					
Sector	Gender	Total	Danish (D)		Ethnic (E)		D - E	95% CI		P(D>E)	Bayes factor (Bf)	
			n	Rate	n	Rate		Low	High		Bf	1 / Bf
Private	Female	96	14	0.146	9	0.094	0.051	-0.040	0.145	0.8686	0.213	4.695
Private	Male	96	13	0.135	11	0.115	0.020	-0.073	0.118	0.6585	0.131	7.634
Public	Female	126	45	0.357	28	0.222	0.133	0.017	0.240	0.9905	2.277	0.439
Public	Male	126	52	0.413	33	0.262	0.148	0.032	0.264	0.9934	3.606	0.277
		444	124	0.279	81	0.182	0.096	0.041	0.151	0.9998	25.149	0.04

TABLE 3. Callback Averages: Sector Differences and Quality

Sector	Quality	Callback						Difference in proportions				Bayes factor (Bf)	
		Danish (D)			Ethnic (E)			D - E	95% CI		P(D>E)		
		Total	n	Rate	Total	n	Rate		Low	High			
Private	Low quality	104	17	0.163	88	10	0.114	0.048	-0.051	0.147	0.8307	0.200	5.000
Private	High quality	88	10	0.114	104	10	0.096	0.018	-0.069	0.109	0.6538	0.120	8.333
Public	Low quality	125	45	0.360	127	26	0.205	0.153	0.042	0.260	0.9968	5.952	0.168
Public	High quality	127	52	0.409	125	35	0.280	0.128	0.014	0.245	0.9837	1.520	0.658
		444	124	0.279	444	81	0.182	0.096	0.041	0.151	0.9998	25.149	0.04

TABLE 4. Callback Averages: Sector Differences and Occupational Categories

Job type	Sector	Total	Callback				Difference in proportions				Bayes factor (Bf)	
			Danish (D)		Ethnic (E)		D - E	95% CI		P(D>E)		
			n	Rate	n	Rate		Low	High		Bf	1 / Bf
Teacher	Private	26	14	0.538	8	0.308	0.216	-0.029	0.467	0.9516	1.304	0.767
Teacher	Public	144	84	0.583	54	0.375	0.205	0.089	0.315	0.9998	76.846	0.013
Physiotherapist	Private	12	1	0.083	2	0.167	-0.071	-0.354	0.201	0.2948	0.378	2.646
Physiotherapist	Public	78	7	0.090	4	0.051	0.038	-0.050	0.119	0.8171	0.155	6.452
Controller	Private	72	5	0.069	5	0.069	0.001	-0.092	0.085	0.5100	0.107	9.346
Controller	Public	24	6	0.250	3	0.125	0.115	-0.109	0.328	0.8528	0.483	2.070
Office	Private	82	7	0.085	5	0.061	0.024	-0.060	0.108	0.7204	0.121	8.264
Office	Public	6	0	0.000	0	0.000	0.001	-0.311	0.334	0.5019	0.265	3.774
		444	124	0.279	81	0.182	0.096	0.041	0.151	0.9998	25.149	0.04

FIGURE 1. Callback rates and Ethnic Origin: Logit and Hierarchical Logit

FIGURE 2. Hierarchical Logit: Callback rates, Ethnic Origin, and Sector Differences

FIGURE 3. Hierarchical Logit: Comparing across Job Types

APPENDIX

A.1. Names used in the experiment

Danish names	Ethnic names
Peter Jensen	Ali Aslam
Jens Nielsen	Mohammad Ashraf
Lars Hansen	Ahmad Afzal
Anne Pedersen	Mariam Yildrim
Kirsten Andersen	Fatima Akhtar
Hanne Christensen	Yasmin Yilmaz

Notes: This table lists the names used in the experiment.

A.2. Traceplots

A.3. Running Means

A.4. Autocorrelation Plots

A.5. Geweke Diagnostics

A.6. Potential Scale Reducing Factor