Pre-Analysis Plan.

Discrimination Against the Entrepreneurs and Ethinc Minorities: Field Experiment in Russia.

October 16, 2017

1 Introduction

Entrepreneurs stay in business despite both lower initial earnings and slower growth of income than in paid employment (Hamilton et al., 2000; Moskowitz and Vissing-Jorgensen, 2002; Åstebro et al., 2004). This puzzling behavior is commonly explained by nonpecuniary taste-based factors like preferences for autonomy or control (see for review Åstebro et al., 2014). Though specific preferences, of course, play a role, the entrepreneurs can stay self-employed simply because they can not find a job since prospective employers could prefer candidates with corporate experience.

Some studies show discrimination of entrepreneurs at the labor market (e.g. Failla et al., 2017). However, these evidence either correlation or inconclusive. Moreover, little is known why employees discriminate the entrepreneurs: is it taste-based discrimination (Becker, 1971) or statistical one (Phelps, 1972; Arrow, 1973)? We provide correspondence experiment (Bertrand and Mullainathan, 2004) in Russia to fill this gap Specifically, we try to understand (1) if entrepreneurial discrimination exists; (2) if it exists, is it statistical or taste based; (3) if providing the goal for applying can reduce the level of discrimination. In addition, we measure the level of ethnic discrimination.

In this pre-analysis plan, we describe the methodology that we will follow to conduct our analysis. The next section provides an overview of the study. The third and fourth section will describe the hypotheses and a specification that we plan to use ot test them. The fifth section reports the power analysis. The final section will discuss the methodology that we plan to use for post-hoc, explorative analysis. The detailed description of experimental design procedure, and sample size can be found in accompanying master thesis by Maria Mavlikeeva (Submitted: 10th of April. Defended: 13 April 2017, University of Kassel).

2 Overview of the Study

We provide the correspondence experiment in Russia. Overall, we send 12.000 fictitious resumes in response to 3.000 real job advertisement in Russia and measure the rate of call-backs as an indicator of interest in the applicant.

2.1 Discrimination against Entrepreneurs

To elicit discrimination towards entrepreneurs we send 8.000 CV's in response to 2.000 vacancies. We randomly assign if (1) a person was self-employed or worked for a company

and (2) whether motivation section in CV is provided.[1]

To assess if the entrepreneurial discrimination is statistically based, we apply for the similar occupations with different skill-levels: High (Skill level 4) or low(skill level 2) skill level job according to International Standard Classification of Occupations (2008). Namely, we send resume on the next vacancies in Skill level 4: Finance Managers (1211), Advertising and Public Relations Managers (1222), Information and Communications Technology Services Managers (1330). In skill level 2: Accounting Associate Professionals (3313); Conference and Event Planners (3332); Information and Communications Technology Operations Technicians (3511).[2]

2.2 Discrimination against Ethnic Minorities

To asses racial discrimination we send 4.000 CV's in the response of 1.000 vacancies.We randomly vary the Slavic and ethnic minorities sounding names for the resumes with equal quality e.g. Alexey Lebedev vs. Ansar Juraev. Also, we vary if the person has work experience (on average 1.5 year) or (s)he recently graduated. We choose broad range of occupations with highest demand and that require or does not face-to-face interaction with clients. Thus, we chose next six occupations for this part of experiment: Advertising and Marketing Professionals (2431, skill level 4), Real Estate Agents and Property Managers (3334, skill level 3), General Office Clerks (4110, skill level 2), Answering Service Operator (4223, skill level 2), Receptionists (general) (4226, skill level 2), Messengers, Package Deliverers and Luggage Porters (9621, skill level 1).[3]

2.3 Experimental Procedure

Practitioners expertise. We interviewed a number HR managers of large and middle companies. They interview showed that (A) applying for a specific position by sending electronic letters with CV attached to the e-mail of employee is acceptable and can lead to call for the interview and (B) they negatively perceive the CV's with an entrepreneurial background. For instance, one of the human resources managers from large recruiting company wrote: "Entrepreneurial experience adversely affects the chance of hiring candidate... Generally, the candidates with this experience are not invited to the interview".

Computer software for correspondence study
We develop a computer program for this experiment. It contains three main functions:

- Resumes generating

- Mailing

- E-mail Responses tracking

CV characteristics database and CV Generation. To create the resumes, we analyzed a large number of real CVs posted online. With the help of these CVs, we created a database for resumes' sections. We also made a database of names, surnames and middlenames based on most popular slavic and ethnic minorities names.

---

[1]For example, the in motivation section applicants write: "I would like to work in a company, where my professional skills and personal qualities could be a significant contribution to the long-term success of the venture."

[2]International Standard Classification of Occupations codes 2008 in brackets.

[3]International Standard Classification of Occupations codes 2008 in brackets.

All the resumes contain the following sections: Heading (name, mobile telephone, e-mail, birth date); Work experience; Education; Professional skills; Personal qualities; Advanced training, Language Skills. The program randomly choose gender of applicant and construct thier full name from corresponding part of names database (according to the type of CV). Next, it takes randomly lines from every section (according to the type of CV) in order to create a resume. The formatting of the resumes is also chosen at random. All the characteristics of the resume stored in MySQL database with unique CV identifier number.

All this helps us to address the Heckman critique (Heckman, 1998) as suggested by Carlsson et al. (2014). Thus, it gives us the possibility to interpret the ratios as suggested by Neumark (2012) and to estimate wage differentials in lines with Lanning (2013).

Mailing. We apply within-subject design: For each vacancy computer program sends four types of the resumes in random order.

All applications contain a cover letter, where candidates express an interest in the posted job advertisement. The exact wording of the cover letter is chosen from a set of cover letters at random.To make the application realistic and targeted to specific vacancy the program contains an electronic form that is filled for every vacancy. The form contained following lines: job vacancy, email of a recipient, web source and the name of contact person (if given). This information is used to generate the cover letter.

"Dear (the name of contact person),

I am confident my experience, academic qualifications and interests would fulfill the requirements of the position (position name), posted on (web site). Please find my resume attached.

Yours sincerely,

(Name of the candidate) "

E-mail Responses tracking. Program automatically track the responses on e-mails that are unique for each resume.

Call-back tracking. Unique phone number is assigned to each treatment and job skill level. We plan to track phone calls on each type of treatment and job skill level. That allows us to have clear results with respect to the main treatments. To get the detailed picture of the characteristics of resume, we plan to match the call-back and corresponding CV.

3 Hypothesis

3.1 Discrimination against Entrepreneurs

Some studies examining entrepreneurs' satisfaction and interviews with human resource managers conducted within the framework of this experiment in Russia give the reason to conclude that:

Hypothesis E 1 (HE1)**.** *The callback rate is lower for self-employed than for candidates with corporate experience.*

We assume, that it is harder for prospective employers to find a qualified candidate for high skilled occupation. The selection process for such job openings usually takes more time. The employers gather resumes during the certain time period, as they need the larger number of candidates for these high positions to choose from. That gives us the reason to assume that:

Hypothesis E 2 (HE2)**.** *The callback rate for low skilled occupations is higher*

It is harder to assess the quality of high skilled applicants then low skilled applicants. Hence, if discrimination against entrepreneurs is statistical then decision makers can largely rely on visible features of CV and discriminate applicants with entrepreneurial experience more if they apply for high-skilled job.

Hypothesis E 3 (HE3). *The rate of entrepreneurial discrimination is higher for high skilled occupations*

As it was mentioned by human resource managers in the interviews, the fact that self-employed candidates are applying for corporate job raises the question why they want to change the career. Some resumes of this experiment contain the motivation section, where applicants try to answer possible employers' questions in advance. Thus we present the following hypothesizes:

Hypothesis E 4 (HE4). *Motivation section in resume increases the callback rate.*

Hypothesis E 5 (HE5). *Motivation section in resume decrease entrepreneurial discrimination.*

3.2 Discrimination against Ethnic Minorities

Hypothesis M 1 (HM1). *The callback rate for candidates with non-Slavic names is smaller than for candidates with Slavic names.*

As it is found by Bertrand and Mullainathan (2004), more experinced candidates receives more call-back. Thus, we assume that the same effect will be found in this study:

Hypothesis M 2 (HM2). *The presence of work experience increases the rate of the callback.*

Based on the evidence presented by Bertrand and Mullainathan (2004) that candidates with ethnic sounding names receive 50 percent fewer callbacks (3.2 percent point less), but one year of epxerience increase the rate of call-back by 0.4 percent point. Thus, we assume:

Hypothesis M 3 (HM3). *Ethnic discrimination is higher than discrimination based on experience.*

Assuming that presence of previous work experience may serve as a positive signal for prospective employers, put it differently, ethnic discrimination has statistical nature, we hypothesize:

Hypothesis M 4 (HM4). *Ethnic discrimination is lower when an experience is present.*

4 Specifications

4.1 Basic Specification

We plan to measure the treatment effect on the probability that potential employer responds to the application either by phone or e-mail.

We plan to estimate the treatment effect on response to application using the mixed-effect probit regression with heteroskedasticity-robust standard errors. This regression takes the next general form:

$$Pr(Response = 1) = \Phi(\beta_0 + \sum_1^n \beta_T T_k + v_i), \tag{1}$$

where $Response$ - dummy variable that equals to 1 when employer responded either via phone or e-mail, Φ is the the standard cumulative normal distribution, T_i - is a treatment dummy for variable of interest k out of n variables, υ_i are random effects for vacancy i.

We plan to estimate the treatment effect using the probit regression with heteroskedasticity-robust standard errors clustered at vacancy level:

$$Pr(Response = 1) = \Phi(\beta_0 + \sum_1^n \beta_T T_k + c_i), \tag{2}$$

where c_i - unobserved clustered effect of vacancy i.

We plan to compare the proportions of callbacks contingent on treatment using simple proportion test and t-test. Finally, we plan to calculate the ratios as suggested by Neumark (2012) and to estimate wage differentials in lines with Lanning (2013).

4.1.1   Discrimination against Entrepreneurs

In case of entrepreneurial discrimination experiment the regression takes the next from:

$$Pr(Response = 1) = \Phi(\beta_0 + \beta_{Self} T_{Ent} + \beta_{Motiv} T_{Motiv} + \beta_H H + c_i), \tag{3}$$

where T_{Ent} – dummy that equals 1 if the applicant is self-employed; T_{Motiv} equals 1 if the candidate includes motivation for applying; H – 1 if high skilled job: ISCO-08 group 1 according to the International Standard Classification of Occupations.

To asses the type of discrimination we plan to use the following regression

$$\begin{aligned} Pr(Response = 1) = \Phi(\beta_0 + \beta_{Self} T_{Ent} + \beta_{Motiv} T_{Motiv} + \beta_H H + \\ + \beta_{Self \times H} T_{Ent} \times H + c_i), \end{aligned} \tag{4}$$

We plan to use the following regression to investigate if motivation decrease rate of entrepreneurial discrimination:

$$\begin{aligned} Pr(Response = 1) = \Phi(\beta_0 + \beta_{Self} T_{Ent} + \beta_{Motiv} T_{Motiv} + \beta_H H + \\ + \beta_{Self \times Motiv} T_{Ent} \times T_{Motiv} + c_i), \end{aligned} \tag{5}$$

We plan to make similar estimations using the probit regression with heteroscedasticity-robust standard errors clustered at vacancy level (see equation 2).

4.1.2   Ethnic Discrimination

To asses ethnic discrimination we plan to use the regression of the next from:

$$Pr(Response = 1) = \Phi(\beta_0 + \beta_{Slav} T_{Slav} + \beta_{Exp} T_{Exp} + c_i) \tag{6}$$

where T_{Slav} – dummy that equals 1 if the resume includes Slavic sounding name; T_{Exp} equals 1 if the candidate has any work experience.

We plan to use the following regression to examine if the ethnic discrimination is higher than discrimination based on experience:

$$Pr(Response = 1) = \Phi(\beta_0 + (\beta_{Slav} - \beta_{Exp})T_{Slav} + \beta_{Slav}(T_{exp} + T_{Slav}) + c_i) \tag{7}$$

To asses if the ethnic discrimination is lower in presence of experience we asses the next regression:

$$\begin{aligned} Pr(Response = 1) = \Phi(\beta_0 + \beta_{Slav} T_{Slav} + \beta_{Exp} T_{Exp} + \\ + \beta_{Slavic \times Exp} T_{Slavic} \times T_{Exp} + c_i) \end{aligned} \tag{8}$$

5

5 Power Analysis

We use computer similuation to asses the power of experiment. First, we test if simulation properly works and we get comprable results. Then, we simulate the experiment number of times and see in how many cases we will reject the null-hypothesis at 5% level using speficication from section 4 and given that the treatment effect is present.

5.1 Simulation Setup

Generally, we assume the rate of call backs $\approx 10\%$. We take the effect size from Bertrand and Mullainathan (2004). Thus, white-to-black ethnicity increase the chance of call by 50% (call back ratio-1.5) and presence of expereince increase the chance of call back by $\approx 20\%$ (call back ratio-1.2).

```
  set.seed(239)
  N<-10000 #Number of Job
  NCV <- rep(c(1:4), N)   # CV number -- Needed to make treatments number
  VAC <- rep(c(1:N), each = 4)   #Vacancy


  TE <- NCV%%2   #Dummy Ethnicity 1-slavic
  TEXP <- ifelse(NCV > 2, 1, 0)   #Dummy Experience 1- Exp
  ord <- c(replicate(N, sample(c(1:4), replace = FALSE)))
  DATA <- as.data.frame(cbind(VAC, NCV, TE, TEXP, ord))

  DATA$call <- rbinom(N * 4, 1, 0.1)
  le <- nrow(subset(DATA, TE == 0 & call == 1))
  DATA$call <- with(DATA, ifelse(TE == 0 & call == 1, rbinom(le, 1, 0.66),
      call))#1/1.5 ~ 0.66
  lexp <- nrow(subset(DATA, TEXP == 0 & call == 1))
  DATA$call <- with(DATA, ifelse(TEXP == 0 & call == 1, rbinom(lexp,
      1, 0.8), call)) #1/1.2 ~ 0.8
DATA$call<-as.factor(DATA$call)
levels(DATA$call)<-c("no","yes")

DATA$TE<-as.factor(DATA$TE)
levels(DATA$TE)<-c("Non-slavic","Slavic")
```

Let's compare the results with Bertrand and Mullainathan (2004).

```
require(AER)
require(xtable)
data(ResumeNames)

print(xtable(prop.table(xtabs(~TE + call, data = DATA), 1), caption = "Simulated Data",
    digits = 3), file = "ta.tex", floating = FALSE)

print(xtable(prop.table(xtabs(~relevel(ethnicity, ref = "afam") +
    call, data = ResumeNames), 1), digits = 3), file = "tb.tex",
    floating = FALSE)
```

	no	yes		no	yes
Non-slavic	0.940	0.060	afam	0.936	0.064
Slavic	0.908	0.092	cauc	0.903	0.097
(a) Simulated Data			(b) Bertrand&Mullainathan		

Table 1: Distribution of Calls by Ethnicity

We get similar results in both tables and the ratio of calls depending on the ethnicity ≈ 1.5 both in the simulation and the data from Bertrand and Mullainathan (2004).

5.2 Discrimination against Ethnic Minorities

We take a code from previous section and make a function "ONE_ET_SIM ". This function

1. Simulate the data from one experiment;

2. Run regression and tests according to specification from section 4;

3. Returns the p-values of the treatment dummies for each test.

```r
ONE_ET_SIM <- function(N) {
    NCV <- VAC <- TE <- TEXP <- ord <- DATA <- NULL
    NCV <- rep(c(1:4), N)  # CV number -- Needed to make treatments number
    VAC <- rep(c(1:N), each = 4)  #Vacancy


    TE <- NCV%%2  #Dummy Ethnicity 1-slavic
    TEXP <- ifelse(NCV > 2, 1, 0)  #Dummy Experience 1- Exp
    ord <- c(replicate(N, sample(c(1:4), replace = FALSE)))
    DATA <- as.data.frame(cbind(VAC, NCV, TE, TEXP, ord))

    DATA$call <- rbinom(N * 4, 1, 0.1)
    le <- nrow(subset(DATA, TE == 0 & call == 1))
    DATA$call <- with(DATA, ifelse(TE == 0 & call == 1, rbinom(le,
        1, 0.66), call))
    lexp <- nrow(subset(DATA, TEXP == 0 & call == 1))
    DATA$call <- with(DATA, ifelse(TEXP == 0 & call == 1, rbinom(lexp,
        1, 0.8), call))


    ### Ethinc Base
    glmer <- glmer(call ~ TE + TEXP + (1 | VAC), data = DATA,
        family = binomial(link = probit))
    geeglm <- geeglm(call ~ TE + TEXP, id = VAC, data = DATA,
        family = binomial(link = probit))



    glmerI <- glmer(call ~ TE + TEXP + TE * TEXP + (1 | VAC),
        data = DATA, family = binomial(link = probit))
    geeglmI <- geeglm(call ~ TE + TEXP + TE * TEXP, id = VAC,
        data = DATA, family = binomial(link = probit))

    # HM1. Ethinicity: Non-slavic sounding name call back rate is
    # smaller than for slavic name.
    pe <- linearHypothesis(glmer, "TE=0", white.adjust = TRUE)[2,
        "Pr(>Chisq)"]
    pecl <- coef(summary(geeglm))["TE", "Pr(>|W|)"]
    pepro <- prop.test(table(DATA$TE, DATA$call))$p.value
    pet <- t.test(DATA$call, DATA$TE, paired = T)$p.value

    # HM2. Experience: Expreince increase the rate of call-back.
    pexp <- linearHypothesis(glmer, "TEXP=0", white.adjust = TRUE)[2,
        "Pr(>Chisq)"]
    pexpcl <- coef(summary(geeglm))["TEXP", "Pr(>|W|)"]
    pexppro <- prop.test(table(DATA$TEXP, DATA$call))$p.value
    pexpt <- t.test(DATA$call, DATA$TEXP, paired = T)$p.value
```

```
# HM3. Ethinc disrimination is higher than discrimination
# based on experience
pevsexp <- linearHypothesis(glmer, c("TE=TEXP"), white.adjust = TRUE)[2,
    "Pr(>Chisq)"]

# HM4. Ehtinc discrimination is lower when experience is
# present.
pexpI <- linearHypothesis(glmerI, "TE:TEXP=0", white.adjust = TRUE)[2,
    "Pr(>Chisq)"]
pexpIcl <- coef(summary(geeglmI))["TE:TEXP", "Pr(>|W|)"]




    l <- c(pe, pecl, pepro, pet, pexp, pexpcl, pevsexp, pexppro,
        pexpt, pexpI, pexpIcl)
    l
}
```

We run this function 50 times (simulate the experiment 50 times) with different number of vacancies: 500, 1000, 1500.

```
require(lme4)
library(geepack)
require(car)
set.seed(239)
# ONE_ET_SIM(100)

pr500ET <- replicate(50, ONE_ET_SIM(500))


pr1000ET <- replicate(50, ONE_ET_SIM(1000))


pr1500ET <- replicate(50, ONE_ET_SIM(1500))
```
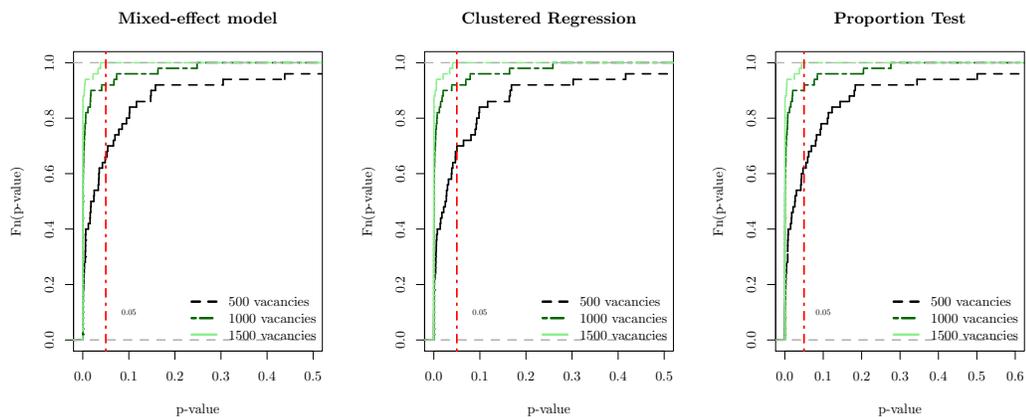

Figure 1: HM1. Cumulative Distribution Function of p-values

We observe on the figure 1 that with the sample of 1000 vacancies (4000 resumes) we reject the **(main)** null-hypothesis about absence of ethnic discrimination in more than

90% of cases (Power=0.9) at the 5% significance level and find evidence in favour of the hypothesis HM1.

5.2.1 Discrimination against Entrepreneurs

Similar to previous section, we assume the general rate of call backs $\approx 10\%$, but we take into account that high job applicants will receive lower number of calls ($\approx 6.25\%$). We assume that entrepreneurial experience decrease the chance of call back by 35% (call back ratio-1.35). We choose this ratio since we assume that enterprenerial experience considered by HR's similar to unemploeyd status but in lower extent. The call back ratio for employed to unemployed people ranges from 1.25 to 1.47 (Bertrand and Duflo, 2016).

```r
ONE_ENTR_SIM <- function(N, call, probit) {
    NCV <- NULL
    job <- NULL
    TE <- NULL
    TG <- NULL
    TC <- NULL
    ord <- NULL
    DATA <- NULL

    NCV <- rep(c(1:4), N)
    VAC <- rep(c(1:N), each = 4)

    TENT <- NCV%%2    #Dummy Self-employed
    TM <- ifelse(NCV > 2, 1, 0)   #Dummy 1- no Motivation
    Hjobtype <- VAC%%2   #1- ISCO group 1 (High skill job)
    ord <- c(replicate(N, sample(c(1:4), replace = FALSE)))

    DATA <- as.data.frame(cbind(NCV, VAC, TENT, TM, Hjobtype,
        ord))

    ## Initial rate of call-back
    DATA$call <- rbinom(N * 4, 1, 0.1)

    ## Self-emlpoyed call-back
    le <- nrow(subset(DATA, TENT == 1 & call == 1))
    DATA$call <- with(DATA, ifelse(TENT == 1 & call == 1, rbinom(le,
        1, 0.75), call))   #1/1.35 ~ 0.75

    ## High skilled job call-back
    lh <- nrow(subset(DATA, Hjobtype == 1 & call == 1))
    DATA$call <- with(DATA, ifelse(Hjobtype == 1 & call == 1,
        rbinom(lh, 1, 0.625), call))   # 1/1.6 ~ 0.625

    ## Self-emlpoyed call-back with motivation (higher)
    lM <- nrow(subset(DATA, TENT == 1 & TM == 1 & call == 0))
    DATA$call <- with(DATA, ifelse(TENT == 1 & TM == 1 & call ==
        0, rbinom(lM, 1, 0.02), call))   #Aprox 1/2 half of self-emploeyemnt effect

    ## Self-emlpoyed High skilled job call-back
    lc <- nrow(subset(DATA, TENT == 1 & Hjobtype == 1 & call ==
        1))
    DATA$call <- with(DATA, ifelse(TENT == 1 & Hjobtype == 1 &
        call == 1, rbinom(lc, 1, 0.75), call))

    glmer <- glmer(call ~ TENT + TM + Hjobtype + (1 | VAC), data = DATA,
        family = binomial(link = probit))
    geeglm <- geeglm(call ~ TENT + TM + Hjobtype, id = VAC, data = DATA,
        family = binomial(link = probit))
```

```r
    glmerIch <- glmer(call ~ TENT + TM + Hjobtype + TENT * Hjobtype +
        (1 | VAC), data = DATA, family = binomial(link = probit))
    geeglmIch <- geeglm(call ~ TENT + TM + Hjobtype + TENT *
        Hjobtype, id = VAC, data = DATA, family = binomial(link = probit))



    glmerIcm <- glmer(call ~ TENT + TM + Hjobtype + TENT * TM +
        (1 | VAC), data = DATA, family = binomial(link = probit))
    geeglmIcm <- geeglm(call ~ TENT + TM + Hjobtype + TENT *
        TM, id = VAC, data = DATA, family = binomial(link = probit))

    # HE1. The rate of call back is lower for Entrepreneur than
    # candidates with corporate experience
    ps <- linearHypothesis(glmer, "TENT=0", white.adjust = TRUE)[2,
        "Pr(>Chisq)"]
    pscl <- coef(summary(geeglm))["TENT", "Pr(>|W|)"]
    pspro <- prop.test(table(DATA$TENT, DATA$call))$p.value
    pst <- t.test(DATA$call, DATA$TENT, paired = T)$p.value

    # HE2. The rate of call back for high skilled occupations is
    # lower
    ph <- linearHypothesis(glmer, "Hjobtype=0", white.adjust = TRUE)[2,
        "Pr(>Chisq)"]
    phcl <- coef(summary(geeglm))["Hjobtype", "Pr(>|W|)"]
    phpro <- prop.test(table(DATA$Hjobtype, DATA$call))$p.value
    pht <- t.test(DATA$call, DATA$Hjobtype, paired = T)$p.value

    # HE3. The rate of enterprenerial discrimination is higher
    # for high skilled occupations
    psh <- linearHypothesis(glmerIch, "TENT:Hjobtype=0", white.adjust = TRUE)[2,
        "Pr(>Chisq)"]
    pshcl <- coef(summary(geeglmIch))["TENT:Hjobtype", "Pr(>|W|)"]


    # HE4.Motivation increase the rate of call-back
    pm <- linearHypothesis(glmer, "TM=0", white.adjust = TRUE)[2,
        "Pr(>Chisq)"]
    pmcl <- coef(summary(geeglm))["TM", "Pr(>|W|)"]
    pmpro <- prop.test(table(DATA$TM, DATA$call))$p.value
    pmt <- t.test(DATA$call, DATA$TM, paired = T)$p.value

    # HE5. Motivation decrease the enterperenail dsicrimination
    psm <- linearHypothesis(glmerIcm, "TENT:TM=0", white.adjust = TRUE)[2,
        "Pr(>Chisq)"]
    psmcl <- coef(summary(geeglmIcm))["TENT:TM", "Pr(>|W|)"]


    l <- c(ps, pscl, pspro, pst, ph, phcl, phpro, pht, psh, pshcl,
        pm, pmcl, pmpro, pmt, psm, psmcl)
    l
}



require(lme4)  # 'Glmer'
library(geepack)  #Clustered robust standard errrors Geeglm
require(car)  #for linear hypothesis(robust)
set.seed(239)
pr1500ENTR <- replicate(50, ONE_ENTR_SIM(1500))
```

```
pr2000ENTR <- replicate(50, ONE_ENTR_SIM(2000))


pr2500ENTR <- replicate(50, ONE_ENTR_SIM(2500))
```
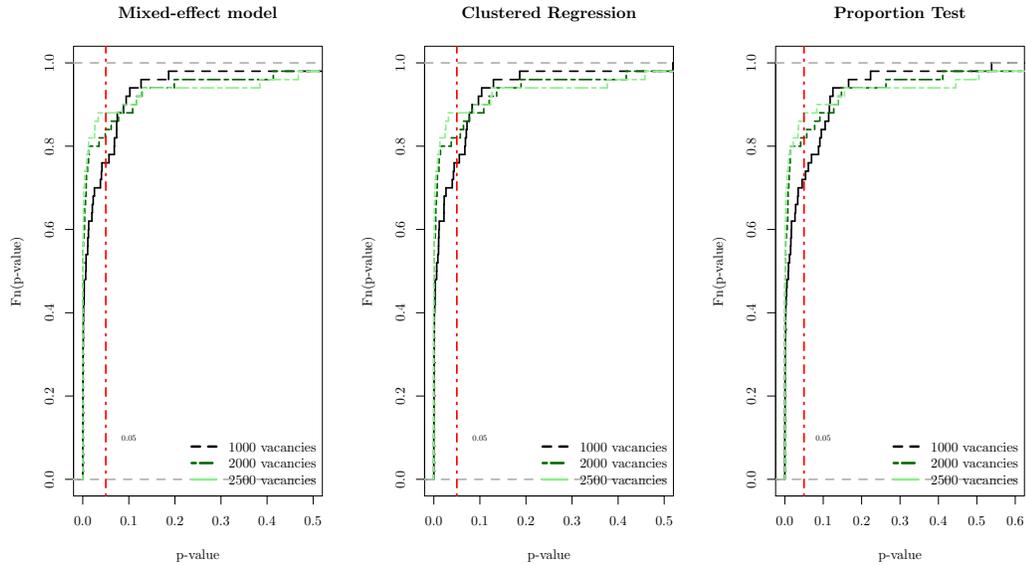

Figure 2: HE1. Cumulative Distribution Function of p-values

We observe on the figure 2 that with the sample of 2000 vacancies (8000 resumes) we reject the **(main)** null-hypothesis about the absence of discrimination against entrepneuers in more than 80% of cases (Power=0.8) at the 5% significance level and and find evidence in favour of the hypothesis HE1.

6   Exploratory analysis

To investiagte which other characterstics of candidates affects the interest of employer we plan to provide further explorative analysis. Namely, we plan to adress the next issues in our analysis:

1. **Heterogeneous treatment effects.** To test for the heterogeneity of effects, we will interact the individual characteristics of CV with the main treatment status: (1) entrepreneurial experience and (2) ethnicity of candidate.

2. **Individual Characteristics of Resume.** We plan to analyse which other randomly assigned characteristics of CV affects the rate of response.

3. **Sections of Resume.** We plan to investigate which section of resume has highest importance for the rate of response.

To deal with multilpe hypothesis testing problem we plan to use lasso regression and random forest aglorithm for variables selection (Imbens and Rubin, 2015). We also plan to

use double-lasso selection procedure (Belloni et al., 2013) to select the control characteristics of CV's and "causual trees" (Wager and Athey, 2015; Athey and Imbens, 2016) to asses hetergenous treatment effects.

References

Arrow, K., 1973. The Theory of Discrimination. In: Discrimination in Labor Markets. pp. 3–33.

Åstebro, T., Herz, H., Nanda, R., Weber, R. a., 2014. Seeking the Roots of Entrepreneurship: Insights from Behavioral Economics. Journal of Economic Perspectives 28 (3), 49–70.
URL http://pubs.aeaweb.org/doi/abs/10.1257/jep.28.3.49

Åstebro, T., Jeffrey, S., Adomdza, G., 2004. Inventor Perseverance After Being Told to Quit : The Role of Overconfidence and Optimism Inventor Perseverance After Being Told to Quit : The Role of Overconfidence and Optimism. Social Sciences, 1–25.

Athey, S., Imbens, G. W., 2016. Recursive Partitioning for Heterogeneous Casaul Effects. Proceedings of the National Academy of Sciences 113 (27), 1–43.
URL http://arxiv.org/abs/1510.04342 http://www.ncbi.nlm.nih.gov/pubmed/27382149

Becker, G. S. G. S., 1971. The economics of discrimination. No. October 1961.
URL http://books.google.es/books?id=rFlemcOSE7EC

Belloni, A., Chernozhukov, V., Hansen, C., 2013. Inference on treatment effects after selection among high-dimensional controls. Review of Economic Studies 81 (2), 608–650.

Bertrand, M., Duflo, E., 2016. Field Experiments on Discrimination. NBER Working Paper Series 22014 (1), 110.
URL http://www.nber.org/papers/w22014

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

Carlsson, M., Fumarco, L., Rooth, D.-o., 2014. Does the design of correspondence studies influence the measurement of discrimination? IZA Journal of Migration 3 (11), 1–17.
URL http://www.izajom.com/content/pdf/2193-9039-3-11.pdf

Failla, V., Melillo, F., Reichstein, T., 2017. Entrepreneurship and employment stability Job matching, labour market value, and personal commitment. Journal of Business Venturing 32 (2), 162–177.
URL file://www.sciencedirect.com/science/article/pii/S0883902617300083

Hamilton, B. H., Journal, S., June, N., Hamilton, B. H., 2000. Does Entrepreneurship Pay ? An Empirical Analysis of the Returns to Self Employment Published by : The University of Chicago Press Stable URL : http://www.jstor.org/stable/10.1086/262131 Does Entrepreneurship Pay ? An Empirical Analysis of the Returns t 108 (3), 604–631.

Heckman, J. J., 1998. Detecting Discrimination. Journal of Economic Perspectives 12 (2), 101–116.
URL http://pubs.aeaweb.org/doi/abs/10.1257/jep.12.2.101

Imbens, G. W., Rubin, D. B., 2015. Causal Inference for Statistics, Social, and Biomedical Sciences: An Introductio. Cambridge University Press.
  URL http://www.amazon.com/Causal-Inference-Statistics-Biomedical-Sciences/dp/0521885884

International Standard Classification of Occupations, 2008. International Standard Classification of Occupations. Isco-08 I, 1–420.

Lanning, J. A., 2013. Opportunities Denied , Wages Diminished : Using Search Theory to Translate Audit-Pair Study Findings into Wage Differentials 13 (2), 921–958.

Moskowitz, T. J., Vissing-Jorgensen, A., 2002. The Returns to Entrepeneurial Investment: A Private Equity Premium Puzzle. The American Economic Review 92 (4), 745—-778.

Neumark, D., 2012. Detecting Discrimination in Audit and Correspondence Studies. Journal of Human Resources 47 (4), 1128–1157.
  URL http://jhr.uwpress.org/lookup/doi/10.3368/jhr.47.4.1128

Phelps, E. S., 1972. The Statistical theory of Racism and Sexism. American Economic Review 62 (4), 659–661.

Wager, S., Athey, S., 2015. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests.
  URL http://arxiv.org/abs/1510.04342