CHRISTOPHER ZERRES

MARKETING

Schriftenreihe „Arbeitspapiere für Marketing und Management“

Herausgeber:
Prof. Dr. Christopher Zerres

Hochschule Offenburg
Fakultät Medien und Informationswesen

Arbeitspapier Nr. 52

Differences in Unemployment due to Sexual Orientation: Evidence from the Swedish Labour Market

Karaarslan, C.

Offenburg, Januar 2021

ISSN: 2510-4799

Impressum

Prof. Dr. Christopher Zerres
Hochschule Offenburg
Fakultät Medien und Informationswesen
Badstraße 24
77652 Offenburg

ISSN: 2510-4799

Inhalt

1 INTRODUCTION

The right to engage in work and choose an occupation to freely work at, is declared a fundamental human right in the EU. Behaviour that restrains somebody from doing so, due to sexual orientation discrimination for example, is prohibited.[1] Inquiries on the dimension of this particular behaviour, as well as the magnitude of harm it causes in the population, is of vital importance for policy makers and the entire civil society. A growing number of research pertaining to labour market outcomes due to sexual orientation has been conducted rececently. Most of the studies have been carried out in western countries, where annual income, hourly wages, labour market participation and employment decisions have been in the focus of researchers.[2] Ahmet, Andersson and Hammarstedt have been the pioneering scientists in this field in Sweden and contributed by extending their inquiries from the individual to the couple level (Ahmed, et al., 2011a) and to field experiments (Ahmed, et al., 2011b) in detecting discrimination against homosexuals.

The present paper aims to contribute to the labour market discrimination literature by estimating the differences in the employment probabilities and in the duration of unemployment by sexual orientation in Sweden using survival analysis techniques. Time-to-event data is rare in social sciences, which is particularly valid for data sets where the sexual orientation of individuals is observable. Due to this scarcity, the present study represents the first paper investigating the effect of sexual preferences on the duration in unemploymnet using survival analysis techniques. In contrast to other estimation methods, survival techniques enable us to incorporate the particular nature of time-to-event data, such as its particular skewness, strict non- negative nature, as well as censoring and truncation. Separately Zero Inflated Negative Binomial regression has been conducted to the duration in unemployment and Probit estimation to the event of getting employed, where differing significant outcomes by sexual orientation have been detected for some specifications.

The remainder of the article is organized as follows. A detailed literature review is provided in section 2. Section 3 comprises the theoretical framework of the investigation, while section 4 contains the methodical framework and research design. Data and descriptive statistics are presented in section 5, followed by the results of the estimation in section 6. Section 7 concludes with the final discussion.

[1] The Charter of Fundamental Rights of the European Union and its respective articles 15 and 21 are pertaining to the fields of work and discrimination, which is signed by Sweden as well. Sweden additionally adopted stringent legislation for the combat against discrimination and the promotion of equal rights and opportunities (s. Discrimination Act 2008:567).

[2] Section 2 of the present paper provides a detailed discussion of the labour market literature on discrimination due to sexual orientation.

2 LITERATURE REVIEW

2.1 LABOUR SUPPLY STATUS AND LABOUR SUPPLY INTENSITY

In contrast to previous studies, Leppel (2009) studies employment probabilities due to sexual orientation. Using US 2000 Census data she finds that homosexual women are more likely to be in the labour force than heterosexual married and unmarried partnered women, by 23 per cent. Lesbians are 20 per cent less likely to be employed than married heterosexual women, but 50 per cent more likely than unmarried partnered heterosexual women. Homosexual men are 71 per cent more likely to be out of the labour force than married heterosexual men and 24 per cent less likely than unmarried heterosexual partnered men. The probability of being unemployed is with 2.7 per cent compared to 1.4 per cent almost twice as high for gays than for heterosexual married men. Compared with unmarried partnered heterosexual men, gay men have an employment advantage of 13 per cent.

Attention has also been paid to differences in labour supply by sexual orientation. Indicated by the fact that hourly wages not necessarily differ between homo- and heterosexuals, not the whole variation in annual earnings could be explained by discrimination. Tebaldi and Elmslie (2006) present the first analysis about differences of labour supply due to sexual orientation. Using the US 2001 Current Population Survey (CPS) they find that homosexual women are 21 per cent more likely to be full- time employed, 13 per cent less likely to work in part- time jobs and supply 7 per cent more hours of work per week than heterosexual married women do. Formerly unemployed lesbians supply more hours of labour than their heterosexual peers. The probability of working full- time is 5 per cent lower for gay than for heterosexual married men. Gay men are 4 per cent more likely to work part- time and supply 8 per cent less hours of work per week. Recently unemployed gay men are also less likely than heterosexual men to be full- time employed and suffer from longer durations in unemployment.

Using US 2000 Census data Black *et al.* (2007) confirm the before mentioned findings. Gay males tend to supply less labour, while lesbian women tend to supply more in terms of hours supplied. This is in line with the finding, that smaller earnings differentials occur when hourly wages are studied instead of annual earnings.

2.2 DISCRIMINATION DUE TO SEXUAL ORIENTATION

Particular explanations have been put forward to the observed differences above, e.g. specialization within the household (Becker, 1981) and discrimination (Becker, 1957; Phelps, 1972). While investigations based on earnings data can only discuss probable explanations, field experiments provide a powerful tool in order to detect discrimination particularly (see Riach & Rich (2002) for the advantages).[3]

Experiments in the labour market were conducted by Adam (1981) in Toronto, Canada for the legal sector and Hebl *et al.* (2002) in Texas metropolis, USA, applying *personal situation tests*. Although evidence of discrimination was reported, as external tests show, these findings were not statistically significant (Ahmed, et al., 2011b). Those studies can only partly reflect the situation in the hiring process, since, beside the general problems anticipated with

[3] See Ahmed (2010a, 2010b), Ahmed and Hammarstedt (2008) and Ahmed et al. (2009, 2010, 2012) for further examples of field experiments in the investigation of discrimination in Sweden.

correspondence tests, where trained individuals are involved (see Riach & Rich (2002)) both investigations were not nationwide implemented, only conducted in a single occupation and thus had very small sample sizes.

Weichselbaumer (2003) utilizes matched assignment procedures in the investigation of discrimination against lesbians and found that heterosexual females receive 123 per cent more responses for applications than gay females in the Viennese labour market, Austria. Drydakis (2009) finds gay men to be less likely than heterosexual men to receive an invitation to job interviews in the Greek labour market by 23 per cent, while Drydakis (2011) detects a 31 per cent call back advantage for heterosexual females in Athens, Greece.

Experiments in the Swedish labour market were undertaken by Ahmed, *et al.* (2011b). They conducted the first nationwide field experiment in ten different types of occupations and find that homosexuals are discriminated against in the hiring process. Gay men are 14 per cent less likely than heterosexual men, and gay females are 22 per cent less likely than heterosexual females to obtain positive responses to their job applications. Discrimination was found in the private but not in the public sector and gay men were mainly discriminated against in typical male- dominated occupations, while lesbians in typical female- dominated occupations. Since the level of discrimination is, in comparison to other European countries, relatively low in Sweden (as discussed in Ahmed *et al*. 2011b), other explanations also become relevant for the findings such as Becker's (1981) theory of specialization within the family, as indicated by the findings in Jepsen and Jepsen (2006), Grossbard and Jepsen (2008) for the US and Ahmed *et al.* (2011a) for Sweden. It has to be mentioned that although the findings in other European countries than Sweden are statistically significant, they suffer in their representativeness also from the limitation to single cities.

Previous research found discrimination against homosexuals and different labour market outcomes by sexual orientation. Implications of discriminating treatment against homosexuals in the labour market as well as other markets were validated for Sweden as well. Wage differentials and discrimination against homosexuals in the hiring process were substantiated. However, the question of how the detected social phenomena might influence the duration in unemployment and employment probability of homosexual individuals in Sweden, haven´t been investigated before. The present work contributes to the labour market literature in this sense. It examines in particular the differences in the probability of being employed as well as the differences in the duration of unemployment by sexual orientation, by conducting survival analysis techniques. Survival data with indicators of sexual orientation pertaining to the labour market are rarely obtainable and mostly existent in Scandinavian countries. Therefore, previous labour market literature utilizing this approach is not existent.

3 THEORETICAL FRAMEWORK

According to standard search and matching models in labour market economics, the process of search for a desired occupation and thus, the duration in unemployment is costly.[4] The cost consists of the foregone potential income while searching, less the unemployment compensation or other transfer payments received, such as financial aid from family members. The lower the potential income from the targeted occupation, the lower the search cost, the

[4] Among other fundamental papers, Ehrenberg and Oaxaca (1976) discuss explicitly the implications of unemployment insurance and supplementary benefits from nongovernmental sources on the duration in unemployment via its (search) cost reducing effect and the correlation with the expected postemployment wages.

longer the search period, thus the duration in unemployment. The search period would be prolonged if *ceteris paribus* an unemployed individual becomes rejected due to his sexual preference, while otherwise be employed. If that were the case, search periods, i.e. durations in unemployment, would be longer for homosexual individuals than for heterosexual if discrimination against homosexual individuals exists in the labour market.

Two main theoretical explanations have been put forward in the literature on labour market discrimination; Becker´s (1957) taste based discrimination approach and the statistical discrimination hypothesis by Phelps (1972). A rather new approach in the labour economics literature is the implicit discrimination hypothesis discussed in Bertrand et al. (2005).

According to Becker (1957) employers could have distaste for employing people with particular sexual preferences and could therefore treat those people disadvantageous in the hiring or working process. In order to employ a member of this group (if at all) employers expect a reward, embodied in a lower wage paid to that group member. The size of the reward corresponds with the degree of distaste. This logic can be extended to employees, who dislike working with members of that particular group and therefore expect a compensating wage surplus; or customers, who expect a compensating lower price for buying products or services from members of that group. An employer, who is interested in the harmony within the firm, could have an incentive to reject the homosexual applicant due to the prejudice existent among the staff.

The second approach is based on information asymmetry and the inability of complete signalling the productivity of employees. Phelps' (1972) statistical discrimination hypothesis assumes profit-maximizing action of employers, based on uncertainty. The employer tries to circumvent this uncertainty about the employee in introducing suppositional group characteristics in the hiring and payment decision. The personal, measurable productivity will be accompanied by assumed group characteristics. If these characteristics signal lower productivity to the employer than the actual productivity is, the employee will suffer from lower wages or longer job search periods, since rejection occurs more often.

Both approaches above presume conscious actions of individuals in decision-making, but this might actually not always be the case. The implicit discrimination approach explains latent discriminatory behaviour, a product of unconscious action. As discussed in Bertrand *et al*. (2005) individuals may act discriminating, without awareness of that action. This unintentional behavior may occur due to time pressure, stress and distracting environment, as it is the case in managerial jobs, where automaticity can take the place of intentional action (Chugh, 2004). The empirical evidence for this occurrence in the hiring process was provided by Rooth (2010). Use was made of a computer based *Implicit Association Test* (IAT) and a *correspondence test*. It was shown that implicit discrimination could indeed be a determinant in the hiring process.

The above-mentioned theories explain the outcome of discrimination on the labour market, but not how the potential employer knows about the sexual preference of the applicant, in order to act discriminating. An answer could be the intended signalling of the sexual preference by the homosexual applicant. Arguments for why an individual could prefer the risk of a disadvantageous treatment if the applicant is expected to act rational in the application process, become vital in this case. One incentive for openly signalling the sexual preference is the avoidance of psychological pressure, stress and fear of being uncovered while hiding the preference (Ragins *et al.,* 2007). A second stimulus for exposing ones sexual preference is self-esteem, the behaviour could result from. This argument is quite reasonable for the individuals in our data set, since the individuals in it live in registered partnerships. Thus, the

sexual preference is exposed towards governmental authorities. When an individual chooses a registered partnership, one could argue that this is not only for legal and financial advantages[5], but also a product of self-confidence and -consciousness. 'Outing' is mostly not a behaviour of practical usefulness, but especially an ideological, political expression of oneself against conservative ideas towards plural lifestyles. This in turn could be correlated with such advantageous characteristic qualities, that the employer prefers those employees rather than avoiding them, since self-esteem is a productive determinant in the working process (Pierce & Gardner, 2004). In contrast to the expected results from discrimination theories, homosexual applicants who are 'brave enough' to expose themselves, could have shorter durations in unemployment, i.e. shorter search periods, since they are preferred due to the surpassing self- esteem and might be due to that even better searchers if we assume higher self-esteem to be correlated with more engagement into the job search (Ellis & Taylor, 1983 and Kanfer *et al.*, 2001).

All reasoning above requires the observance of sexual preference, either unintended or explicitly, in order to derive conclusions for homosexual individuals in the search process. Since we are unable to verify, nor reject any of these reasoning, we have to ask whether differing outcomes in the labour market could appear due to sexual orientation, even when the employer does not observe the sexual preference. A theory based on self- selection due to assumed discrimination is presented by Ahmed (2008). As opposed to former approaches, Ahmed (2008) assumes non- discriminating employers and shifts the cause for different outcomes on the labour market by particular characteristics to the applicant. The model is based on self- categorization and leads to self- conforming outcomes, which leads to a scenario that could be summarized as 'self- fulfilling' prophecy (Merton, 1948). Two types of workers are assumed, those who are thought to be discriminated against by the employer, i.e. type b, and those who are not, type g. All employers act profit- maximizing and do not prefer one applicant to another. There are high paying job offers and low paying job offers, while the productivity is constant among the applicants. Type b workers do not prefer applying to high paying jobs, where they believe to be discriminated against, in order to avoid frustration. Type g workers are aware of this behavior and do not fear competition by type b workers when applying to higher paying jobs and also do not prefer applying to lower paying jobs, since potential competition is more severe there, i.e. against type b and potentially against type g workers, and income is low. Since employers observe this behaviour, more employers will prefer lowering the wage, in order to attract workers of the expected constant productivity for lower costs. Thus, more job offers are present at a lower paying rate than at a higher.

Two rather surprising findings can be derived from this model. The first is that type b workers, i.e. homosexual individuals in our case, are expected to earn less on average, since they select themselves into lower paying jobs and thus a systematic difference in labour market outcomes appears by sexual preference. Since this lower income is expected by the applicant, according to basic search theory the opportunity cost of staying in unemployment is lower for homosexual individuals than for heterosexual and thus, *ceteris paribus* the duration of unemployment will be longer for homosexual individuals than for heterosexual, since the search is less costly for them. The second finding is counteracting to the first, i.e. since lower paying jobs are more present than higher paying, these are faster to get in and less time is required to search for them, thus, homosexual individuals have shorter search periods, i.e. shorter durations in unemployment.

[5] Sweden introduced the right of registered partnership for homosexual couples in 1994, which includes almost all rights and obligations married heterosexual couples have. These rights were extended to the adoption of children in 2002. (For a more detailed discussion and the effects of these legal actions on the labour market situation of same-sex couples see *Alden et al.* (2015).)

Another consideration suggesting similar conclusions, without the assumption of homophobic behavior, is based on income inequalities by gender that leads to differences by sexual orientation. As research shows, men still earn more than women in the same occupation at the same level of education in Europe (Eurostat, 2017). Thus, a family consisting of two workingwomen should *ceteris paribus* have less disposable income, than a family consisting of a woman and a man or even of two men, both individuals working. Accordingly, we could expect that an unemployed woman cohabiting with another woman should have less financial support opportunities from the partner, than a woman living with a man. On the other hand, a man coupled with a woman should have less financial support by the working partner than a homosexual male. The higher the income of the partner, the higher the actual household earning when unemployed and the lower the individual cost of search and thus, the longer the duration in unemployment. This leads to the following conclusion: Coupled lesbian individuals have shorter durations in unemployment than heterosexual married women, and gay men have longer durations in unemployment than heterosexual men. The reasoning regarding the durations in unemployment can be extended to the probability of being employed if longer durations in unemployment are assumed to be associated with lower likelihoods of being employed. *Ceteris paribus,* the longer an individual needs for the job search, the longer he stays out of the working process and is not employed, thus, the higher the probability of unemployment. Correspondingly, if homosexual individuals are more rejected in the application process than heterosexual, *ceteris paribus* homosexual individuals will have lower likelihoods of employment. A third occurrence is the fusion of both phenomena, i.e. even though homosexual individuals might not differ from heterosexual by search durations or employment probabilities on its own, they might have disadvantages when the occurrence of employment is put into relation to the duration of unemployment. Hence, due to discrimination or other reasons mentioned above, homosexual individuals might have less employment incidences in a particular search interval or might have the same employment incidences but within a longer search period.

Following Rooth (2010) and in order to sum up the discussion, we can conclude that the probability of getting a job for individual *i* from the group of homosexual applicants *j* can be formalized as follows:

$$\Pr\left(Employed = 1\right)_{ij} = X\beta_j + \delta_j^A + \delta_j^{EP} + \delta_j^{ES} + \delta_j^I \; ; \; j = Gay$$

where X is a vector of human capital characteristics and β_j is the return to those characteristics. δ_j^A measures the assumed level of discrimination by the applicant, $\delta_j^{EP}$is a measure of degree for the explicit preference discrimination (taste-based discrimination) against homosexuals, $\delta_j^{ES}$is a measure for explicit statistical discrimination and δ_j^I measures the degree of implicit discrimination. Hence, the probability that a homosexual unemployed becomes employed depends, beside personal characteristics, on the personally assumed level of discrimination and the different degrees of 'external' discrimination.

Similarly, the duration in unemployment t for individual *i* from the group of homosexual applicants *j* depends on the human capital characteristics comprised in vector X, negatively on the opportunity cost of search *OC*, where β_{1j} and β_{2j} are the returns to human capital characteristics and to the opportunity cost of search respectively, additionally on the personally

assumed degree of discrimination δ_j^A and the different types of discriminatory treatment, δ_j^{EP}, $\delta_j^{ES}$and δ_j^I which we formalize as follows:

$$t_{ij} = X\beta_{1j} + OC\beta_{2j} + \delta_j^A + \delta_j^{EP} + \delta_j^{ES} + \delta_j^I \; ; \; j = Gay$$

We won't be able to test for and distinguish between the particular discriminating forces in the Swedish society during this paper, but we can assume that the discriminating attitudes against homosexual females and males consist of a mixture of the above-mentioned behaviours.

Hypotheses

Based on the discussion above and contrary to previous findings in the literature, the following hypotheses are formulated for married individuals or those living in civil unions, for men and women separately. *Ceteris paribus*, we expect that,

1. Homosexual men have longer durations in unemployment.
2. Homosexual women have shorter durations in unemployment.
3. Homosexual men have a lower employment probability.
4. Homosexual women have a higher employment probability.
5. Homosexual men have a lower incidence rate of employment within a particular interval, i.e. *Hazard Ratio* smaller than one by sexual orientation.
6. Homosexual women have a higher incidence rate of employment within a particular interval, i.e. *Hazard Ratio* larger than one by sexual orientation.

4 METHODICAL FRAMEWORK AND RESEARCH DESIGN

The phenomena to be explained are the duration in unemployment and the event of getting employed. Thus, the research design is clustered around three stages, i.e. the modelling of a continuous dependent variable, a categorical dependent one and the fusion of both. We start with the duration in unemployment and use a *Zero- Inflated Negative Binomial* (*ZINB*) model.[6] Whether days in unemployment differ by sexual orientation is the main aspect of our interest. While time is assumed to be of continuous nature, in our data we observe how many days of unemployment an individual had, which is constrained by zero and three hundred and sixty-five, and the increment of one day. Correspondingly, we can count the number of individuals a particular day in unemployment had been assigned to. As obtainable from the descriptive statistics, our dependent variable follows a negative binomial distribution and the standard deviation is much larger than the mean, which makes the usage of *Ordinary Least Squares* not quite meaningful.[7] We use a count model, which incorporates the overrepresentation of zero- counts in our data set and also accounts for the particular distribution of our

[6] Further details of the model selection process are provided in the appendix section 2.
[7] Using *OLS* in the presence of overdispersion can cause the estimates to be inefficient, inconsistent and biased (Long and Freese, 2006, p. 349).

data. Without going into full depth, following an unobserved latent variable, the individuals in our data are divided into two categories, i.e. those whose probability of getting zero days of unemployment is equal to one $(A_i = 0)$ and those who have a probability of getting unemployed according to their individual characteristics $(A_i = 1)$, which is determined by negative binomial regression. The probability generating process is separated for both groups and mixed by the fraction of the group affiliation. Accordingly, the probability of having particular days in unemployment is expressed as follows:

$$\Pr(y_i|\boldsymbol{x}_i, A_i = 1) = \frac{\Gamma(y_i+\alpha^{-1})}{y_i!\,\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu_i}\right)^{\alpha^{-1}}\left(\frac{\mu_i}{\alpha^{-1}+\mu_i}\right)^{y_i}$$

where the probability Pr(·) of days in unemployment y_i for individual i given the individual characteristics comprised in vector $\boldsymbol{x}_i$ and while being in subgroup A , is dependent on the degree of dispersion α, days in unemployment y_i , individual mean μ_i and the gamma function with the corresponding gamma distribution $\Gamma(\cdot)$. μ_i , which is also called the observed heterogeneity effect (Long and Freese, 2006, p. 356), comprises the individual characteristics of the individual:

$$\mu_i = \exp\left(\alpha_i + \beta_i \boldsymbol{X}_i + \lambda_i \boldsymbol{D}_i + \vartheta_i \boldsymbol{I}_i + \varphi_i \boldsymbol{O}_i + \theta_i Homosexual_i + \varepsilon_i\right)$$

where human capital variables such as age and educational attainment are comprised in vector $\boldsymbol{X}_i$, socio- geographic variables such as region of residence, children in the household and immigrant background are depicted in vector $\boldsymbol{D}_i$, the different types of income and occupations are comprised in the vectors $\boldsymbol{I}_i$ and $\boldsymbol{O}_i$ respectively, and the sexual orientation is comprised in the dummy variable $Homosexual_i$ with heterosexual as the reference group. $Homosexual_i$ equals one if an individual is homosexual. α_i is the intercept term and β_i, λ_i, ϑ_i, φ_i and θ_i are the returns to the corresponding vectors, while ε_i is the error term. By incorporating $Homosexual_i$ we are able to see how much the duration in unemployment is *ceteris paribus* prolonged by sexual orientation. We include age^2 due to the non-linear effect of ageing on the duration in unemployment. As research on ageism confirms, becoming older has a negative impact on employability after a certain age (Anxo *et al*., 2017, p.50). The older an applicant is, the higher the likelihood of being employed, but at a decreasing rate. Thus, *ceteris paribus* we can expect that the duration in unemployment will be shortened by age, but at a decreasing rate, which is incorporated by the non-linear effect of the variable age^2. The different types of income comprised in $\boldsymbol{I}_i$ such as disposable family income, unemployment compensation, as well as social contributions to families, are expected to prolong the unemployment duration, since the search becomes less costly by these amounts. Income from own work, before unemployment or after, is expected to decrease the duration, since it signals the possible foregone earnings, i.e. the opportunity cost of search. Different types of occupation are expected to have different returns to work. The incomes from blue-collar work might differ from white-collar work; the agricultural sector might categorically differ in terms of income from those predominantly present in metropolitan regions. Thus, the presumed income while targeting a particular occupation in the search process influences the

opportunity cost of search. Therefore the vector $\boldsymbol{O}_i$ incorporates dummy variables that indicate whether an individual is in the agricultural, manufacturing or service sector, as well as whether the income is caused by construction, administration or self-employment.[8]

In our second step we model the dichotomous dependent variable, i.e. the event of becoming employed. We utilize *Probit* estimation and obtain marginal effects at the average levels of the explanatory variables and the effects of discrete changes for categorical covariates. We are interested in the effect of homosexuality on the probability of employment and will obtain the discrete change in the probability of employment for homosexual individuals, when all other variables are held at their mean values. In this way we will see how sexual orientation influences the 'average individual' in the data. Additional to the parameterization in the *Zero-Inflated Negative Binomial* estimation (ZINB), we include days in unemployment as an explanatory variable, but delete the occupation dummies, since perfect prediction in the measurement process makes an estimation for the vector $\boldsymbol{O}_i$ not possible, while theoretically still influential, since the probability of becoming employed is expected to differ by occupational category, its relative scarcity and the differing requirements to get into a particular job. As theory suggests, days in unemployment are costly. This might have an incentive effect for the job-searching individual, since the faster one gets into a job, the lower the cost due to search activity. Thus, the probability of employment might increase by passing days. But the more time passes, the more the discouraging effect of unsuccessful search might become influential. Hence, a positive effect at a decreasing rate can be expected for unemployment days and thus, it is included as an explanatory vector in its linear and squared form, interacted with the sexual preference. As stated in the theoretical discussion, the own opportunity cost of staying unemployed for coupled individuals also depends on the income of the working spouse. Thus, it is expectable that the effect of the variable might be different for those individuals with high earning partners than for low earning, in our case for those individuals who are coupled with women, since women are assumed to earn systematically less than men. The following model is estimated

$$\Pr(Employed = 1)_i = \alpha_i + \beta_i \boldsymbol{X}_i + \lambda_i \boldsymbol{D}_i + \vartheta_i \boldsymbol{I}_i + \theta_i Homosexual_i + \psi_i HUDays_i + \varepsilon_i$$

where the probability of being employed for individual i, i.e. $\Pr(Employed = 1)_i$, depends on the same explanatory variables as in the *ZINB* with '*Homo* x *Days in Unemployment*' as an additional vector, i.e. $HUDays_i$, which also incorporates its squared term, while ψ_i is the corresponding coefficient. The variable income is not included in this approach due to perfect prediction.

In our third and last step we use survival analysis techniques, which enable us to estimate the changing probabilities of becoming employed over time, i.e. not a parameter estimate anymore, but the estimation of a function. The additional gain of these methods lies in its property to fuse time and event into one analytical feature (Therneau & Grambsch, 2001). With the first two methods we obtain point estimates that is an overall snapshot of days in unemployment and the probability of employment for the whole data set and the whole year. Survival analysis techniques enable us to observe the differences in the probability of becoming employed within the year 2003 and whether the probability of becoming employed is significantly different in some intervals of the year and not in others, by sexual orientation. In

[8] The full set of variables and the according explanations are obtainable from Appendix 1.

order to gain a first glimpse of the time-to-event data, we use unadjusted Kaplan-Meier survivor curves and 'let the data talk for itself' (Cleves et al., 2016). This procedure serves as a descriptive measure and provides us with the unemployment experience by sexual orientation, when neither assumptions about the effects of the covariates nor the distribution of failure times are made. Since covariates of our interest are present, we utilize a parametric baseline hazard with Gompertz specification and test against semiparametric and other parametric approaches, which is presented in detail in Appendix 3. When an appropriate distributional assumption is made, in our case Gompertz, parametric models yield more efficient estimates than semi-parametric and incorporate censoring of data into the analysis, while nonparametric techniques are unable to do so (Cleves *et al*., 2016, p. 6, p.234). The following proportional hazards model is estimated for different parameterizations:

$$h(t|\boldsymbol{x}_j) = \exp(\gamma t)\exp\left(\beta_0 + \beta_i \boldsymbol{X}_i + \lambda_i \boldsymbol{D}_i + \vartheta_i \boldsymbol{I}_i + \varphi_i \boldsymbol{O}_i + \theta_i Homosexual_i + \varepsilon_i\right)$$

where the hazard $h(\cdot)$, the incidence of becoming employed within a particular time span, is determined by time *t* and the individual characteristics of individual *j*, comprised in vector $\boldsymbol{x}_j$. $\exp(\gamma t)\exp(\beta_0)$ is the baseline hazard of the Gompertz model, where γ is the shape parameter of the baseline hazard and β_0 the intercept term, estimated together with the covariates comprised in the vectors of exp(·), which follows the parameterization in the *ZINB* model.

5 Data and Descriptive Statistics

The data used comes from the LOUISE database (Longitudinal Integration Database for Health Insurance and Labour Market Studies) provided by Statistics Sweden. Our data contains information about demographic characteristics, labour market characteristics and other socio- economic variables of all homosexual individuals aged 25- 64, who were living in civil unions in Sweden by the year 2003. A group of randomly selected married heterosexual individuals in the same age interval was included for comparison. 5370 individuals are incorporated in the analysis, among whom 2661 are heterosexual and 2627 homosexual. 484 individuals were registered unemployed for at least one day in 2003, among which 211 individuals were heterosexual and 273 homosexual. 111 individuals were censored, i.e. had positive durations in unemployment but no observable event of employment, thus, right- censoring of the data is present and the exact time in unemployment is unknown. Left- truncation has to be assumed in our data set, since only a minority of individuals might have started to face unemployment with the beginning of the year 2003. These limitations are generally existent in social studies drawing the survival data from non-experimental trials and even in those. We assume the truncation phenomenon equally existent for all four subgroups and assume valid conclusions for the comparison of these. The limitation of the data to coupled and registered individuals is vital since every conclusion made can only be derived for a selected subgroup of homosexual and heterosexual individuals in the Swedish society.

Table 1: Descriptive Statistics for Homo- and Heterosexual Individuals (25- 64 Years of Age) by the Year 2003

	Males		*Females*	
	Heterosexual	*Homosexual*	*Heterosexual*	*Homosexual*
Age (years)	48.93	45.10	47.25	40.97
Schooling (years)	11.76	12.70	11.90	12.97
Children (%)	46.89	0.51	47.59	.23.35
Immigrant background (%)	17.93	26.13	16.72	16.67
Metropolitan Area (%)	56.41	83.56	51.13	73.16
Northern Region (%)	8.25	2.62	11.67	6.40
Days in Unemployment	11.09	10.99	8.29	13.90
Std.D.Days in Unemployment	48.47	45.56	41.63	51.92
Employed (%)	89.42	84.22	82.08	87.10
Income*	2717.03	2311.92	1593.54	1886.51
Disposable Family Income*	4099.06	3785.77	4427.55	3301.13
Unemployment Insurance *	40.40	36.25	37.09	51.52
Social Contributions *	21.51	6.34	24.39	15.94
Agricultural Sector (%)	2.25	0.26	0.68	0.75
Manufacturing Sector (%)	22.88	5.81	7.30	8.10
Construction Sector (%)	8.25	0.38	0.53	1.32
Service Sector (%)	38.11	41.60	24.17	31.17
Healthcare Sector (%)	5.78	0.20	24.62	20.43
Public Administration (%)	5.10	5.18	5.42	8.76
Self-Employed (%)	11.70	6.26	4.47	6.21
Number of Individuals	1,333	1,565	1,328	1,062

* Displayed in hundreds of Swedish Kronor (SEK)

Table 1 presents the descriptive statistics regarding socio- demographic characteristics as well as labour market and income indicators of the individuals in our investigation. Table 2 presents descriptive measures of survival data. Homosexual individuals in civil unions are on average younger than married heterosexual individuals and women are younger than men within both groups of sexual orientation. An important feature that has to be taken care of within the analysis is that homosexual individuals have on average higher educational attainment than heterosexual individuals, which is considered as a confounding measure in our analysis. Another such confounding important occurrence is the non- equal spread of homosexual individuals in the geographical areas of Sweden. While 56 per cent of male and 51 per cent of female heterosexuals live in metropolitan areas as the counties of Stockholm, Gotaland, Skane and Vastra, 84 per cent of the homosexual males and 73 per cent of the homosexual females live within these areas. As discussed in Black et al. (2007) and Ahmed and Hammarstedt (2009), this dissimilar distribution of homosexuals to metropolitan regions could be due to more liberal attitudes towards homosexuals within these areas compared to rural ones and hence, regional fixed effects have to be taken into account in the analysis, (Clain and Leppel, 2001), (Arabsheibani et al., 2004). For our purposes, it is important to consider that these regional differences in attitudes may have differing impacts on the duration in unemployment as well as the probability of getting employed, that is, shorter durations in unemployment in metropolitan areas due to more employers with liberal, non- discriminating attitudes towards homosexuals or higher employment probabilities.

Another considerable difference is the immigrant share among males. While 26 per cent of homosexual males have an immigrant background, this share is with 18 per cent eight points lower for heterosexual. The share for females is 16 per cent for both groups. Since immigrants have higher unemployment rates in Sweden and are discriminated against in the hiring process (Carlsson and Rooth, 2012), this characteristic might have a strong confounding effect on the duration in unemployment and the likelihood of employment. 48 per cent of the heterosexual women and 47 per cent of the married heterosexual men take care of children in their households. Almost one fourth of the homosexual females, but only 0.5 per cent of the homosexual males take care of children. Following Becker (1981) children in the household could have incentive effects in the job search process for men and disincentive effects for women. Since specialization within the family could exist for homosexual couples as well (Ahmed *et al.*,2011a), absence of children in the homosexual male household could lower the feeling of acuteness in the search process and thus, could prolong the search, since a small tummy to feed is missing.

While the average days in unemployment show with approximately eleven days no difference by sexual orientation for males, this picture changes with an interesting magnitude for females. Homosexual females have with 14 days seventy-five per cent longer durations in unemployment than heterosexual with 8 days. To the contrary, lesbian individuals earn 18 per cent more than heterosexual women and heterosexual men earn 18 per cent more than gays, when the lower value is treated as benchmark. In line with this pattern is the employment fraction. 89 per cent of heterosexual men were employed for at least one day, compared to 84 per cent for homosexual men. Homosexual women had with 87 per cent a higher ratio than homosexual men and heterosexual women with 82 per cent. Disposable family income, which is the combined income of the partners, reaches with 442755 SEK the highest value of all subgroups for heterosexual married women and is therefore 34 per cent higher than the income of homosexual coupled women, which is also the lowest income of all four categories. This difference is with eight per cent in favour of heterosexual men much lower for males. As discussed in the theoretical part, income on the household level is one of the key- determinants of the search length. While the individual wage indicates the foregone earning and is thus theoretically associated with higher opportunity costs of search the higher it is; higher total household incomes are assumed to have a compensating effect on the unemployed partner.

The different sectors show a particular allocation pattern to the occupational categories. Lesbian women show a higher share in those occupations, where also heterosexual men are predominantly present and a lower, where gays are more present than heterosexual men, except for the service sector and public administration. This could be interpreted as the reflection of a particular differing characteristic lesbian have. They might be less reluctant to work in a male dominated environment and more robust to heterosexual *machismo*. This is particularly assumable in the construction sector, where the employment rate is 250 per cent higher for lesbians, compared to heterosexual females. The same dedication might also show up in the share of *self-made women,* since lesbian are with 6.21 per cent compared to 4.47 almost fifty per cent more present when starting an own business.

Table 2: Descriptive Measures of Survival Experience for Individuals (25- 64 Years of Age) who were Unemployed for at Least one Day in 2003

	Males		*Females*	
	Heterosexual	*Homosexual*	*Heterosexual*	*Homosexual*
Average unemployment (days)	138.1	118.6	105.9	115.4
Median unemployment (days)	138	112	94	106
25th percentile	48	43	44	54
75th percentile	243	240	226	191 198
Average hazard rate	0.0055	0.0067	0.0064	0.0071
Average hazard rate ratios	0.82	1.22	0.90	1.11
Number of Individuals	107	145	104	128
Censored Individuals	25	29	34	23

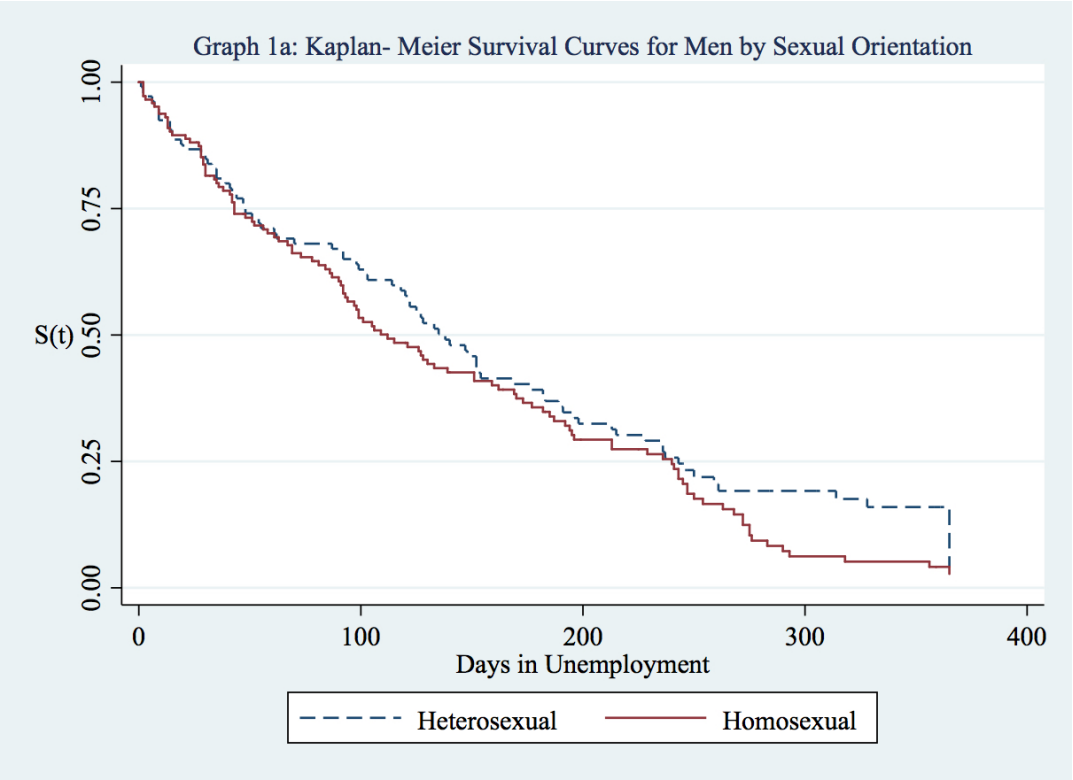
Table 2 shows that heterosexual males have with 138 days the highest overall average of unemployment days in the group of unemployed and heterosexual females the lowest with 106 days. Homosexual men have lower average unemployment days than heterosexual men and homosexual women higher average unemployment days than heterosexual women do, when only individuals are taken into account, who were unemployed for at least one day in 2003. Conclusions based on these descriptives have to be derived with caution, since no adjustments for selective measures were made, but the median unemployment duration points into the same direction. The time at which the probability of remaining in unemployment becomes fifty per cent is with 94 days the shortest for heterosexual women and with 138 days the longest for heterosexual men. This indicates that half of the unemployed heterosexual females are in job after 94 days, while it takes 138 days for heterosexual males. Homosexual men have with 112 days an advantage of almost one fifth compared to heterosexual males, and compared to heterosexual females it takes almost 13 per cent more time until fifty per cent of the unemployed homosexual females are employed.

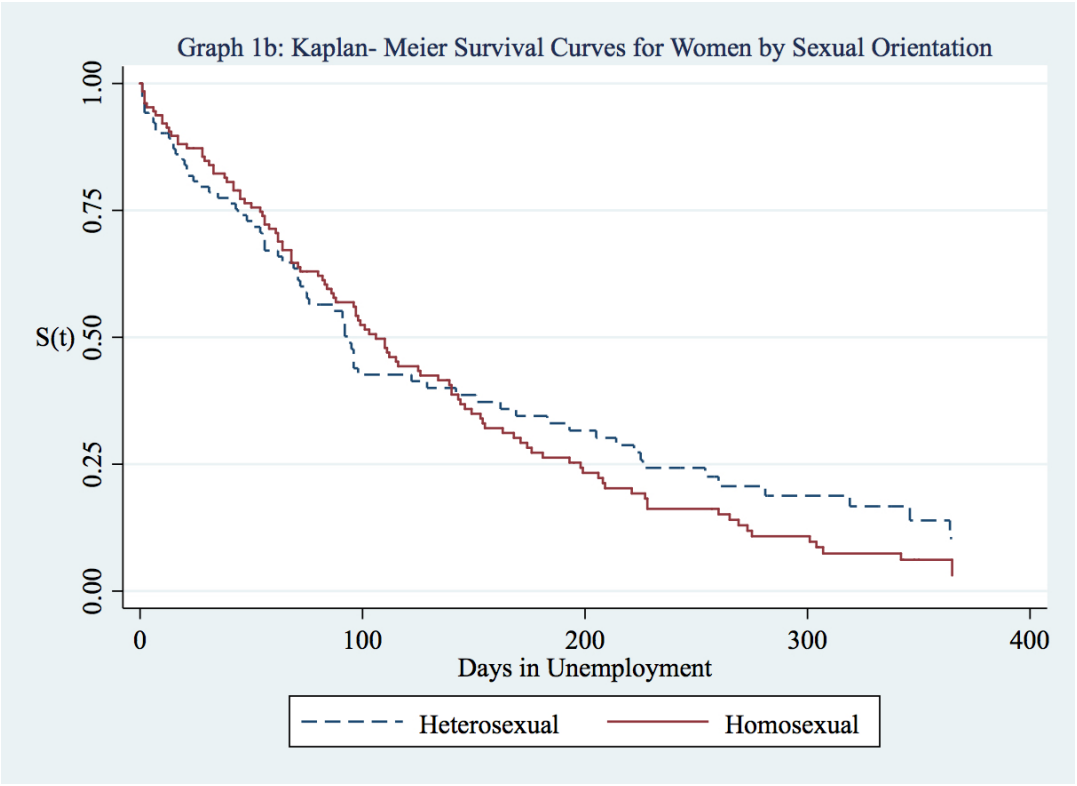
This picture changes, when the 25th percentile and 75th percentile are also considered. While for males the pattern is similar for every percentile, the pattern for females is ambiguous. The duration until a particular percentile is longer for heterosexual males than for gays. The duration until 25 per cent and 50 per cent of the homosexual females got employed is with 54 and 106 days longer than it takes for heterosexual females with 44 and 94 days respectively. When the 75th percentile is reached, 198 days were required for homosexual females and 226 days for homosexual females. Hence, a shift took place. This shift explains the outcome of the third descriptive measure, the average hazard rate, also called the average incidence rate that is defined as the quotient of the total number of events in a group and the sum of the observed survival times. Hence, “How many events of employment occur in which overall time span of unemployment?” The more employments happen in a time interval or the shorter the duration in unemployment in which a particular amount of events of employment happen, the higher the average hazard rate, the better the performance of the group. Homosexual females have with 0.0071 the highest rate of incidences whereas heterosexual males the lowest rate with 0.0055. Homosexual males have with 0.0067 an advantage of twenty-two per cent compared to heterosexual males, whereas heterosexual females have with 0.0064 a disadvantage of eleven per cent in relation to homosexual females.

Graphs 1a and 1b provide graphical representations of the unemployment experience. Kaplan- Meier curves do not assume any functional form and depict the changing employment probabilities over time without adjustment by any covariates. Thus, it allows the “data

talk for itself" and hence, can be used as a descriptive measure (Cleves *et al*., 2016). Graph 1a shows that the survivor curves for men are entangled into each other on an erratic path in the first quarter, diverge in the second, narrow in the third quarter and drift apart in the fourth. The survivor curve for homosexuals lies after the first quarter consistently below the curve for heterosexuals, thus at any point in time of the follow- up period homosexual males have a lower probability of being unemployed than heterosexual males, when no adjustment is undertaken. The probabilities of being employed drift most apart towards the end of the observation, where an abrupt employment performance takes place between the 250th and 290th days for homosexual men but not for heterosexual. The survivor curves for women are almost identical approximately up to day 145, where the curve for homosexual women crosses the curve for heterosexuals and lays below the one for heterosexuals for the rest of the observational time. The widening gap between the curves indicates that the employment performance for homosexual women is higher after an equal path in the first time.

Graphs 2a and 2b show histograms of days in unemployment for our data. The histogram in 2a clearly shows the overrepresentation of zero counts in our data set, which justifies the choice of a zero- inflated model. Graph 2b excludes the zero counts and shows the negative binomial distribution of positive days in unemployment. Beside the statistical tests for model selection provided in Appendix 2, the observable distribution leads us to the usage of *Zero-Inflated Negative Binomial* regression.




Graph 1b: Kaplan- Meier Survival Curves for Women by Sexual Orientation
S(t)
0.00
0.25
0.50
0.75
1.00
0
100
200
300
400
Days in Unemployment
Heterosexual
Homosexual


Graph 2a: Histogram of Days in Unemployment with Zero Counts
Density
0
.01
.02
.03
.04
.05
0
100
200
300
400
Days in Unemployment

6 Results

6.1 Results of Zero- Inflated Negative Binomial Regression for Days in Unemployment

The results from the estimation of the five different specifications for the Zero- Inflated Negative Binomial regression are presented in table 3a and 3b for men and women separately. While the upper section shows the results for the negative binomial regression for those individuals who were unemployed for at least one day, the lower section presents the findings for the logistic regression for the *certain zero* group, i.e. the probability of being employed for the whole observational time. The coefficient estimates are presented in exponentiated format, in order to make the interpretation more intuitive.

As both table show, the expected numbers of days in unemployment do not vary significantly by sexual orientation, for men and women equally. Irrespective of the statistical significance, the estimated values show an increasing prolongation of days in unemployment the more vectors are controlled for and amounts to a factor value of 1.09 for men and 1.16 for women. As obtainable from the lower section of table 3b, being homosexual does significantly decrease the probability of being in the *certain zero days of unemployment* group for women and commutes in the odds ratio (not the log of odds) of 0.67, while the significance level shrinks to ten per cent. Although the probability of employment will be discussed in detail in section *6.2*, this finding provides a hint on the drawback homosexual individuals face. While educational attainment has a marginally significant impact at the ten per cent level for men in specification five, migration background has a significant impact on days in unemployment for men but not for women when all vectors are controlled for. The expected days in unemployment increase *ceteris paribus* by the factor 1.3 when a man has an immigrant back-

ground, i.e. the prolongation of expected unemployment days by thirty per cent compared to non- migrant individuals when an individual is in the not always employed group. Correspondingly, the probability of falling in the certainly employed group is significantly lower for men and women with migration background in all three specifications where migration background is controlled for. The odds of falling in this category is lower by a factor of 0.6 for men and with 0.54 even lower for women in specification five.

Theory suggests that own income from work, prior unemployment or after, has a shortening effect on the duration in unemployment, as opposed to the extending effect of external transfer payments and social contributions while being unemployed. While the estimate for income is statistically significant at the five per cent level for males in specification five, its effect is with 0.99989 almost one and thus, does not change the duration in a multiplicative manner. However, payments from unemployment insurance as well as other social contributions have statistically significant impacts on days in unemployment in the theoretically expected direction for both sexes. The monetary values are measured in 100 Swedish Kronor (SEK), thus, an increase of insurance payments by 1000 SEK, approximately 100 Euros, increases the duration in unemployment for men by the factor 1.011814. The factor value for the other social contributions is 1.008012. The additive effect of both variables, when both payments increase by 1000 SEK, is approximately two per cent. The same effect is observable in specification five for women with 1.014 for unemployment insurance payments and 1.009 for other social contributions. The duration in unemployment prolongs *ceteris paribus* for women by 1.4 per cent for every 1000 SEK increase received from unemployment insurance, irrespective of the specification. As expected, income and disposable family income both increase the probability of being in the certainly employed group, while unemployment insurance and social contributions decrease the probability. While some of the values in this vector are not statistically significant for women but for men and vice versa, all estimated values and thus the odds are approximately one and thus, the probability of being in the certain employed group does not change practically by these categories.

Rather surprising are the findings for the occupational vector in specification five. Irrespective of the particular occupation and contrary to our theoretical consideration, there are no statistically significant impacts of the particular occupational choices neither on the duration in unemployment nor the probability of being in the *never unemployed* group for women. For men, it is only working in the service sector that increases the duration in unemployment by the factor 1.32, while the other occupations do not have a significant impact. If this single finding is interpreted from the perspective of search theory, one can conclude that the prolonging effect of working in the service sector is anticipated, since service sector provides lower payments for men on average compared to the other sectors and thus aiming for this occupation lowers the opportunity cost of search. Although only the manufacturing sector is statistically significant at the five per cent level, also construction, service and healthcare occupations indicate a decreasing probability of being certainly employed, while public administration and being self- employed does increase the probability. While entering public administration has higher entry requirements in general and also provides a more secure working environment regarding the dismissal options, being self- employed is assumed to be a more risky option. Surprisingly, the most secure and most risky occupations both point to an increasing probability of being in the certainly employed group, more for women than for men, but all at a statistically insignificant level. Being once established in the self- employed occupation might be a *once and for all* option. Either one is successful as a self- employed and has a running business, or one is not

Table 3a: Exponentiated Coefficient Estimates from Zero- Inflated Negative Binomial Regression for Days in Unemployment, Men

	Specification				
	I	II	III	IV	V
Homosexual	.85791247	.89070505	.99667013	1.0466129	1.0851204
Age		1.0112435	1.0083603	1.0411646	1.0515645
Age²		.99994136	.99999592	.99954092	.99945419
Schooling		.99759515	.99912527	1.0418153	1.0430815*
Children			1.1549151	.94794735	.93796585
Immigrant			1.0356083	1.2202909*	1.2992207**
Metropolitan Area			.91001418	1.083697	1.0533695
Northern Region			.9751784	1.4032249	1.4507427
Income				.9999132*	.99989975**
Unemployment Ins.				1.0011725***	1.0011814***
Social Contributions				1.0007104**	1.0008012***
Disposable Family Inc.				.99995932	.99995652
Housing Allowance				1.00062	1.0003387
Agricultural Sector					1.7528603
Manufacturing Sector					1.2612928
Construction Sector					1.3691123
Service Sector					1.3226475**
Healthcare Sector					.92123092
Public Administration					1.4773477
Self-employed					.90414539
Constant	137.38429***	95.680989***	91.029577***	17.442613***	11.139593**
Inflated					
Homosexual	.85371332	.93932211	.94153531	1.0583755	1.0084648
Age		1.1030812*	1.0613424	.98642154	.97399294
Age²		.99935659	.99973724	1.0005951	1.0007118
Schooling		1.0610405**	1.0454687*	.93384319*	.9127476**
Children			1.0704253	1.2955858	1.4037961
Immigrant			.44078746***	.60329395**	.59361096**
Metropolitan Area			1.4343904**	1.0903342	1.0422388
Northern Region			.69576144	.58781093	.56424544
Income				1.0003714***	1.0004769***
Unemployment Ins.				.989476***	.98973596***
Social Contributions				.99840107***	.99823589***
Disposable Family Inc.				1.0001533**	1.0001605**
Housing Allowance				.99982985	.99986685
Agricultural Sector					.89680451
Manufacturing Sector					.39356026**
Construction Sector					.34287526*
Service Sector					.62529301*
Healthcare Sector					.6436001
Public Administration					1.1801533
Self-employed					1.395428
Constant	11.389775***	.24826148	.74643584	9.0940945	20.609416*
Statistics					
Number of Obs.	2898	2898	2898	2898	2898
Log- Likelihood	-2326.0383	-2304.3839	-2286.269	-1898.3287	-1889.7462
χ^2	1.5415701	2.5217727	3.4785429	83.900752	92.795326
BIC	4691.9354	4696.4573	4724.0018	4027.839	4122.2788

* Statistically significant at 10%; ** Statistically significant at 5% ; *** Statistically significant at 1%

Table 3b: Exponentiated Coefficient Estimates from *Zero- Inflated Negative Binomial* Regression for Days in Unemployment, Women

	Specification				
	I	II	III	IV	V
Homosexual	1.090731	1.1596124	1.0516547	1.1606919	1.1591323
Age		.94805351	.95702091	.97753057	.96890559
Age²		1.0007059	1.0005648	1.0001417	1.0002369
Schooling		1.0038488	.99602156	.96254212	.96408737
Children			.8792006	.81995555	.7993739
Immigrant			.90222733	1.1189862	1.0837174
Metropolitan Area			1.0918162	1.2775649*	1.2796245*
Northern Region			.76380307	1.0814913	1.0733428
Income				.99991662	.99991091
Unemployment Ins.				1.0013571***	1.0013917***
Social Contributions				1.0009359**	1.0009153**
Disposable Family Inc.				1.0000099	1.00001
Housing Allowance				1.0003054	1.0004517
Agricultural Sector					1.4304529
Manufacturing Sector					1.0528687
Construction Sector					.36185243
Service Sector					1.0769474
Healthcare Sector					1.078917
Public Administration					.60076876
Self-employed					.92062488
Constant	104.22201***	244.50579***	260.03915***	133.36253***	156.18351***
Inflated					
Homosexual	.62028944***	.75280353*	.66078835**	.66853483*	.6691906*
Age		1.0982201	1.1025223	.98658266	.96237114
Age²		.99951567	.99940912	1.0006583	1.0009091
Schooling		1.1020982***	1.0773248**	.98568642	.97394629
Children			.89660327	.86076726	.85580719
Immigrant			.48625564***	.53455397***	.54442825***
Metropolitan Area			1.2740205	1.0222265	1.0318847
Northern Region			1.0279049	1.1319318	1.1432021
Income				1.0005252***	1.000553***
Unemployment Ins.				.98983652***	.99029591***
Social Contributions				.99965075	.9997045
Disposable Family Inc.				1.0001349	1.0001384
Housing Allowance				.99591983**	.99553185**
Agricultural Sector					.59937913
Manufacturing Sector					.53849611
Construction Sector					1.2514223
Service Sector					.74645447
Healthcare Sector					1.0737335
Public Administration					2.3800864
Self-employed					2.1902006
Constant	11.568727***	.15214006	.23420588	6.3650294	12.958567
Statistics					
Number of Obs.	2390	2390	2390	2390	2390
Log- Likelihood	-2077.9505	-2047.7447	-2036.8231	-1714.7313	-1708.4106
χ^2	.37087773	2.0044067	4.4723287	85.490755	90.190802
BIC	4194.7963	4181.059	4221.4481	3655.055	3751.3203

* Statistically significant at 10%; ** Statistically significant at 5% ; *** Statistically significant at 1%

6.2 Results from Probit Estimation for the Probability of Becoming Employed

In table 4a and 4b we obtain marginal effects at the average values of the covariates after *probit* estimation of employment probabilities for males and females respectively. In all specifications the probability of becoming employed is *ceteris paribus* lower for homosexual males compared to heterosexual at the one per cent significance level and varies between minus five and minus eight percentage points depending on the particular specification. For females the probability of becoming employed is significantly higher by five percentage points when the individual is homosexual, and no other covariates are incorporated. This effect turns statistically insignificant and negative when other vectors are controlled for. Specification two indicates that differences between homo- and heterosexual women in the employment probability are highly driven by the differences in human capital characteristics. When these are incorporated, the probability shrinks by almost seven percentage points and lowers only by 1.5 points more after the adjustment for the other vectors. In specification five when days in unemployment are incorporated, the negative effect of being homosexual on employment probability turns statistically significant at the ten per cent level and increases by 0.7 percentage points in absolute value compared to specification four, where different kinds of income, socio-geographical and human capital variables are controlled for. The interaction term in specification five is positive and significant at the five per cent level, just as the squared term, which is negative. This concave relation indicates that the probability of employment changes positively for lesbians by passing days at a decreasing rate. That is, the longer the days in unemployment, the higher the probability of becoming employed. The squared term is too small in magnitude for being considered, but when the parameter estimates for *Homosexual* and the interaction term of *Days in Unemployment* and sexual orientation are put into an equation, a threshold of 23 days can be calculated, after which the probability of employment becomes positive for homosexual women. This again underlines the importance of passing time in unemployment for the probability of employment. The negative impact of being homosexual on the probability of becoming employed is with -6.8 percentage points more than twice as high in absolute value for men, than for women with -3 percentage points.

Regardless of the particular specification, the vector containing the human capital variables such as age and educational attainment has a statistically significant impact for both sexes at the one per cent level. As theoretically expected, the likelihood of becoming employed increases with age, which is at around three percentage points per year for males and four for females. The squared term of age has a negative sign for males and females, which indicates the expected concave impact of age on employability. The older an individual gets, the more employable he becomes, but at a decreasing rate. When the first derivative of the probability of employment is taken with respect to age, we find out that 87 and 84 years of age are the thresholds for men and women respectively, after which age has a negative impact on employability. These findings are significant in a statistical sense, but not much in a practical, since most of the people become retired much earlier in Sweden. While one more year of schooling increases the probability of employment by 0.7 percentage points for men, it increases by 2.6 percentage points for women. Thus, the returns to schooling are higher for women than for men. Having immigrant background decreases the probability of employment more severely than any other single factor and with 15 percentage points more for women than for men with 13 percentage points. Both values are statistically significant at the one per cent level regardless of any specification.

Table 4a: Marginal Effects from *Probit* Estimation for Employment Probability, Men

	Specification I	II	III	IV	V
Homosexual	-.052051***	-.0779529***	-.0652209***	-.0644269***	-.0678315***
Age		.0398014***	.0328267***	.0261089***	.0264032***
Age²		-.0004538***	-.0003921***	-.0003274***	-.0003294***
Schooling		.0128063***	.011514***	.0070429***	.0070131***
Children			.002529	.0116288	.0128775
Immigrant			-.1816484***	-.1278855***	-.1279635***
Metropolitan Area			-.0083065	-.0215525	-.0209117
Northern Region			-.0109629	-.0190103	-.0189479
Unemployment Ins.				-.0000528*	-.0000719**
Social Contributions				-.0002567***	-.0002617***
Disposable Family Inc.				.0000215***	.0000216***
Housing Allowance				-.0001694	-.0001631
HomoUnempDays					.0006203
HomoUnempDays²					-1.51e-06
Statistics					
Number of Obs.	2898	2898	2898	2898	2898
Log- Likelihood	1132.429	-1074.846	-1011.666	-949.437	-948.037
χ^2	16.863	119.445	257.335	275.113	273.051
BIC	2280.802	2189.550	2095.078	2002.507	2015.651

* Statistically significant at 10%; ** Statistically significant at 5%; *** Statistically significant at 1%

Table 4b: Marginal Effects from *Probit* Estimation for Employment Probability, Women

	Specification I	II	III	IV	V
Homosexual	.050215***	-.0164149	-.0238756	-.0236344	-.0300253*
Age		.0422737***	.0398019***	.0370617***	.0373234***
Age²		-.0004914***	-.0004711***	-.0004447***	-.0004461***
Schooling		.0324917***	.0284375***	.0261351***	.0264226***
Children			.0054384	.0171694	.0187284
Immigrant			-.1733559***	-.1483515***	-.148993***
Metropolitan Area			.028662*	.0273108*	.0271495*
Northern Region			-.0044626	-.0129502	-.0130896
Unemployment Ins.				-.0001054***	-.0001171***
Social Contributions				-.0002381***	-.0002471***
Disposable Family Inc.				-6.22e-07	-6.24e-07
Housing Allowance				-.000251	-.0002567
HomoUnempDays					.0013319**
HomoUnempDays²					-4.41e-06**
Statistics					
Number of Obs.	2390	2390	2390	2390	2390
Log- Likelihood	-1032.754	-912.707	-878.739	-853.872	-851.114
χ^2	11.293	229.550	289.104	314.235	312.697
BIC	2081.066	1864.310	1827.489	1808.871	1818.915

* Statistically significant at 10%; ** Statistically significant at 5%; *** Statistically significant at 1%

6.3 Results from Survival Analysis

Table 5a and 5b present the *Hazard Ratios*, i.e. the incidence rate ratios of becoming employed for the different explanatory variables, by man and women separately. As mentioned previously, the length of time the event of employment occurs within, is explicitly incorporated into the analytical process. No statistically significant findings by sexual orientation are present, neither for men nor for women. Rather surprising findings are those for the covariate indicating the educational attainment for men. While the value for women is statistically insignificant, it is larger than one and thus, the incidence of finding a job increases *ceteris paribus* with every additional year of schooling. For men, the same variable is less than one and significant at the five per cent level. Thus, the value of 0.9 indicates a decreasing hazard by every additional year of education. This value becomes even smaller, but statistically insignificant when the squared term of education is incorporated and decreases to 0.87.

Much more striking are the findings for immigrant. The hazard rate ratios for immigrant are 0.58 for men and 0.65 for women in specification five, respectively at the five per cent and ten per cent levels of significance. Hence, the incidence of employment shrinks by 42 per cent for men and is 35 per cent lower for women when the individual has an immigrant background. While this drawback amounted in the probit estimation to -13 and -15 percentage points for man and women respectively, these values triple and double in absolute value when analysis time is explicitly incorporated. Thus, the findings become more severe when survival analysis techniques are used.

While our income vector has statistically significant variables, the values for men and women are around 1 and thus, do not change the incidence of employment practically. The estimates for the occupational vector provide a much different picture, which also deviates in its severity from the findings in the *Zero- Inflated Negative Binomial* model. While working in the manufacturing, construction and service sector more than double the incidence rate of employment for men, working in the healthcare sector increases the incidence rate of employment by almost five times. While some of the occupational values are not significant at the common levels of significance, working in the construction sector, which is a less common phenomenon for women, increases the incidence of employment by the factor seven, which is the highest occupational value for women. Followed by public administration with the factor 4.6 and healthcare with 2.8, the particular rootedness within a sector at least double the employment incidence for women. As theoretically discussed, those occupational choices where the average expected income is higher than in others, such as when healthcare and service sector are compared or public administration and service, will shorten the time in search, i.e. will increase the incidence of employment within a particular time interval. Of course, the diverse range of jobs within every occupation relativizes this conclusion, but even the finding for women and the construction sector points to the same if it is assumed that women do not work on the building site within the construction sector but more in the administrative and engineering part.

The incorporation of the occupation vector increases the hazard rate for homosexual men by ten points and makes it practically one and increases it for women by seven points, while both values remain statistically insignificant. This shows that a great part of the differences in employment ratios are not driven by sexual orientation, but at least to some extend by the occupational choices individuals typically make. The gamma values for all of our specifications for man and for specifications four and five for women are statistically different from zero, which thereby tells us that our *Gompertz* specification of the baseline hazard was correct and not the *Exponential* for instance, to which the model would reduce otherwise.

In all our three estimation methods, the model selection criteria provided at the bottom of the table, such as the *Bayesian Information Criterion* (BIC), point consistently into the same direction, i.e. the more parsimonious models in specification four are preferable to specification five, and four is more preferable in its complexity to the first three specifications. Either the addition of the occupation vector does not turn the main variable of our interest, i.e. sexual orientation, statistically significant, but the magnitude still changes in the *ZINB* and *Gompertz,* or the value changes marginally by the inclusion of days in unemployment as an explanatory variable.

Table 5a: Hazard Ratios from *Gompertz* Estimation, Men

	Specification				
	I	II	III	IV	V
Homosexual	1.2515699	1.1811471	1.0208013	.89978373	.99719
Age		1.0467774	1.0482949	1.0111523	.97189943
Age²		.9993899	.99933855	.99978513	1.0002396
Schooling		1.0005189	1.0052897	.90235638**	.90003901**
Children			.89456354	1.0365762	1.1152244
Immigrant			.76314021	.56071981***	.57684367**
Metropolitan Area			1.0958729	.79475834	.79473635
Northern Region			.95288379	.50851907*	.62320079
Income				1.0002131***	1.0001676**
Unemployment Ins.				.99803218***	.99802366***
Social Contributions				.99771337***	.99809595**
Disposable Family Inc.				1.0001061	1.0001454*
Housing Allowance				.99987705	.9991895
Agricultural Sector					2.1017864
Manufacturing Sector					2.7656777**
Construction Sector					2.219635*
Service Sector					2.3081164**
Healthcare Sector					4.643672***
Public Administration					1.9552414
Self-employed					1.7343301
Constant	.00426998***	.00199542***	.00228758***	.02288117*	.01827719**
γ	1.0021775**	1.0023241**	1.0023016**	1.0083189***	1.009082***
Statistics					
Number of Obs.	252	252	252	252	252
Log- Likelihood	-350.97324	-349.97547	-348.26844	-264.14889	-248.90049
χ^2	2.4285605	4.424096	7.8381594	176.07726	206.57407
BIC	718.53477	733.12752	751.83117	611.23922	619.44841

* Statistically significant at 10%; ** Statistically significant at 5% ; *** Statistically significant at 1%

Table 5b: Hazard Ratios from *Gompertz* Estimation, Women

	Specification				
	I	II	III	IV	V
Homosexual	1.1239117	1.0116356	1.0431252	.8446384	.9145688
Age		1.1615475*	1.1621303*	1.1322097	1.018771
Age²		.99816902*	.99812994*	.998396	.99974291
Schooling		1.0085652	1.0179664	1.0675479	1.0725583
Children			.99841169	.91101099	1.133797
Immigrant			1.0075121	.65209435*	.64787717*
Metropolitan Area			.89495659	.63441432*	.72147322
Northern Region			1.5192714	1.167981	1.2774729
Income				1.0002478**	1.000299***
Unemployment Ins.				.99775007***	.99776119***
Social Contributions				.9982216**	.99834414*
Disposable Family Inc.				1.0000966	1.0001218
Housing Allowance				.99992945	1.0000487
Agricultural Sector					2.3935765
Manufacturing Sector					2.1334692*
Construction Sector					6.923015*
Service Sector					2.2097086***
Healthcare Sector					2.7872141***
Public Administration					4.5878119**
Self-employed					2.5512201*
Constant	.00602431***	.00031969***	.00029825***	.000491***	.0011282***
γ	1.0004983	1.00061	1.0007907	1.008153***	1.0086007***
Statistics					
Number of Obs.	232	232	232	232	232
Log- Likelihood	-335.84052	-333.08804	-331.49294	-258.32581	-243.92332
χ^2	.57545493	6.0804233	9.2706096	155.60488	184.40985
BIC	688.02125	698.8565	717.45326	598.35268	607.67486

* Statistically significant at 10%; ** Statistically significant at 5% ; *** Statistically significant at 1%

7 DISCUSSION

The detection and fight against discrimination is a vital point on the agenda of the European Union in general and the Swedish in particular. Previous research found discrimination against homosexual individuals in Sweden, with smaller magnitude in comparison to other countries. The present study has been devoted to the analysis of differences in the duration and probability of unemployment by sexual orientation, using data of coupled homosexual individuals and married heterosexual. Six hypotheses have been tested for. Based on theoretical considerations homosexual men were assumed to have longer durations in unemployment than heterosexual men and homosexual women shorter durations than heterosexual women. In the same manner, homosexual men were hypothesized to have a lower probability of unemployment, while homosexual women were assumed to have a higher one, when compared to their heterosexual peers. The same pattern of odds was assumed to appear, when duration in and the probability of unemployment was fused into the *hazard ratio*.

A three-stage modelling approach has been utilized. *Zero- Inflated Negative Binomial* regression has been conducted to the duration in unemployment, while *Probit* analysis to the probability of employment. For the first time, survival analysis techniques have been used for modelling labour market outcomes by sexual orientation. A parametric *Gompertz* baseline hazard has been determined, which enables us to fuse unemployment duration and employment event into one analytical feature, which increases the power of the analysis and leads to different results in comparison to the usage of the single component methods, as previous studies did. Statistically significant differences have been found regarding the probability of employment, but not the duration in unemployment or the hazard ratio. The probability of employment decreases by seven per cent for males and three per cent for females, when the individual is homosexual, while the finding for women is significant at the ten per cent level. The negative difference to heterosexual women vanishes significantly the longer the duration in unemployment takes. When both parameter estimates are put into an equation, we detect a threshold of 23 days, after which the probability of employment is significantly higher for homosexual women. The conduction of the parametric survival model points to no differences by sexual orientation, when unemployment probability and duration are set into relation. It is observable that differences in duration and probability of employment are driven by income and employment factors, as search theory suggests and by human capital variables. Migration background has a significant impact on the unemployment probability and duration in a statistical sense, as well as in magnitude, which has been substantiated by a broad range of literature for Sweden and other countries.

8 ACKNOWLEDGEMENTS

The author is particularly grateful to Professor Dominique Anxo and Professor Lina Alden, Department of Economics and Statistics at Linnaeus University, for fruitful discussions and great advice. Also grateful to Adonia Odessa and Fidel Deniz Can, who are always my muse.

9 REFERENCES

Adam, B. D. (1981). Stigma and employability: Discriminatin by sex and sexual orientation in the Ontario legal profession. *Canadian Review of Sociology and Anthropology , 18* (2), pp. 216-221.

Ahmed , A. M. (2008). If you Believe that Discrimination Exists, it will . (B. Publishing, Ed.) *The Manchester School , 76* (6).

Ahmed, A. M. (2010a). Muslim discrimination: Evidence from two lost letter experiments. *Journal of Applied Social Psychology , 40*, pp. 888-898.

Ahmed, A. M. (2010b). What is in a surname? The role of ethnicity in economic decision making. *Applied Economics , 42*, pp. 2715-2723.

Ahmed, A. M., & Hammarstedt, M. (2008). Discrimination in the rental housing market: A field experiment on the internet. *Journal of Urban Economics , 64*, pp. 362- 372.

Ahmed, A. M., & Hammarstedt, M. (2009). Detecting discrimination against homosexuals: Evidence from a field experiment on the internet. *Economica , 76*, pp. 588- 597.

Ahmed, A. M., Andersson, L., & Hammarstedt, M. (2008). Are lesbians discriminated against in the rental housing market? Evidence from a correspondence testing experiment. *Journal of Housing Economics , 17*, pp. 234- 238.

Ahmed, A. M., Andersson, L., & Hammarstedt, M. (2009). Ethnic discrimination in the market place of small business transfers. *Economics Bulletin , 29*, pp. 3050-3058.

Ahmed, A. M., Andersson, L., & Hammarstedt, M. (2010). Can discrimination in the housing market be reduced by increasing the information about the applicants? *Land Economics , 86*, pp. 79-90.

Ahmed, A. M., Andersson, L., & Hammarstedt, M. (2011a). Inter- and intra-household earnings differentials among homosexual and heterosexual couples. *British Journal of Industrial Relations, 49*, pp. 258-278.

Ahmed, A. M., Andersson, L., & Hammarstedt, M. (2011b). Are gay men and lesbians discriminated against in the hiring process? *Southern Economic Journal* .

Ahmed, A. M., Andersson, L., & Hammarstedt, M. (2012). Does age matter for employability? A field experiment on ageism in the Swedish labor market. *Applied Economics Letters , 19*, pp. 403-406.

Aldén, L., Edlund, L., Hammarstedt, M., & Mueller-Smith, M. (2015). Effect of Registered Partnership on Labor Earnings and Fertility for Same-Sex Couples: Evidence From Swedish Register Data. (S. US, Ed.) *Demography , 52* (4), pp. 1243–1268.

Anxo , D., Ericson , T., Herbert, A., & Rönnmar, M. (2017) *To Stay or not to Stay. That is the Question.*

Arabsheibani, R. G., Marin, A., & Wadsworth, J. (2004). In the pink: Homosexual-heterosexual wage differentials in the UK. *International Journal of Manpower , 25*, pp. 343 - 354.

Becker, G. S. (1957). *The Economics of Discrimination.* Chicago, Illinois, USA: University of Chicago Press.

Becker, G. S. (1981). *A Treatise on the Family.* Cambridge, Massachusetts, USA: Harvard University Press.

Bertrand , M., Chugh, D., & Mullainathan, S. (2005). Implicit Discrimination. *The American Economic Review , 95* (2), pp. 94-98.

Black, D. A., Sanders, S. G., & Taylor, L. J. (2007). The economics of lesbian and gay families. *The Journal of Economic Perspectives , 21*, pp. 53-70.

Carlsson, M., & Rooth, D.-O. (2012). Revealing taste- based discrimination in hiring: A correspondence testing experiment with geographic variation. *Applied Economics Letters* (19), pp. 1861-1864.

Chugh, D. (2004). Societal and Managerial Implications of Implicit Social Cognition: Why Milliseconds Matter. *Social Justice Research , 17* (2), pp. 203–222.

Clain, S. H., & Leppel, K. (2001). An investigation into sexual orientation discrimination as an explanation for wage differences. *Applied Economics , 1*, pp. 37-47.

Cleves , M., Gould, W. W., & Marchenko, Y. V. (2016). *An Introduction to Survival Analysis Using Stata, , Revised Third Edition.* (S. LLC, Ed.) Texas, USA: Stata Press.

Drydakis, N. (2009). Sexual orientation discrimination in the labour market. *Labour Economics , 16*, pp. 364- 372.

Drydakis, N. (2011). Women`s sexual orientation and labor market outcomes in Greece. *Feminist Economics , 17* (1), pp. 89-117.

Ehrenberg , R. G., & Oaxaca , R. L. (1976). Unemployment Insurance, Duration of Unemployment, and Subsequent Wage Gain. *The American Economic Review , 66* (5).

Ellis, R. A., & Taylor , S. M. (1983). Role of Self-Esteem Within the Job Search Process. *Journal of Applied Psychology , 68* (4), pp. 632-640.

Eurostat. (2017). *ec.europa.eu.* Retrieved February 15, 2018 from eurostat Statistics Explained : http://ec.europa.eu/eurostat/statistics-explained/index.php/Gender_pay_gap_statistics#Gender_pay_gap_levels

Grossbard, S. & Jepsen, L. K. (2008). The economics of gay and lesbian couples: Introduction to a special issue on gay and lesbian households. *Review of Economics of the Household , 6*, pp. 311-325.

Hebl, M. R., Foster, J. B., Mannix, L. M., & Dovidio, J. F. (2002). Formal and interpersonal discrimination: A field study of bias toward homosexual applicants. *Personality and Social Psychology Bulletin , 28* (6), pp. 815-825.

Jepsen, C. A., & Jepsen, L. K. (2006). The sexual division of labor within households: comparisons of couples to roommates. *Eastern Economic Journal , 32*, pp. 299-312.

Kanfer, R., Wanberg, C. R., & Kantrowitz, T. (2001). Job Search and Employment: A Personality-Motivational Analysis and Meta-Analytic Review. *Journal of Applied Psychology , 86*, pp. 837-855.

Leppel, K. (2009). Labour force status and sexual orientation. *Economica , 76*, pp. 197-207.

Long , S. J., & Freese, J. (2006) *Regression Models for Categorical Dependent Variables Using Stata* (2 ed.). Texas, USA: Stata Press.

Merton , R. K. (1948). The Self-Fulfilling Prophecy. *The Antioch Review , 8* (2), pp. 193-210.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *American Economic Review , 62*, pp. 659–661.

Pierce, J. L., & Gardner, D. G. (2004). Self-Esteem Within the Work and Organizational Context: A Review of the Organization-Based Self-Esteem Literature. *Journal of Management , 30* (5), pp. 591–622.

Ragins, B., Singh , R., Cornwell , R., & John, M. (2007). Making the Invisible Visible: Fear and Disclosure of Sexual Orientation at Work. (A. P. Association, Ed.) *Journal of Applied Psychology , 92* (4), pp. 1103-1118.

Riach, P., & Rich, J. (2002, November). Field experiments of discrimination in the marketplace. *The Economic Journal , 112*.

Rooth, D.-O. (2010, June). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics , 17* (3), pp. 523–534.

Tebaldi, E., & Elmslie, B. (2006). Sexual orientation and labour supply. *Applied Economics , 38*, pp. 549-562.

Therneau, T. M., & Grambsch, P. M. (2001). *Modeling Survival Data: Extending the Cox Model* (2 ed.). Springer.

Weichselbaumer, D. (2003). Sexual orientation discrimination and hiring. *Labour Economics , 10* (6), pp. 629-642.

10 APPENDIX 1: VARIABLES

Table A1: Dependent and Explanatory Variables Used in Estimation Procedures

Dependent Variables:	Explanation:
$\Pr(y_i \mid \boldsymbol{x}_i, A_i = 1)$	Dependent variable in the Zero- Inflated Negative Binomial count model. Probability of the occurrence of a particular duration in unemployment when the individual is not in the certain- zero group.
$\Pr(Employed = 1)_i$	Dependent variable in the Probit estimation. Probability of becoming employed.
$h(t \mid \boldsymbol{x}_j)$	Dependent variable in Gompertz regression. The hazard, incidence of becoming employed within a particular duration of unemployment.
Independent Variables:	
Homosexual	1 if individual is homosexual 0 otherwise
Age	Individual's age in years
Age²	Squared age of individual in years
Schooling	Individual's educational attainment measured in years
Children	1 if children below the age of eighteen are present in the household 0 otherwise
Immigrant	1 if individual is foreign born

	0 otherwise
Metropolitan Area	1 if individual resides in the counties of Stockholm, Vastra, Gotaland or Skane 0 otherwise
Northern Region	1 if individual resides in the counties of Vasternorrland, Jamtland, Vasterbotten or Norrbotten 0 otherwise
Income	Individual's yearly earnings measured in hundreds of SEK (Swedish Krona)
Unemployment Ins.	Received payment from unemployment insurance in SEK (Swedish Krona)
Social Contributions	Sum of different kinds of transfer payments. Not necessarily caused by previous work. Yearly amount measured in SEK (Swedish Krona)
Disposable Family Inc.	Total sum of yearly disposable family income measured in hundreds of SEK (Swedish Krona)
Housing Allowance	The amount is stated in hundreds of SEK. The variable is summarized at family level and includes both housing allowance for families with children, housing allowance for young people, housing allowance for other families and single without children.
Agricultural Sector	1 if individual is employed in the agricultural sector 0 otherwise
Manufacturing Sector	1 if individual is employed in the manufacturing sector 0 otherwise
Construction Sector	1 if individual is employed in the construction sector 0 otherwise
Service Sector	1 if individual is employed in the service sector 0 otherwise
Healthcare Sector	1 if individual is employed in the healthcare sector 0 otherwise
Public Administration	1 if individual is employed in public administration 0 otherwise
Self-employed	1 if individual is self- employed 0 otherwise

11 APPENDIX 2: *ZERO- INFLATED NEGATIVE BINOMIAL* REGRESSION– MODEL SELECTION

Table A2 provides a compact comparison between four common count models, generated via the Stata command countfit. It shows how the *Zero- Inflated Negative Binomial* (ZINB) model is preferred over the *Poisson* regression model (PRM), the *Negative Binomial* regression model (NBRM) and the Zero- Inflated Poisson (ZIP) model, which is obtainable from the last subsection of the comparative blocks within the table, where ZINB is compared to the pertaining model. Table A3 shows that the ZINB is preferable over OLS. All provided model selection criteria point to the same direction. The *Bayesian Information Criterion* (BIC) is with 3751 for the *ZINB* much lower than 27699 for the *OLS* and thus, our choice in favour of the count model is also confirmed in the direct comparison via statistical model selection criteria.

These findings are equally valid for men and women, while here only the data for men is utilized.

Table A2: Comparison between Poisson, Negative Binomial and Zero- inflated versions of both

PRM	**BIC= 75408.633**	**AIC=75283.226**	**Prefer**	**Over**	**Evidence**
vs	BIC= 4866.118	dif= 70542.515	NBRM	PRM	Very strong
NBRM	AIC= 4734.739	dif= 70548.487	NBRM	PRM	
	LRX2=70550.487	prob= 0.000	NBRM	PRM	p=0.000
vs	BIC= 13914.472	dif= 61494.162	ZIP	PRM	Very strong
ZIP	AIC= 13663.657	dif= 61619.569	ZIP	PRM	
	Vuong= .	prob= .	ZIP	PRM	p=.
vs	BIC= 4122.279	dif= 71286.355	ZINB	PRM	Very strong
ZINB	AIC= 3865.492	dif= 71417.734	ZINB	PRM	
NBRM	**BIC= 4866.118**	**AIC= 4734.739**	**Prefer**	**Over**	**Evidence**
vs	BIC= 13914.472	dif= -9048.354	NBRM	ZIP	Very strong
ZIP	AIC= 13663.657	dif= -8928.918	NBRM	ZIP	
vs	BIC= 4122.279	dif= 743.839	ZINB	NBRM	Very strong
ZINB	AIC= 3865.492	dif= 869.247	ZINB	NBRM	
	Vuong= .	prob= .	ZINB	NBRM	p=.
ZIP	**BIC= 13914.472**	**AIC=13663.657**	**Prefer**	**Over**	**Evidence**
vs	BIC= 4122.279	dif= 9792.193	ZINB	ZIP	Very strong
ZINB	AIC= 3865.492	dif= 9798.165	ZINB	ZIP	
	LRX2= 9800.165	prob= 0.000	ZINB	ZIP	p=0.000

Table A3: Comparison between *Zero- Inflated Negative Binomial* and *Ordinary Least Squares* regression

	ZINB	OLS
Homosexual	.14767169	-.0321585
Age	-.0315881	-.41038188
Age²	.00023691	.00265898
Schooling	-.03657335	.57853379*
Children	-.22392648	-2.1629138
Immigrant	.08039716	4.9231741***
Metropolitan Area	.24656665	.77649496
Northern Region	.07077793	5.2219738*
Income	-.00008909	-.00113823***
Unemployment Ins.	.00139078***	.19833701***
Social Contributions	.0009149*	.05395878***
Disposable Family Inc.	.00001001	.0000138
Housing Allowance	.00045158	.06084512**
Agricultural Sector	.35799111	3.4253557
Manufacturing Sector	.05151849	2.6661762
Construction Sector	-1.0165188	5.1057794
Service Sector	.07413056	1.852295
Healthcare Sector	.07595779	-.99066959
Public Administration	-.50954518	2.5820091
Self-employed	-.08270262	.25439262
Constant	5.0510317***	8.5754368
Statistics		
Number of Obs.	2390	2898
Log- Likelihood	-1708.4106	-13766.032
AIC	3502.8213	27574.064
BIC	3751.3203	27699.471

12 Appendix 3: Survival Analysis – Model Selection and Parameterization Process

In this section we will justify the particular modelling approach for the survival model via statistical model selection criteria. Since different kind of covariates are present in our data set, a nonparametric survival model cannot be utilized and a choice has to be made among semiparamteric and parametric models, i.e. respectively the baseline hazard has to be left unparameterized or has to be specified with a particular functional form. We start with the selection among the most popular six parametric models, i.e. the *Generalized Gamma*, *Exponential*, *Weibull*, *Lognormal*, *Loglogistic* and *Gompertz* models and will observe which of them provides the best fit. After that, we will compare our chosen model with the less constrained semiparametric Cox PH model and will select among the two via the Schwarz/Bayesian Information Criterion (BIC), while other criteria will be provided.

Following Cleves *et al.* (2006, p.279), we start with the *Generalized Gamma* model and use the property that the *Lognormal*, *Weibull* and *Exponential* models are nested in the *Generalized Gamma*, which is a three- parameter baseline model and its distribution depends on β_0, κ and σ. If κ equals 1, the model could be *Weibull*. If κ and σ simultaneously equal 1, we assume the model to be *Exponential* and if κ equals zero, we assume a *lognormal* distribution. The full set of variables is utilized for the subsection of data for men, while all findings below are valid for women as well.

Table A4: Results for *Generalized Gamma* regression

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Homosexual	-.050048	.1496239	-0.33	0.738	-.3433054	.2432093
Age	.0193587	.0407193	0.48	0.634	-.0604497	.0991672
Age²	-.0001604	.000477	-0.34	0.737	-.0010953	.0007746
Schooling	.0570254	.0215782	2.64	0.008	.0147328	.0993179
Children	-.0325181	.174143	-0.19	0.852	-.3738321	.3087959
Immigrant	.3375385	.1043025	3.24	0.001	.1331093	.5419677
Metropolitan Area	.1211448	.1175558	1.03	0.303	-.1092603	.3515499
Northern Region	.1701142	.2159755	0.79	0.431	-.2531899	.5934184
Income	-.000076	.0000352	-2.16	0.031	-.000145	-6.93e-06
Unemployment Ins.	.0009387	.000155	6.05	0.000	.0006348	.0012425
Social Contributions	.0009077	.0003678	2.47	0.014	.0001869	.0016285
Disposable Family Inc.	-.0000895	.00004	-2.24	0.025	-.0001679	-.000011
Housing Allowance	.0000427	.0010214	0.04	0.967	-.0019593	.0020447
Agricultural Sector	-.5440413	.4570141	-1.19	0.234	-1.439772	.3516898
Manufacturing Sector	-.738801	.2186953	-3.38	0.001	-1.167436	-.310166
Construction Sector	-.5466844	.233199	-2.34	0.019	-1.003746	-.0896229
Service Sector	-.5708928	.1802816	-3.17	0.002	-.9242382	-.2175474
Healthcare Sector	-.9504911	.2042264	-4.65	0.000	-1.350768	-.5502147
Public Administration	-.5034673	.2564276	-1.96	0.050	-1.006056	-.0008785
Self-employed	-.3749802	.1585277	-2.37	0.018	-.6856888	-.0642717
Constant	4.229843	.9040308	4.68	0.000	2.457975	6.001711
ln(σ)	-.6147324	.1263795	-4.86	0.000	-.8624317	-.367033
κ	1.7129	.3382523	5.06	0.000	1.049937	2.375862
σ	.5407856	.0683442		.4221343	.6927868	.5407856

As table A4 shows we obtain the p- value for the Wald test for H0: κ = 0 directly from the output and reject the null since the p-value is 0.000, hence, we filter out a lognormal model. Now we test whether κ equals 1.

Table A5: Test for κ equals one

H0: κ = 1
χ^2 (1) = 4.44
Prob > χ^2 = 0.0351

We reject the null hypothesis of $\kappa = 1$, since the p-value of 0.0351 in table A5 indicates a rejection at the five per cent level and thus, the model is not assumed to be Weibull. In the next step we use a Wald test and test for $\kappa = 1$ and $\sigma = 1$ simultaneously, which is the same as ln of σ equals zero and obtain the following results in table A6.

Table A6: Simultaneous test for κ equals one and ln(σ) equals zero

H0: $\kappa = 1$ & ln(σ) = 1
χ^2 (2) = 43.35
Prob > χ^2 = 0.0000

Following the p-value for the test, we reject the null hypotheses of $\kappa = 1$ and $\sigma = 1$ and reject an exponential distribution. Thus, the "risk" of getting employed, in our case the luck of getting employed, is not constant over time, which is assumed when the model is exponential. In the next step the non- nested *Log- Logistic* and *Gompertz* models as well as the semiparamteric *Cox PH* will be compared which is shown in table A7.

Table A7: Comparison between *Log- Logistic, Gompertz* and Semiparametric *Cox PH*

	Log- Logistic	*Gompertz*	*Cox PH*
Homosexual	.01368975	-.00281396	-.04129247
Age	.02775668	-.02850295	-.00936117
Age²	-.00032022	.00023959	.00003797
Schooling	.05437681	-.10531717**	-.10082344**
Children	-.15150084	.10905568	.0673947
Immigrant	.2415826	-.55018399**	-.4506015**
Metropolitan Area	.15632142	-.22974485	-.23148177
Northern Region	.36284247	-.47288651	-.45529996
Income	-.00018121*	.00016756**	.00014547*
Unemployment Ins.	.00143411***	-.00197829***	-.00198699***
Social Contributions	.00089196*	-.00190586**	-.00203257***
Disposable Family Inc.	-.00007923	.00014542*	.00014783*
Housing Allowance	.00119055	-.00081083	-.00120848
Agricultural Sector	-.14181291	.74278766	.87074029
Manufacturing Sector	-.3821419	1.0172857**	1.0319693**
Construction Sector	-.47639336	.79734279*	.85945734*
Service Sector	-.53357043**	.83643179**	.85079584***
Healthcare Sector	-.87122511***	1.5355054***	1.5339692***
Public Administration	-.4781647	.67051368	.56900601
Self-employed	-.05672956	.5506212	.48409479
Constant	3.482649**	-4.0021016**	
ln(γ)	-.67832309***		
γ		.00904096***	
Statistics			
Number of Observations	252	252	252
AIC	616.6652	541.80097	1673.7974
BIC	694.31264	619.44841	1744.386

* Statistically significant at 10%; ** Statistically significant at 5% ; *** Statistically significant at 1%

Both criteria, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), have their lowest values for the Gompertz model, which indicates that the Gompertz model is the best fitting among the parametric models. When compared to the more flexible semiparametric *Cox PH,* the parametric Gompertz model is still the preferable one.

13 AUTHOR INFORMATION

Can Karaarslan is an economist working in research, lecturing and counselling. His main focus of research is in the fields of Labour Economics, Environmental Economics, Climate Economics, Political Economy and Microeconometrics. Can Karaarslan is member of the American Economic Association, Econometric Society, Society of Labor Economists and the Association of Environmental and Resource Economists.