Unobserved Heterogeneity and Labor Market Disparities against Homosexuals*

Miguel Sarzosa†
Purdue University

December 6, 2019

Abstract

Sexual minorities have historically been subject to many kinds of discrimination. Prejudicial treatment in the labor market could arguably be one of them. Despite that, economic literature has remained mostly silent on the topic. This paper fills that void by leveraging on a novel longitudinal data set that collects detailed information on sexual orientation. I develop an empirical strategy that exploits the fact that sexuality is not a dichotomous trait but rather a wide assortment of sexual preferences. I use empirical models that rely on the identification of unobserved heterogeneity, in the forms of skills and sexual orientation, to allow schooling, employment, and income to be endogenously determined. I find that, after controlling for differences in skills distributions, there are no income gaps against employed homosexuals. However, I find that homosexuals are 10–20 percentage points less likely to be employed than heterosexuals. These gaps are larger among men and the college educated. The results suggest that selection on the employment margin contributes to the elimination of the observable income gaps as the average employed homosexual is more skilled than the average employed heterosexual.

Keywords: Wage gap, discrimination, unobserved heterogeneity, skills.

JEL codes: J15, J24, J71

*This research was in part supported by the W.E. Upjohn Institute for Employment Research through their Early Carrer Research Grant program. I would like to thank Sergio Urzua for his guidance during this research project, Bruce Elmslie and the participants of the UMD Economics Department Seminar (2015, College Park, MD), the Easter Economic Association Meetings (2015, Washington, DC), the Economics Summer Seminar at Purdue (2015, West Lafayette, IN), and the European Economic Association Meetings (2017, Lisbon, Portugal) .

†403 W. State Street, West Lafayette, IN, 47907; phone: (765) 494-4343; fax: (765) 494-9658; email, msarzosa@purdue.edu. The author did not have access to information leading to the identification of individuals. The data analysis was carried out in a secure server.

1 Introduction

Ever since Becker (1957), labor market discrimination has been a popular topic in economics.[1] In particular, labor economists have devoted substantial effort in analyzing labor market discrimination against women and racial minorities.[2] Some of their findings contributed to the implementation of a number of antidiscriminatory policies around the world. However, much less is known regarding labor market discrimination toward sexual minorities, despite the progress made by this population in securing social rights in recent years (Mucciaroni, 2009; Frank, 2014). Some literature has documented the existence of a sexual minority wage gap. In particular, this literature has shown that homosexual men earn less than their heterosexual counterparts by about 15–30 percent (Badgett, 1995; Black et al., 2003; Carpenter, 2007). Evidence on women is less conclusive. Some studies suggest gay/bisexual women earn less than heterosexual women (Badgett, 1995); others argue the opposite (Berg and Lien, 2002; Black et al., 2003, 2007; Antecol et al., 2007).

Although these studies have made important contributions to our understanding of sexual minorities' labor market outcomes, there are at least two major problems with the existing literature. First, they do not consider that schooling and occupational choices are endogenous and themselves possibly affected by the sexual preferences as well.[3] And second, the income gap might respond to

[1] Defining labor market discrimination is not a simple matter. In this paper, I take a broad view of discrimination defining it in terms of members of the minority group (i.e., the group that suffers discrimination) being subject to inequitable treatment that may lead to disparate outcomes (Cain, 1986).

[2] See, for example, Cain (1986) survey article that collected the extensive evidence available by then on racial and gender labor market discrimination; Mincer and Polachek (1974) link between the family decisions on differential human capital investments across genders and earnings' gap observed later in life, which relates to Neal and Johnson (1996) findings where they account for most of the black-white wage gap by differences in cognitive skills; more recently, Urzua (2008) introduces differentials in abilities rather than test scores as partial determinants of the black-white wage gap, and Lang and Lehmann (2012) who review the existing literature on racial discrimination and evaluate the capacity different theoretical models have to explain the empirical regularities of discrimination we observe today; finally, see Blau and Kahn (2017) for a recent review of the current state of the literature on the gender wage gap, where they suggest that human capital factors are no longer able to explain it.

[3] Evidence suggest that college outcomes are in general different for homosexual men (Carpenter, 2009); that cohabiting homosexuals have significantly higher levels of education than heterosexuals (Allegretto and Arthur, 2001; Antecol et al., 2007); that gay cohabiting men supply less labor than married heterosexual men, while

unobservables, namely skill differentials, not only in productivity terms but also in the way people choose their schooling level or occupation. This paper overcomes these two issues by using an empirical strategy that controls for unobserved heterogeneity, and based on them allows choices and labor market outcomes to be endogenous.

Another problem that plagues this literature is the lack of appropriate data sets. Papers on sexual-preference discrimination have relied mainly on two sources of data: the General Social Survey (GSS) or the census.[4] Each has its own problems. On one hand, the GSS provides a relatively detailed description of sexual preferences but has a very small number of members of sexual minorities that hinders empirical work on the topic (for instance, the sample used by Badgett (1995) comprises only 34 homosexual women and 47 homosexual men). On the other hand, the census provides a larger sample size, but the definition of who belongs to a sexual minority has important setbacks. Researchers using the census identify homosexuals from people who cohabit with a person of the same gender (see, for instance, Clain and Leppel, 2001; Allegretto and Arthur, 2001; Antecol et al., 2007; Antecol and Steinberger, 2013). Therefore, those identified comprise a selected sample that excludes people who have chosen to remain single (Black et al., 2007). The cultural and legal restrictions imposed to gay marriage give the grounds to claim that the selection into marrying/cohabiting is very different for homosexuals than for heterosexuals, and that it may rely heavily on unobservables that also affect the schooling and occupational choices and the wages earned.[5] I overcome these problems by using a novel longitudinal Norwegian data set that provides

lesbians supply more labor than heterosexual women (Tebaldi and Elmslie, 2006)

[4]See Black et al. (2007) for a detailed description of the data sets that contain data on sexual preferences in the United States. Carpenter (2007) uses the National Health and Nutrition Examination Surveys but faces the same problems I describe for the papers that use the GSS. Tebaldi and Elmslie (2006) use the Current Population Survey. These data face issues analogous to the ones I describe for the papers that use the census.

[5]Antecol and Steinberger (2013) show that the gap between primary and secondary earners within a household of cohabiting lesbians is smaller than than found in a heterosexual household. This could relate to differences in selection into marrying/cohabiting across sexual orientation, and thus, be one of the many vehicles through which limiting the sample to partnered adults may be biasing the labor market outcome estimates.

very detailed information about sexual preferences, allows me to have an adequate sample size, and importantly, allows me to measure skills as a latent variable. This way, I do not have to restrict the study to nonsingle homosexuals, and I am able to avoid the source of confoundedness that arise from such a selected sample.

An additional contribution of this paper is that it considers homosexuality a continuous latent variable. This novel approach goes in line with the fact that, widely accepted in psychology, sexuality is not a dichotomous trait but rather a complex assortment of sexual preferences (Laumann et al., 1994). The treatment of homosexuality as an unobserved—to the econometrician—continuous variable provides an opportunity to better understand the effect that different degrees of homosexuality can have on the labor force.

Furthermore, I extend the model of potential outcomes with unobserved heterogeneity used in Heckman et al. (2006) and Urzua (2008) to incorporate the endogenous choice of employment, and thus address selection issues that could affect the estimation of an income gap—namely, that the absence of an income gap does not rule out the existence of discrimination. In fact, it could be the results of discrimination at hiring (Brown, 1984). Income gaps may not be evident within the employed population because only the highly skilled within the minority will manage to get a job. Therefore, quantifying the effect of an entrance barrier to the labor market for sexual minorities becomes particularly relevant.

Gaps against sexual minorities in the probability of being employed could materialize through both the demand and the supply sides of the labor market. Regarding the former, one can extend the Beckerian model of discrimination to include the fact that if the *compensation*—the gap between the wages the employer is willing to pay majority workers over minority ones—needed by the employer to outweigh her distaste for minority workers is big enough, it could cause the wage

offered to minorities to be less than their reservation wage.[6] However, current economic models struggle to reconcile discrimination stemming from the demand side with the fact that homosexuality is often veiled and to some extent easier to strategically conceal than gender and race. This makes more difficult for prejudiced employers to discriminate against sexual minorities. French reports extensively document members of sexual minorities hiding their sexual identity during the hiring precess (Falcoz, 2008).

Supply side-genarated disparities may be driven by the fact that sexual minorities often have to face hostile workplace cultures (Frank, 2014), and thus, their reservation wages may be higher in order to compensate for that inconvenience when taking on a job.[7] A report from the European Commission found evidence suggesting that being a member of a sexual minority negatively affects working conditions, salaries and career opportunities (van Balen et al., 2010). Seventy percent of the respondents in a German study report using information regarding gay-friendly workplaces when looking for a job (Frohn, 2007). Workplace hostility or the aversion to a hetero-normative workplace culture may be grater among workers placed higher in the homosexuality spectrum. This would yield a negative relation between the likelihood of being employed and a person's degree of homosexuality, even if the latter is not observed by employers. Although both demand and supply side channels differ profoundly, they amount to unfair disparities as they cause minority agents to behave differently in the labor market, obtaining different outcomes from it.

I find evidence on the existence of labor market disparities against sexual minorities. My results

[6]This intuition on the effect of discrimination on employment is general enough to include specific mechanisms explored recently in the literature like uncertainty in the subjective evaluation of minority workers by majority managers (Ritter and Taylor, 2011).

[7]This angle relates to the one explored by Bond and Lehmann (2018) where black workers lower their reservation wage when they are matched with firms that have black supervisors, as it signals lower firm's prejudice. Black workers are willing to work for less in these firms in order to secure a less prejudiced environment that would yield lower probabilities of being fired.

indicate that the disparities exist at the employment margin and not at the earnings margin for those employed. While I find no income gaps among the employed, I find that the overall employment rate is between 10 and 20 percentage points lower for homosexuals than for heterosexuals.[8] I find that the disparities are stronger among men than women. The endogenous selection into employment reduces the income gaps we observe among the employed. Employment probability disparities cause employed homosexuals to be, on average, more skilled than the average employed heterosexual, and the former face higher labor market returns to those skills. Finally, I find that labor market disparities against sexual minorities differs across genders. Among homosexual men, the college educated ones are less likely to find a job. Among homosexual women, those with no college degree are the ones who do.

The paper is organized as follows. Section 2 introduces the definitions of homosexuality used in the paper and presents evidence on the income gap. Section 2.1 describes the data and background. Section 3 introduces the empirical strategy. Section 4 presents the first set of results and discusses whether the differences between homosexuals and heterosexuals in terms of income, schooling choices and occupational choices respond to differences in skills. Section 5 presents the analysis of homosexuality as a latent factor. Finally, Section 6 concludes.

[8]This challenges the argument of Cain (1986) that states that there should not be employment rate gaps against minorities without the existence of earning gaps, because if minorities are consistently rejected and not hired, they must bid for lower paying jobs. It provides some suggestive evidence indicating that disparities may come from the self selection into employment as the reservation wages of homosexuals seem to be higher than that of comparable heterosexuals; in response, probably, to the psychic cost of facing a hostile workplace environment.

2 Data and Background

2.1 Data

This paper uses a novel data set that has not been explored by the economic literature. The Young in Norway (YiN) is a longitudinal data set product of a research project headed by the Program for Adolescent Research (Ungforsk) in 1990 (Strand and von Soest, 2008). YiN collects detailed information about the characteristics, attitudes, and choices of a cohort of Norwegians. In particular, it has four features that makes it the best data set currently available to study sexual minority labor market discrimination. First, it is longitudinal; therefore, I am able to observe respondents' characteristics when young. This improves over existing literature that explores income gaps of sexual minorities using nonlongitudinal surveys, which in turn prevents researchers from differentiating away time-invariant unobservables that might bias the results. Second, it collects scores on personality traits and academic performance as children. I use these measures to identify the unobserved skills, a key feature of my empirical strategy. Third, it collects detailed information about sexual preferences and practices. This way, I do not run into problems with small samples, I can include single people in the analysis, and I can construct a model with a continuous measure of unobserved heterogeneity on homosexuality. Finally, YiN collects information about the job market including ISCO88 coded occupations and yearly income. See the descriptive statistics Table 1.

YiN has four waves of data. The first wave was collected in 1992, when 12,287 students of lower and upper secondary school pupils, from 67 schools in grades 7–12 (age 13–18) were interviewed. The sampling allowed for each grade to be represented equally, and the sampling unit was the school. Wave 2 was collected two years later. By then, the older members of the sample had

Table 1: Descriptive Statistics

		Overall		Males		Females		Heterosex.	
Variable	N	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Demographic									
Age *(Wave 4)*	2958	27.465	1.844	27.384	1.788	27.531	1.886	27.538	1.860
Female *(%)*	2958	0.548	0.498					0.465	0.499
Father's Educ (Wave 1, %)									
University	2958	0.153	0.361	0.153	0.360	0.154	0.361	0.148	0.355
3yr College	2958	0.090	0.286	0.092	0.289	0.088	0.283	0.086	0.281
Vocational	2958	0.138	0.345	0.134	0.341	0.141	0.348	0.141	0.349
Mother's Educ (Wave 1, %)									
University	2958	0.073	0.260	0.073	0.259	0.073	0.261	0.064	0.244
3yr College	2958	0.204	0.403	0.203	0.403	0.204	0.403	0.199	0.399
Vocational	2958	0.038	0.191	0.037	0.190	0.038	0.192	0.039	0.193
Household Composition and SES (Wave 1)									
Siblings	2958	1.720	1.095	1.694	1.094	1.742	1.096	1.727	1.062
Both Parents *(%)*	2958	0.772	0.420	0.777	0.416	0.767	0.423	0.789	0.408
Father:High	2958	0.268	0.443	0.271	0.445	0.266	0.442	0.263	0.441
Father:Skilled	2958	0.139	0.346	0.135	0.341	0.143	0.350	0.141	0.349
Mother:High	2958	0.069	0.253	0.076	0.264	0.064	0.244	0.060	0.238
Mother:Skilled	2958	0.014	0.118	0.012	0.109	0.016	0.126	0.014	0.116
Scores (Wave 1)									
Soc. Accept.	2892	0.000	0.720	-0.012	0.717	0.010	0.722	0.012	0.714
Sympathetic	2891	-0.001	0.731	-0.173	0.735	0.140	0.697	-0.014	0.737
Self-esteem	2893	0.000	0.733	0.142	0.722	-0.117	0.722	0.038	0.730
Leadership	2836	-0.002	0.581	0.034	0.599	-0.031	0.564	-0.007	0.572
Grades	2726	0.066	1.897	-0.145	1.923	0.241	1.857	0.096	1.931
Scholastic	2892	-0.001	0.673	0.066	0.677	-0.056	0.665	0.025	0.672
College & Labor Mkt (Wave 4, %)									
Any College	2958	0.622	0.485	0.561	0.496	0.673	0.469	0.617	0.486
Active	2626	0.863	0.344	0.886	0.318	0.845	0.362	0.884	0.320
Employed	2266	0.946	0.226	0.936	0.245	0.954	0.209	0.951	0.215
Full Time	2626	0.626	0.484	0.757	0.429	0.524	0.500	0.661	0.474
Inc. Empl	1934	273.9	118.8	299.6	129.2	252.2	104.6	284.0	118.9
Inc. Full Time	1502	292.6	115.4	310.9	123.9	271.2	100.5	302.0	114.7
Non-heterosexuals (Wave 4, %)									
Sexual Interest	644	0.220	0.414	0.114	0.318	0.307	0.461		
Self-perception	591	0.202	0.401	0.129	0.336	0.261	0.440		
Same-gender Sex	350	0.119	0.324	0.102	0.302	0.133	0.340		

Note: *University* (*universitetet*) includes programs of four or more years. *3yr College* corresponds to the *høyskole*. *Both Parents* equals to 1 if the person—at wave 1—lived with both parents. *High SES* describes parents with a high administrative or academic job. *Soc. Accept.* (social acceptance), *Sympathetic* (capable of building close friendships), *Self-esteem*, *Leadership* and *Scholastic Competence* are measures that come from SPPA scales (Harter, 1988). *Grades* correspond to a demeaned aggregation of school grades (Math, Norwegian and English). *Any College* equals to 1 for those with any kind of tertiary education in colleges (*høyskole*) or universities. *Employed* presents the proportion of people employed out of the whole sample, and not out of only those active. The same holds for *Full Time*. *Inc. Empl.* presents the pre-tax income of people employed for those who report such income to be greater than zero. The same holds for *Inc. Full Time*. *Sexual Interest* equals to 1 if attracted in any degree to people of the same gender. *Self-perception* equals to 1 for individuals who consider them homosexuals in any degree. *Same-gender sex* corresponds to whether a person has had sex with someone from the same gender. Income is in levels. *Source*: Own calculations. *Source*: Own calculations.

already graduated. Thus, the younger cohorts answered wave 2 in school (n=3,844), while the older ones were contacted through the mail. The study was originally planned to end after wave 2. However, a spin-off of the study continued in 1999 but only for the younger cohorts, those who had answered the wave 2 questionnaire in school and not by mail. 2,923 people were interviewed. Wave 4 was collected between May 2005 and August 2006, with the target population being the same as in wave 3. In this wave, data were collected for 2,890 people. There is a great overlap between those wave 3 and wave 4 respondents; in total, 2,562 responded both surveys. Following advice from the experts at NOVA, the agency responsible for the data, I do not include in my analysis respondents that were not followed up in waves 3 or 4. This leaves a final sample of 2,991.

2.2 Sexual Minorities' Rights in Norway

Norway leads European nations in several rankings measuring the protection of rights for sexual minorities (ILGA Europe, 2019). With the enactment of a 1981 law, Norway became the first country in the world to ban discrimination based on sexual orientation in the provision of goods and services (Hennum, 2001). As early as 1993, same-sex couples were given equal rights to heterosexual couples with the exception of the adoption of children. Adoption of step-children was allowed in 2002 and adoption of children in general became legal in 2009 (Anderssen and Hellesund, 2009). Advocacy groups consider Norway the best European country for workers that belong to a sexual minority. They do so on the basis of having a strong legal framework that protects minority rights and a high tolerance for sexual minorities within the society in conjunction with low unemployment rates and high wages in the general population. Thus, any evidence of labor market disparities I find for young Norwegians in 2006 could be considered as a lower bound for other countries and societies that find themselves behind Norway in the path towards equality.

2.3 Definition of Homosexuality According to the Data

Defining homosexuality is a difficult task. Hetero/homosexuality is not a dichotomous indicator but a spectrum of preferences that go from being attracted or having had sex with people from the same gender to being attracted or having had sex with people from the opposite gender, with a complex assortment of sexual preferences in between these two extremes (Laumann et al., 1994).[9] Defining homosexuality according to the data becomes even more complex as it can be drawn from three different concepts (Laumann et al., 1994; Badgett, 1995) . First, the extent to which each gender generates sexual interest (i.e., sexually attracted to, sexual fantasies about) to the respondent. Second, whether the respondent considers herself homosexual to a particular extent. And third, the gender of the respondent's sexual partners. It should be noted that one definition does not imply any of the other two, particularly for individuals of the homosexual or heterosexual extremes.

YiN inquires about the respondents' homosexuality under the three different definitions, and it does so by using a rich set of possible categories to fully capture the wide range of sexual preferences under each definition.[10] Table 2 presents the distribution of sexual preferences according to each definition. The table is divided in two panels. The top panel presents the tabulations between the self-perception and the sexual-interest scales for those who have not had sexual relations with a same-sex partner. The bottom panel does the same but for those who have had sexual relations with a same-sex partner. The first thing to note is that although 67.7%(=76.94%*88.1%) of the respondents find themselves in the heterosexual extreme of the spectrum, the remaining 36.3% are scattered across the different intermediate categories. This lends support to the conceptual

[9]That is why in this paper I sometimes use the term "nonheterosexual" to refer to people whose sexual preferences are away from the heterosexual extreme but anywhere else in the spectrum of preferences.

[10]The exact questions and possible answers can be found in Appendix A.

Table 2: Distribution of Hetero/Homosexuality Scales by Definition

	Sexual Interest Scale							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**Total**
Self-rate Scale	%	%	%	%	%	%	%	%
No Same-gender Sexual Relations (88.1%)								
1	76.94	7.43	0.63	0.04	0.00	0.00	0.16	85.19
2	1.41	4.75	0.75	0.08	0.00	0.00	0.00	6.99
3	0.04	0.35	0.47	0.08	0.00	0.00	0.00	0.94
4	0.00	0.00	0.00	0.12	0.04	0.00	0.00	0.16
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	5.85	0.75	0.08	0.04	0.00	0.00	0.00	6.72
Total	84.24	13.28	1.93	0.35	0.04	0.00	0.16	100.00
At Least One Same-gender Sexual Relation (11.9%)								
1	26.65	12.03	1.15	0.29	0.00	0.00	0.00	40.11
2	2.29	16.62	5.73	0.57	0.00	0.00	0.00	25.21
3	0.00	2.87	8.31	0.57	0.29	0.00	0.00	12.03
4	0.00	0.57	0.29	1.15	0.29	0.00	0.00	2.29
5	0.00	0.00	0.00	0.00	1.15	0.29	0.00	1.43
6	0.00	0.00	0.00	0.00	0.29	3.15	0.29	3.72
7	2.58	2.01	0.29	0.29	0.00	1.15	8.88	15.19
Total	31.52	34.10	15.76	2.87	2.01	4.58	9.17	100.00

Note: *Sexual Interest Scale* responds to the question "Are you sexually interested in men or women (sexually attracted to, sexual fantasies about)?". The scale is increasing in the sense that men aswering "women only" and women answering "men only" have a 1, and women aswering "women only" and men answering "men only" have a 7. See intermediate categories in Appendix A. *Self-rate Scale* responds to the question "How would you rate youself on a scale from absolutely heterosexual to absolutely homosexual/lesbian"? The scale is increasing in the sense that answers "Only heterosexual" have a 1, and "Only homosexual/lesbian" have a 7. See intermediate categories in Appendix A. *Source*: Own calculations.

framework that considers sexuality as a non-dichotomous trait,; a continuum of degrees of preferences that can be manifested in three dimensions that although positively correlated—correlation between the SI and SP scales among men is 0.64, among women 0.28—are not bounded together.

Table 2 shows that discretizing the sexual preferences measures requires the imposition of *ad hoc* thresholds, one for each definition. One way to use the data is to build three dichotomous variables that split the sample between heterosexual and non-heterosexual respondents. One can construct the first measure according to sexual interests (SI) and considers to be homosexual any individual who is attracted in any degree to people of the same sex. A second measure could be

constructed according to self-perception of homosexuality (SP) and considers as homosexual any individual who believes she is homosexual in any degree. Finally, a third measure (SX) responds to whether a person has had sex with someone from the same gender. Note that in the interest of sample size, in creating SI, SP and SX, I imposed the thresholds so as to define as homosexual anyone that claims to be nonheterosexual in any degree.

The bottom panel of Table 1 shows the proportion of respondents who are classified as homosexual according to each definition. Interestingly, for every definition women are more likely to be classified as homosexuals, but the difference in the number of men and women identified as homosexual is drastically smaller in the SX definition than in the ones that use self-perception or sexual interest.

Discretizing the measures of sexual preferences is arbitrary and results in some loss of information. To overcome this issue, this paper also incorporates the treatment of homosexuality as a continuous latent variable that can be measured through three manifest scores. Thus, I allow people's *degree of homosexuality* to be an unobserved continuum that affects the answers given to each question regarding sexual preferences. That way, I can reconcile the fact that people may be *less heterosexual* according to one score but *more heterosexual* according to other. That is because the latent *degree of homosexuality* could, for instance, be enough to be manifested in one score, but not in others. This approach, explored in Section 5, also allows the scores to be affected by family contexts that could influence the likelihood of someone *manifesting* a degree of nonheterosexuality.

2.4 Skills

The empirical strategy used in this paper relies on the identification of unobserved skills. Such identification and the subsequent estimations based on this unobserved heterogeneity pose heavy

requirements on sample size. The number of observations required increases geometrically with the number of unobserved dimensions of skill to be estimated. The fact that the sample size of homosexuals in every definition is not large prevents me from estimating models with more than one unobserved skills factor. This factor will comprise both cognitive and noncognitive skills in one construct. Thus, it will collect the common variation between the two dimensions. Although, labor market could reward differently cognitive and non-cognitive skills, in this paper, I cannot inquire about those differences. Instead, I rely on the fact that cognitive and non-cognitive skills are positively and strongly correlated (Heckman and Masterov, 2007; Sarzosa, 2015).

To identify the unobserved skills factor, I use six scores; all of them were collected in the first wave of the survey, when the respondents were in their early to mid teens. The first four measures come from the Self-Perception Profile for Adolescents (SPPA) (Harter, 1988). These are self evaluations that take into account the perceived competence and the relative importance the respondent gives to specific areas of life (Wichstraum, 1995). First, two of the measures relate to agreeableness, one in the area of social acceptance—being able to get others to accept and like oneself—and another one measuring how easy it is for the respondent to build close social relationships. The third measure relates to self-esteem; similar to the widely used Rosenberg self-esteem scale, but its wording is more appropriate to adolescents. The fourth score measures scholastic competence, understood as the perceived cognitive competence related to schoolwork (Harter, 2012). The fifth score relies on a scale that measures qualities like leadership and strong-mindedness—makes decisions easily, is independent, willing to take a stand. Finally, the sixth scale uses junior high school grades in Math, Norwegian, and English. Unfortunately, grades in the Norwegian education system follow a scale with only five categories. Thus, I aggregate the grades of the three subjects in order to have enough smoothness and continuity in the manifest

variable that enters the measurement system.

2.5 Income Measures

The YiN data set is was not conceived as a labor force survey and thus does not contain a measure of hourly wage. Instead, it asks about pretax personal yearly income. In addition, time worked is not recorded. Hence, halftime workers could be earning less just because they are working less hours. To overcome this, I compare the results I obtain on the whole sample with those I obtain on the sample of full-time workers, for whom the income reported is more likely to come from their work.[11] Note that income measures were collected when the cohort in the sample was around 30 years old. Thus, this study focuses on labor market discrimination at relatively early stages in their career. However, evidence shows that wage gaps in other contexts widen with age (Manning and Swaffield, 2008; Bertrand et al., 2010). Hence, finding gaps at early ages is indicative of gaps later in life.[12]

2.6 The Income Gap

The typical estimation found in the literature on income discrimination for sexual minorities relies on a linear model for labor market outcomes of the form

$$Y = \alpha + \gamma Gay + \beta \mathbf{X} + \sum_{s=1}^{S} \phi_s D_s + \varepsilon \tag{1}$$

[11] In addition, there is a chance part of the income comes from rehabilitation, disability, or unemployment benefits. However, I find that homosexuals are equally as likely to receive social benefits as heterosexuals, regardless of the definition of homosexuality. Therefore, there is no reason to believe that any leakage in the income measures constructed could bias the results in either direction.

[12] In fact, Weinberger and Kuhn (2010) suggest that the reduction of the gender wage gap observed in the last half of the twentieth century responds more to disappearing differentials in the factors that determine initial wage than changes in relative wage growth after labor market entry.

where Y is the income measure of interest, $\mathbf{X}$ is a set of observable controls (e.g., age, experience, geographical region), and D_s is a dummy variable that represents the schooling level where $s = \{0, \ldots, S\}$. To capture differences between the effect for lesbians from the effect of homosexual men, Equation (1) is often augmented with an interaction between Gay and gender, or the estimations are done separately for each gender. In essence, the parameter of interest of these types of models is γ, which the existing literature interprets as the difference in income between two individuals that have the same observable characteristics and have achieved the same education levels but differ in their sexual preferences.

However, the estimation of (1) requires substantial econometric considerations. The most important one is that unobservable characteristics may influence simultaneously schooling or occupational decisions and income, and those choices themselves can be influenced by the sexual preference of the individual. Models like (1) also assume that ϕ_s does not differ across sexual preferences. Although, we are not interested in the estimates of ϕ_s, this assumption does affect the estimate of γ due to the endogeneity issues just described.

For comparison purposes, Panel A in Table 3 presents the reduced form results of models of the type described by Equation (1) for each definition of homosexuality. Column 1 presents the unconditional difference in means. Employed Norwegian homosexuals, on average, earn between 8.7 (=23.76/273.9) and 11.4 (=31.17//273.9) percent less than heterosexuals. Adding age and gender controls in Model 2 reduces this gap by about half. However, the specification that most resembles the one used in the existing literature is that of Model 3 with the introduction of a dummy variable that takes the value of 1 if the respondent had any kind of tertiary education. Also, it incorporates an interaction term between gay and female to identify a differential effect on lesbians. The income gap among men in this specification ranges from 9.5 percent to 12.3 percent

Table 3: Gap in Yearly Income

	Model 1		Model 2		Model 3		Model 4	
Definition	Coeff.	Std.Err.	Coeff.	Std.Err.	Coeff.	Std.Err.	Coeff.	Std.Err.
Panel A: Dichotomous Variables								
SI	-23.76***	(6.66)	-10.85*	(6.56)	-33.11***	(12.68)	-31.19**	(13.25)
SI×*female*					30.08**	(14.81)	26.83*	(15.34)
Joint Signif.					-3.03	(7.66)	-4.36	(7.77)
SP	-31.17***	(6.92)	-17.88***	(6.72)	-33.67***	(11.90)	-30.09**	(12.53)
SP×*female*					23.23	(14.41)	21.24	(15.04)
Joint Signif.					-10.44	(8.14)	-8.85	(8.35)
SX	-25.05***	(8.52)	-18.94**	(8.15)	-25.98**	(12.87)	-24.52*	(13.35)
SX×*female*					11.79	(16.65)	11.03	(17.11)
Joint Signif.					-14.18	(10.54)	-13.49	(10.73)
Panel B: Combined Index								
Homosex. Category								
2.	-24.45***	(9.23)	-1.12	(9.34)	-0.60	(9.46)	-2.60	(9.52)
3.	-47.42***	(11.71)	-15.58	(11.64)	-13.99	(11.75)	-10.22	(12.22)
4.	-35.51***	(10.64)	-37.63***	(10.22)	-46.64***	(12.76)	-47.71***	(13.37)
5.	-43.88***	(13.02)	-34.87***	(12.51)	-46.11***	(17.08)	-39.10**	(17.79)
6.	-5.28	(14.25)	1.51	(13.69)	21.83	(22.39)	32.06	(23.70)
Homosex. Category×*female*								
4.					24.52	(21.37)	26.34	(21.72)
5.					24.55	(25.01)	20.81	(25.98)
6.					-31.13	(28.31)	-38.68	(29.57)
Joint Signif.								
4.					-22.12	(17.14)	-21.38	(17.11)
5.					-21.56	(18.32)	-18.29	(18.97)
6.					-9.30	(17.32)	-6.62	(17.70)
Age & gender			X		X		X	
College					X			
Test Scores							X	

Note: *** p<0.01, ** p<0.05, * p<0.1. All estimations were drawn from different regressions with the dependent variable being the yearly income among the employed population in wave four of the survey (average 273.58). Age enters in a quadratic polynomial. The College variable takes the value of 1 when the person has undertaken any kind of tertiary education in colleges (*høyskole*) or universities (*universitetet*). This includes three- or four-year college studies or studies in the university with the same length. The estimations with test scores include six scores, one related to school grades (Math, Norwegian, and English), two of them relate to agreeableness and the ease to make friends, the others relate to positivism, leadership, and scholastic competence. All of them collected in the first wave of the survey. *Homosex. Category* is an index that combines the three homosexuality scales tabulated in Table 2. See footnote 14 for details. Number of observations in Model 1: 1,911, Model 2: 1,911, Model 3: 1,911, and Model 4: 1,810. *Source*: Own calculations.

depending on the homosexuality definition. Model 4 presents the estimation results of a model *a la* Neal and Johnson (1996). That is, introducing test scores as proxies of abilities instead of the schooling attainment.[13] The results are very similar to those obtained in Model 3. These gaps amount to half of what similar estimations have found in the U.S. labor market. This could respond to differences in the types of samples used in the analyses or to structural features in the Norwegian labor market *vis-a-vis* the United States. Interestingly, I find no income gaps among women. Furthermore, the mean comparisons of college attendance presented in Table J.1 in the Appendix J show that there seems to be no systematic evidence of either group being more likely to go to a tertiary education institution. If anything, these results suggest that lesbians are less likely to go to college than heterosexual women.

Panel B in Table 3 presents the estimates of a model like (1) where the dichotomous definitions of homosexuality have been replaced with an index combining the three dimensions of homosexuality.[14] The resulting index is categorical with people in higher categories reporting higher scores in the homosexuality scales. Each category, except the base one, accounts roughly for 5 percent of the population. The base category in the regressions comprises the two-thirds of the population that considered themselves heterosexual, who only have sexual interest in and have only had sexual relations with members of the opposite sex. The results show that once I introduce controls, the

[13] Neal and Johnson (1996)—when exploring racial wage gaps—argue that they can bypass the endogeneity between race and education attainment on the measurement of income gaps using the stock of skills accumulated up until mandatory schooling ends instead of controlling for schooling attainment *per se*. They argue that the educational and occupational choices made during adulthood that end up affecting income are themselves consequences of the stock of skills—proxied by test scores—agents had when they started making those decisions. Hence, this approach claims that, conditional on the skill levels, the income gap is fixed across educational levels or occupational sectors. As Urzua (2008) shows in the case of racial income gaps, I show that this approach is not appropriate to describe the existing labor market differences between homosexuals and heterosexuals, as the size of these gaps differ across educational choices.

[14] I create this index based on the three homosexuality scales tabulated in Table 2 using a principal components analysis. I take the first component and group its prediction in ventiles. Given that two-thirds of the sample are heterosexuals, they fill the first 14 ventiles. Ventiles 17 and 18 were merged because the latter had relatively few observations with income.

income gaps against men in categories 4 and 5 are larger than those against men in categories 2 and 3—which turn out to be non-statistically different from zero. These gaps among to 17 percent. I find no gaps among women. An interesting feature of Panel B in Table 3 is that those in the highest homosexuality category have no income differences with heterosexuals. I will return to this in the next subsection.

2.7 The Importance of Endogenous Selection into Employment

The findings in Table 3 are obtained from a selected sample—namely, those who are employed. There are many reasons to believe those employed are not representative of the whole sample, as they may have observable and unobservable characteristics that contribute to their sorting (or not) into employment (Mroz, 1987), including—probably—homosexuality. In fact, Panel A in Table 4 shows significant gaps in the probability of being employed against non-heterosexuals. The unconditional difference in employment probability is 7 percent greater among heterosexuals. The employment gap is greater among men. It ranges between 6.7 to 16.7 percent depending on the homosexuality definition and on the controls used. The gap among women is smaller but significant. It ranges between 4 and 6.4 percent.

Panel B in Table 4 considers homosexuality as a categorical index. It shows that selection into employment is strongly affected the degree of homosexuality: people in higher categories are less likely to be employed than those in lower categories. Those in the top category are 17 percent (men) and 10 percent (women) less likely to be employed than heterosexuals. This employability gap is twice as big as the one among people in categories 4 and 5 and helps explain the lack of an income gap against those in the top category (Panel B in Table 3). The employed among them are such a selected sample that only those with relatively high wages work, thus eliminating the

Table 4: Gap in Employment

	Model 1		Model 2		Model 3		Model 4	
Definition	Coeff.	Std.Err.	Coeff.	Std.Err.	Coeff.	Std.Err.	Coeff.	Std.Err.
Panel A: Dichotomous Variables								
SI	-0.087***	(0.02)	-0.083***	(0.02)	-0.150***	(0.03)	-0.167***	(0.04)
SI×*female*					0.098**	(0.04)	0.123***	(0.04)
Joint Signif.					-0.052**	(0.02)	-0.044**	(0.02)
SP	-0.071***	(0.02)	-0.061***	(0.02)	-0.098***	(0.03)	-0.103***	(0.04)
SP×*female*					0.053	(0.04)	0.062	(0.04)
Joint Signif.					-0.046**	(0.02)	-0.041*	(0.02)
SX	-0.070***	(0.02)	-0.065***	(0.02)	-0.067*	(0.04)	-0.089**	(0.04)
SX×*female*					0.002	(0.05)	0.050	(0.05)
Joint Signif.					-0.064**	(0.03)	-0.038	(0.03)
Panel B: Combined Index								
Homosex. Category								
2.	-0.052**	(0.026)	-0.049*	(0.027)	-0.038	(0.027)	-0.033	(0.028)
3.	-0.064**	(0.032)	-0.049	(0.033)	-0.056*	(0.034)	-0.050	(0.035)
4.	-0.053*	(0.030)	-0.045	(0.030)	-0.073**	(0.037)	-0.093**	(0.039)
5.	-0.093***	(0.036)	-0.080**	(0.036)	-0.076	(0.050)	-0.085	(0.053)
6.	-0.140***	(0.038)	-0.138***	(0.037)	-0.155**	(0.061)	-0.170***	(0.064)
Homosex. Category×*female*								
4.					0.094	(0.062)	0.122*	(0.063)
5.					-0.007	(0.071)	0.028	(0.074)
6.					0.022	(0.077)	0.065	(0.080)
Joint Signif.								
4.					0.021	(0.050)	0.029	(0.050)
5.					-0.083*	(0.050)	-0.057	(0.052)
6.					-0.133***	(0.047)	-0.104**	(0.049)
Age & gender			X		X		X	
College					X			
Test Scores							X	

Note: *** p<0.01, ** p<0.05, * p<0.1. All estimations were drawn from different regressions with the dependent variable takes the value of 1 if the person is employed and 0 otherwise. Age enters in a quadratic polynomial. The College variable takes the value of 1 when the person has undertaken any kind of tertiary education in colleges (*høyskole*) or universities (*universitetet*). This includes three- or four-year college studies or studies in the university with the same length. The estimations with test scores include six scores, one related to school grades (Math, Norwegian, and English), two of them relate to agreeableness and the ease to make friends, the others relate to positivism, leadership, and scholastic competence. All of them collected in the first wave of the survey. *Homosex. Category* is an index that combines the three homosexuality scales tabulated in Table 2. See footnote 14 for details. Number of observations in Model 1: 2,562, Model 2: 2,562, Model 3: 2,562, and Model 4: 2,428. *Source*: Own calculations.

income gap for those who work from that segment of the population.

Table 4 provides evidence on the fact that differences in the labor market against sexual minorities may take place at the time of getting a job and not entirely by working for a lower wage. This result will be fully explored in the subsequent chapters, paying especial attention to how endogenous choices that are critical in determining income—like employment and schooling—could be affected by differences in characteristics, both observable and unobservable, across sexual preference groups.

3 Empirical Strategy

3.1 Modeling Outcomes

Norway has a relatively high tertiary education enrollment rate; around 62 percent of people in the sample go to a tertiary education institution of some sort. A vast literature on returns to schooling shows that differences in access to tertiary education is critical in determining difference in earned income (Mincer, 1958; Becker and Chiswick, 1966; Card, 2001). Therefore, part of the sexual preference income gap could be explained by differences between homosexuals and heterosexuals in enrollment rates in tertiary education. In the same way, the differences shown in Section 2.7 between homosexuals and heterosexuals in the likelihood of being employed could determine part of the sexual preference income gap. Thus, I model the income observed in the YiN sample at their late 20s following a potential outcomes framework that endogenizes schooling and employment choices inspired by the Roy model (Roy, 1951).

An important feature of the Roy model is that individuals must choose between two sectors, for example, treated and not treated depending on their comparative advantage (Heckman and

Honore, 1990). Therefore, both the decision and the outcomes are endogenous based on observable and unobservable characteristics of the individuals. This feature is key in the development of counterfactuals in Section 4.2. Works like Murnane et al. (1995), Heckman et al. (2006) and Espinoza et al. (2014) show that unobserved heterogeneity (e.g., skills) is as relevant in determining life choices and adult outcomes as other characteristics like gender or household composition. Thus, I model the income Y, schooling D, and employment E (i.e., $E = 1$ if employed and $E = 0$ if not) as the results of observable exogenous or predetermined characteristics $\mathbf{X_Y}$, $\mathbf{X}_D$ and $\mathbf{X}_E$ (e.g., gender, tenure, local labor market) and the unobserved heterogeneity $\mathbf{\Theta}$, which in the case of this paper can contain up to two dimensions: skills and the latent degree of homosexuality. That is,

$$\mathbf{Y} = \begin{bmatrix} D \\ E_1 \\ Y_1 \\ E_0 \\ Y_0 \end{bmatrix} = \begin{bmatrix} \mathbb{1}\left[\mathbf{X}_D\beta^{Y_D} + \alpha^{Y_D}\mathbf{\Theta} + e^D > 0\right] \\ D\left(\mathbb{1}\left[\mathbf{X}_E\beta^{E_1} + \alpha^{E_1}\mathbf{\Theta} + e^{E_1} > 0\right]\right) \\ D\ E_1\left(\mathbf{X}_Y\beta^{Y_1} + \alpha^{Y_1}\mathbf{\Theta} + e^{Y_1}\right) \\ (1-D)\left(\mathbb{1}\left[\mathbf{X}_E\beta^{E_0} + \alpha^{E_0}\mathbf{\Theta} + e^{E_0} > 0\right]\right) \\ (1-D)\,E_0\left(\mathbf{X}_Y\beta^{Y_0} + \alpha^{Y_0}\mathbf{\Theta} + e^{Y_0}\right) \end{bmatrix} \tag{2}$$

where $\mathbb{1}[A]$ denotes an indicator function that takes a value of 1 if A is true, and $(e^D, e^{E_1}, e^{Y_1}, e^{E_0}, e^{Y_0})$ is a vector of error terms with distributions $f_{e^{y_m}}(\cdot)$—for $y^m = \{D, E_1, Y_1, E_0, Y_0\}$—that are orthogonal to each other once we control for $\mathbf{X}_{y^m}$ and the unobserved heterogeneity $\mathbf{\Theta}$.

The system for outcomes in Equation (2) is comprised by five rows, where one row contains a binary choice D (e.g., going to college $D = 1$ or not $D = 0$) and the other four contain the outcomes of interest that depend on such choice (e.g., income and employment for those that went to college (Y_1, E_1) and wage and employment for those that did not (Y_0, E_0)). Note that $\mathbf{X}_E$, $\mathbf{X}_D$

and $\mathbf{X}_Y$ could share elements. An exclusion restriction (i.e., a source of variation present in $\mathbf{X}_D$ or $\mathbf{X}_E$ but not in $\mathbf{X}_Y$) would secure identification of the model without the need to rely on its nonlinearity.

3.2 Identifying Unobserved Heterogeneity

Given that $\mathbf{\Theta}$ is unobserved, the estimation of (2) is not straightforward. To begin with, unobserved heterogeneity factors are not well defined entities with established scales or units of measurement. Instead they are latent variation that needs to be inferred from variation captured in manifest variables they affect (Bartholomew et al., 2011). The manifest variables are assumed to be the result of a linear production function of scores whose inputs are both observable characteristics and the unobserved heterogeneity of the form

$$\mathbf{T} = \mathbf{X}_T \beta^T + \alpha^{\mathbf{T}} \mathbf{\Theta} + \mathbf{e}^{\mathbf{T}} \tag{3}$$

where $\mathbf{T}$ is a $L \times 1$ vector of measurements (e.g., test scores, answers to the sexual preferences questionnaire), $\mathbf{X}_T$ is a matrix with all observable controls that affect each score, and $\mathbf{e}^{\mathbf{T}} \perp (\mathbf{\Theta}, \mathbf{X}_T)$ is a $L \times 1$ vector of mutually independent error terms that have associated distributions $f_{e^h}(\cdot)$ for every $h = 1, \ldots, L$. Hence, I consider the manifest measures we observe on the data ($\mathbf{T}$) and the true latent heterogeneity ($\mathbf{\Theta}$) as two different metrics. In particular, in the case of the answers to the sexual preferences questionnaire, I am able to estimate a latent factor clean from characteristics or contexts that may have affected the answers we do observe, but should not change the true underlying homosexuality.

I use measurement system (3) (i.e., the variation observed in the manifest variables) to identify

$f_\theta(\cdot)$, the underlying distribution of the unobserved factors. I use a two-component mixture of normals $\theta \sim \omega\mathcal{N}(\mu_1, \sigma_1^2) + (1-\omega)\mathcal{N}(\mu_2, \sigma_2^2)$ to estimate the true underlying distribution $f_\theta(\cdot)$ for each dimension of unobserved heterogeneity, in case they are nonnormal (Hansen et al., 2004; Heckman et al., 2006; Espinoza et al., 2014).[15] The estimation of the parameters that govern $f_\theta(\cdot)$ is crucial for the estimation the complete semistructural model because such distributions allow me to integrate away the latent heterogeneity in all the outcomes, choices, and scores associated with the model, while still being able to retrieve the coefficients related with it in every equation.

Identification of $f_\theta(\cdot)$ requires the identification of the loadings $\alpha^{\mathbf{T}}$. Carneiro et al. (2003) show that their identification requires three assumptions: first, that $L \geq 2k+1$, where k is the number of unobserved factors in the model; second, that in the case when $k > 1$, the latent factors are orthogonal to each other; and third, that we normalize one of the loadings associated to each factor to 1. To observe how the assumptions work for identification, take for instance the case when $k = 1$, requiring $L \geq 3$. Thus, system (3) would become

$$T_1 = \mathbf{X}_T\beta^1 + \alpha^1\theta + e^1$$

$$T_2 = \mathbf{X}_T\beta^2 + \alpha^2\theta + e^2$$

$$T_3 = \mathbf{X}_T\beta^3 + \theta + e^3$$

where $\alpha^3 = 1$ is the nummeraire. This implies that higher values of θ correspont to higher values of T_3. Loadings α^h for $h = 1, \ldots, L-1$ need not be positive. Note that $\frac{COV(T_1,T_2|\mathbf{X}_T)}{COV(T_1,T_3|\mathbf{X}_T)} = \alpha^2$ and $\frac{COV(T_1,T_2|\mathbf{X}_T)}{COV(T_2,T_3|\mathbf{X}_T)} = \alpha^1$. Then, having identified the loadings, we can write model (3) as $\mathbf{J} = \mathbf{\Theta} + \tilde{\mathbf{e}}^{\mathbf{T}}$

[15] The mixture of normals allows me to not only replicate a wide range of distributions but also integrate numerically using the Gauss-Hermite quadrature, which is particularly useful for calculating $\int f(X)\,dx$ when $X \sim \mathcal{N}(\mu, \sigma^2)$ (Judd, 1998). For a more detailed analysis regarding the estimation of the skills' distributions see Sarzosa and Urzua (2016).

where $\mathbf{J} = \frac{\mathbf{T} - \mathbf{X}_T \beta^T}{\alpha^T}$ and $\tilde{\mathbf{e}}^{\mathbf{T}} = \frac{\mathbf{e}^{\mathbf{T}}}{\alpha^T}$, and together with the fact that $\mathbf{e}^{\mathbf{T}} \perp \mathbf{\Theta}$ rely on the Kotlarski Theorem which shows that using the joint distribution of $(J_1, J_2, \ldots, J_L)$ one can nonparametrically identify the distributions of $\mathbf{\Theta}$ and $\tilde{\mathbf{e}}^{\mathbf{T}}$—and thus $\mathbf{e}^{\mathbf{T}}$—up to one normalization (Kotlarski, 1967).

The complete structure of the model, including the parameters that describe the distributions of the underlying factor $f_\theta(\cdot)$, is estimated using maximum likelihood estimation (MLE). The empirical strategy assumes that after controlling for the unobserved heterogeneity $\mathbf{\Theta}$, the remaining error terms are independent (i.e., $e^D \perp e^{Y_1} \perp e^{Y_0} \perp e^{E_1} \perp e^{E_0} \perp e^{T_1} \perp \cdots \perp e^{T_L}$) and thus can be modeled through independent contributions to the likelihood function. That is:

$$\mathcal{L} = \prod_{i=1}^{N} \int \left[\begin{array}{c} \left[\left(1 - f^D\left(\mathbf{x}_{iD}, D_i, \zeta\right)\right) \left[f_{e^{Y_0}}\left(\mathbf{x}_{iY}, Y_{i0}, \zeta\right) f_{e^{E_0}}\left(\mathbf{x}_{iE}, E_i, \zeta\right) \right]^{E_{i0}} \left(1 - f_{e^{E_0}}\left(\mathbf{x}_{iE}, E_i, \zeta\right)\right)^{1 - E_{i0}} \right]^{1 - D_i} \\ \left[f^D\left(\mathbf{x}_{iD}, D_i, \zeta\right) \left[f_{e^{Y_1}}\left(\mathbf{x}_{iY}, Y_{i1}, \zeta\right) f_{e^{E_1}}\left(\mathbf{x}_{iE}, E_i, \zeta\right) \right]^{E_{i1}} \left(1 - f_{e^{E_1}}\left(\mathbf{x}_{iE}, E_i, \zeta\right)\right)^{1 - E_{1i}} \right]^{D_i} \\ f_{e^1}\left(\mathbf{x}_{iT_1}, T_{i1}, \zeta\right) f_{e^2}\left(\mathbf{x}_{iT_2}, T_{i2}, \zeta\right) \times \cdots \times f_{e^L}\left(\mathbf{x}_{iT_L}, T_{iL}, \zeta\right) \end{array} \right] f_\theta(\zeta)\, d\zeta$$

4 Results: Homosexuality as Dichotomous Variables

4.1 Separate Distributions: Unobserved Abilities

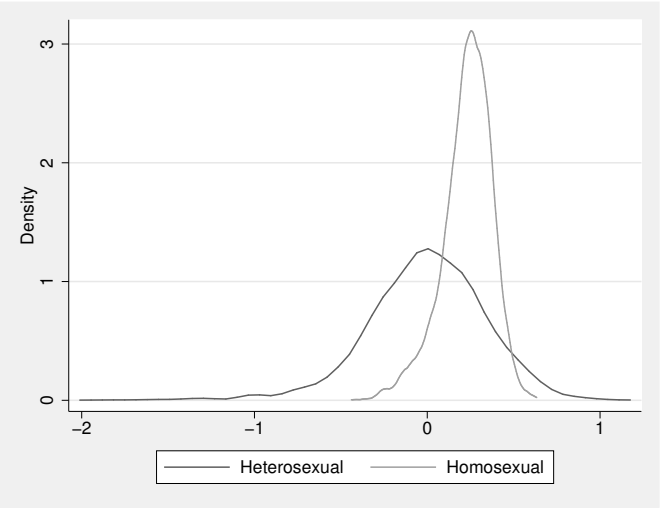
In this subsection, I present the results of estimating model (2)-(3) separately for heterosexuals and homosexuals. That way, I estimate a complete set of different parameters for each group, including the skills' distributions and the labor market returns to observable and unobservable characteristics. Thus, I allow for differences in characteristics and the returns to those characteristics.[16]

Figures 1 presents the estimated skill distributions for heterosexuals and homosexuals according

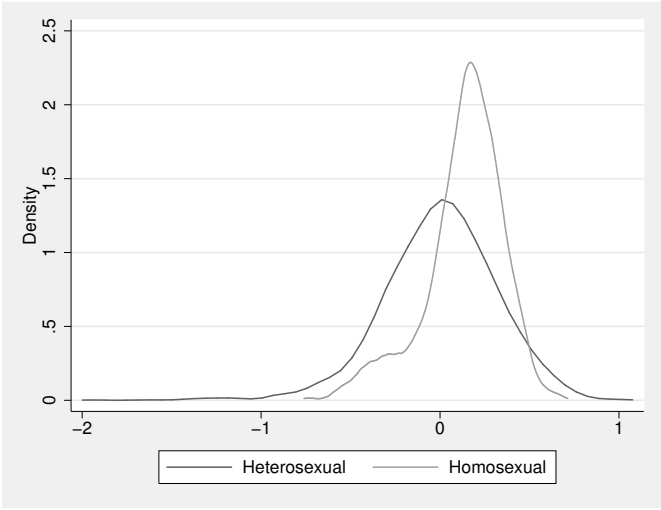
[16] One can estimate model (2)-(3) taking the population as a whole and capturing the effect of being part of a sexual minority by including an indicator variable in matrices $\mathbf{X}_D$, $\mathbf{X}_E$ and $\mathbf{X}_Y$. However, that approach—explored in Appendices B and C—restricts the estimates of the model to be the same for homosexuals and heterosexuals. If there are any differences in the skill distributions, or in the labor market returns to observable and unobservable characteristics, the results of such a model could be misleading.

Figure 1: Distribution of Unobserved Abilities by Sexual Preference

(a) According to SI Definition



(b) According to SP Definition



Note: Estimated from measurement systems presented in Tables H.2 to H.5 in Appendix H. Each estimation use the mixing of two normals. Left densities are estimated for those defined as homosexual and heterosexual according to the SI definition. The distributional parameters for the skills of the homosexuals are given by $\mu_1 = -0.153$, $\sigma_1 = 0.174$, $\mu_2 = 0.044$, $\sigma_2 = 0.107$, and the mixing probability is 0.223. The distributional parameters for the skills of the heterosexuals are given by $\mu_1 = -0.585$, $\sigma_1 = 0.483$, $\mu_2 = 0.032$, $\sigma_2 = 0.296$, and the mixing probability is 0.051. Right densities are estimated for those defined as homosexual and heterosexual according to the SP definition. The distributional parameters for the skills of the homosexuals are given by $\mu_1 = -0.450$, $\sigma_1 = 0.142$, $\mu_2 = 0.057$, $\sigma_2 = 0.157$, and the mixing probability is 0.113. The distributional parameters for the skills of the heterosexuals are given by $\mu_1 = -0.690$, $\sigma_1 = 0.450$, $\mu_2 = 0.024$, $\sigma_2 = 0.290$, and the mixing probability is 0.033. The mixing probability refers to the weight given to the second normal (i.e., the one described by parameters μ_2 and σ_2) in every mixture. The weight given to the first normal (i.e., the one described by parameters μ_1 and σ_1) in every mixture is given by one minus the mixing probability. *Source*: Own calculations.

to each definition of homosexuality.[17] The skill distributions of homosexuals are centered slightly to the right of heterosexuals' skill distributions.[18] Importantly, none of these differences in the means of skill distributions between homosexuals and heterosexuals are statistically significant. Skills among homosexuals are more tightly distributed. While the standard deviation of skills among

[17]Because of sample size issues I can only estimate this kind of models for homosexuals according to the SI and SP definitions of homosexuality.

[18]By construction, all the estimated distributions of unobserved ability are centered at zero. This comes from the fact that I estimate $f_\theta(\cdot)$ using a mixture of two normals that will be combined using weights ω. Given that all equations contain intercepts, the mean of the last normal μ_2 is one such that $\omega\mu_1 + (1-\omega)\mu_2 = 0$. That is, $\mu_2 = {}^{\omega\mu_1}/_{\omega-1}$. However, Urzua (2008) shows that with a minor linearity assumption, one can use the difference between the estimated constants of the test score equations that contain the nummeraire loadings to infer the difference in means of the skill distributions of two groups. Figure 1 is depicted in a way such that the centers of the skills distributions of homosexuals are shifted to represent the difference in means with respect to those of heterosexuals.

heterosexuals is around 0.33, it is between 0.12 to 0.22 for homosexuals. This interesting difference in the distribution of unobserved skills comes from the fact that the pair-wise correlations across manifest variables are lower among the non-heterosexual samples than among the heterosexual ones. It will have important consequences on schooling and occupational sorting and also in the income obtained when adults.

Effect of Skills on Outcomes. Table D.1 in the Appendix present the estimates of the sexual preference-specific coefficients associated with skills for income, employment and tertiary education. They show that the returns to skills do differ between heterosexuals and homosexuals, and that skills have significant positive returns only for those that undertook some sort of tertiary education. Skills have stronger effects on income for homosexuals than for heterosexuals. In particular, while a one standard deviation increase in skills among college-educated heterosexuals causes yearly income to increase by around 8.8 percent to 11.7 percent, the yearly income increase due to a one standard deviation increase in skills among college educated homosexuals ranges between 14 and 39 percent.[19]

Goodness of Fit of the Model. Tables D.2 and D.3 in Appendix C present evidence on the models' goodness-of-fit on the first and second moments for annual income, college enrollment and employment. They are a product of 40,000 simulations of the model based on as many different draws from the distributions of skills. The models do very well in predicting the means and standard deviations for each definition of income and homosexuality. The performance of the models predicting income for homosexuals is remarkable considering the reduced sample size the

[19]To obtain these figures, I take the loadings from Table D.1, normalize them using the standard deviation of the skills distributions, and divide them by the relevant yearly income mean that can be found in Table D.4. For instance, the returns to one SD of skills among college-educated heterosexuals under the definition SI is 76.439*0.33/287.018=8.8 percent.

models are facing.

4.2 Gaps in Labor Market Outcomes

Given that the model does a good job fitting the data, one can use it to simulate the outcomes that homosexuals would have obtained if their skills came from the heterosexual skills distribution and/or if they had the observable characteristics heterosexuals have.[20] This way, I can tease out the differences in labor market outcomes that are due to differences in the skills available across the groups from the differences that are due to plain discrimination. It is evident from Figure 1 that the distribution of skills among homosexuals has a higher mean, but a lower variance than that of heterosexuals.[21] This makes the counterfactual exercise even more compelling: it is not clear the direction outcomes will change—or if they will change at all—as there will be on average a lower stock of skills, but at the same time, there will be people with substantially higher and lower stocks of skills than before.

Let

$$Y^H\left(\mathbf{X}_Y^{H_X}, \mathbf{X}_E^{H_X}, \mathbf{X}_D^{H_X}, \theta^{H_\theta}\right) = D^H\left(\mathbf{X}_D^{H_X}, \theta^{H_\theta}\right) E_1^H\left(\mathbf{X}_E^{H_X}, \theta^{H_\theta}\right) Y_1^H\left(\mathbf{X}_Y^{H_X}, \theta^{H_\theta}\right) + \left[1 - D^H\left(\mathbf{X}_D^{H_X}, \theta^{H_\theta}\right)\right] E_0^H\left(\mathbf{X}_E^{H_X}, \theta^{H_\theta}\right) Y_0^H\left(\mathbf{X}_Y^{H_X}, \theta^{H_\theta}\right)$$

be the income given by observable characteristics $\mathbf{X}_Y^{H_X}$, $\mathbf{X}_E^{H_X}$ and $\mathbf{X}_D^{H_X}$ and unobservable skills θ^{H_θ}. The supra-index denotes the sexual preference group (i.e., $H_\cdot = \{S, G\}$, where the elements labeled with S correspond to heterosexuals and the ones labeled with G correspond to homosexuals).

[20] See the differences in observable characteristics across groups in Table J.3 in Appendix G.

[21] Why the moments of the distributions of skills would differ is outside the scope of this paper. It may be due to the different contexts homosexuals have to face while growing up, being more likely to be bullied or harassed, which end up affecting the stock of skills with which they reach adulthood (Sarzosa, 2015).

Hence, for instance, income for heterosexuals is given by $Y^S\left(\mathbf{X}_Y^S, \mathbf{X}_E^S, \mathbf{X}_D^S, \theta^S\right)$, while income for homosexuals is given by $Y^G\left(\mathbf{X}_Y^G, \mathbf{X}_E^G, \mathbf{X}_D^G, \theta^G\right)$. These are the values presented in Tables D.2 and D.3. I can use this notation to identify the simulation of counterfactuals. For instance, $Y^G\left(\mathbf{X}_Y^G, \mathbf{X}_E^G, \mathbf{X}_D^G, \theta^S\right)$ denotes the simulated income earned by homosexuals when given skills drawn from the heterosexuals' skills distribution but keeping the observable characteristics as they are. Consequently, the simulated income earned by homosexuals when given the skills and observable characteristics of heterosexuals is denoted by $Y^G\left(\mathbf{X}_Y^S, \mathbf{X}_E^S, \mathbf{X}_D^S, \theta^S\right)$.

The fact that choices $D^H\left(\mathbf{X}_D^{H_X}, \theta^{H_\theta}\right)$ and $E_D^H\left(\mathbf{X}_E^{H_X}, \theta^{H_\theta}\right)$ are endogenous is an important feature of the model that affects the simulation of counterfactuals. That is, when counterfactuals are simulated, their new results are a product not only of direct changes in the outcome equations $Y_1^H\left(\mathbf{X}_Y^{H_X}, \theta^{H_\theta}\right)$ and $Y_0^H\left(\mathbf{X}_Y^{H_X}, \theta^{H_\theta}\right)$ but also of indirect changes through the schooling and employment decisions. For instance, when simulating the model for homosexuals but giving them skills drawn from the heterosexuals's distribution, the choices become $D^G\left(\mathbf{X}_D^G, \theta^S\right)$, $E_0^G\left(\mathbf{X}_E^G, \theta^S\right)$ and $E_1^G\left(\mathbf{X}_E^G, \theta^S\right)$. Thus, education and employment decisions are taken based on observable and unobservable characteristics, and once these change, people might choose differently. Hence, the measure of the total income gaps between the two groups need to incorporate the differences that arise from both the direct and indirect channels. That is, the different gaps measured—depending on the counterfactual used—are provided by

$$
\begin{aligned}
&E\left[Y^S\left(\mathbf{X}_Y^S, \mathbf{X}_E^S, \mathbf{X}_D^S, \theta^S\right)\right] - E\left[Y^G\left(\mathbf{X}_Y^{H_X}, \mathbf{X}_E^{H_X}, \mathbf{X}_D^{H_X}, \theta^{H_\theta}\right)\right] \\
&\quad = \Pr\left[D^S\left(\mathbf{X}_D^S, \theta^S\right) = 1\right] \Pr\left[E_1^S\left(\mathbf{X}_E^S, \theta^S\right) = 1 \left| D^S\left(\mathbf{X}_D^S, \theta^S\right) = 1\right.\right] E\left[Y_1^S\left(\mathbf{X}_Y^S, \theta^S\right) \left| D^S\left(\mathbf{X}_D^S, \theta^S\right) = 1\right.\right] \\
&\quad + \Pr\left[D^S\left(\mathbf{X}_D^S, \theta^S\right) = 0\right] \Pr\left[E_0^S\left(\mathbf{X}_E^S, \theta^S\right) = 1 \left| D^S\left(\mathbf{X}_D^S, \theta^S\right) = 0\right.\right] E\left[Y_0^S\left(\mathbf{X}_Y^S, \theta^S\right) \left| D^S\left(\mathbf{X}_D^S, \theta^S\right) = 0\right.\right] \\
&- \Pr\left[D^G\left(\mathbf{X}_D^{H_X}, \theta^{H_\theta}\right) = 1\right] \Pr\left[E_1^G\left(\mathbf{X}_E^{H_X}, \theta^{H_\theta}\right) = 1 \left| D^G\left(\mathbf{X}_D^{H_X}, \theta^{H_\theta}\right) = 1\right.\right] E\left[Y_1^G\left(\mathbf{X}_Y^{H_X}, \theta^{H_\theta}\right) \left| D^G\left(\mathbf{X}_D^{H_X}, \theta^{H_\theta}\right) = 1\right.\right] \\
&- \Pr\left[D^G\left(\mathbf{X}_D^{H_X}, \theta^{H_\theta}\right) = 0\right] \Pr\left[E_0^G\left(\mathbf{X}_E^{H_X}, \theta^{H_\theta}\right) = 1 \left| D^G\left(\mathbf{X}_D^{H_X}, \theta^{H_\theta}\right) = 0\right.\right] E\left[Y_0^G\left(\mathbf{X}_Y^{H_X}, \theta^{H_\theta}\right) \left| D^G\left(\mathbf{X}_D^{H_X}, \theta^{H_\theta}\right) = 0\right.\right]
\end{aligned}
\tag{4}
$$

Notice that differences in schooling or employment choices across groups (i.e., $\Pr\left[D^S\left(\mathbf{X}_D^S, \theta^S\right) = 1\right]$ versus $\Pr\left[D^G\left(\mathbf{X}_D^{H_X}, \theta^{H_\theta}\right) = 1\right]$ and $\Pr\left[E_D^S\left(\mathbf{X}_E^S, \theta^S\right) \middle| D^S\left(\mathbf{X}_D^S, \theta^S\right) = D\right]$ versus

$\Pr\left[E_D^G\left(\mathbf{X}_E^{H_X}, \theta^{H_\theta}\right) \middle| D^G\left(\mathbf{X}_D^{H_X}, \theta^{H_\theta}\right) = D\right]$) are important in determining the overall income gap because they mediate the way income differences between homosexuals and heterosexuals for each level of education or employment choice enter in the overall gap calculation.

Table D.4 in the Appendix presents the results using two counterfactual simulations on college enrollment, employment, and yearly income.[22] For reference, I present the fitted values of the model for heterosexuals (i.e., $\Pr\left[D^S\left(\mathbf{X}_D^S, \theta^S\right) = 1\right]$, $\Pr\left[E^S\left(\mathbf{X}_E^S, \theta^S\right) = 1\right]$ and $E\left[Y^S\left(\mathbf{X}_Y^S, \theta^S\right)\right]$). I present both the censored and the uncensored income means. With the label *basic*, I present the fitted values for the homosexual sample keeping their own characteristics (i.e., $\Pr\left[D^G\left(\mathbf{X}_D^G, \theta^G\right) = 1\right]$, $\Pr\left[E^G\left(\mathbf{X}_E^G, \theta^G\right) = 1\right]$ and $E\left[Y^G\left(\mathbf{X}_Y^G, \theta^G\right)\right]$). Under the label *Equalizing Gender*, I present the simulations correcting for the fact that there are more women than men defined as homosexuals (see Table 1), and if women earn consistently less than men, the gaps I calculate could be confounding the effect of belonging to a sexual minority with the effect of being a woman. I shut down the gender wedge by giving every observation the average of the sample's gender distribution. That way, I prevent gender discrimination biasing the comparisons. Under the label *Het. Unobs*, I present the simulated college enrollment, employment, and income means of homosexuals when the gender distribution and the unobservables are taken from the heterosexual sample (i.e., $\Pr\left[D^G\left(\mathbf{X}_D^G, \theta^S\right) = 1\right]$, $\Pr\left[E^G\left(\mathbf{X}_E^G, \theta^S\right) = 1\right]$ and $E\left[Y^G\left(\mathbf{X}_Y^G, \theta^S\right)\right]$).

Table 5 presents the estimated gaps between heterosexuals and homosexuals under the counterfactuals described above and using the intuition of Equation (4). Table 5 shows that relative to heterosexuals, homosexuals are 6.5 percent more likely to achieve tertiary education, they are

[22] Results using full time employment are available upon request.

Table 5: Gap of Annual Income and College Enrollment of Homosexuals Under Different Assumptions

	College Enrollment		
	Homosexuals		
	Basic	*Eq Gender*	*Het. Unobs*
SI	-0.065***	-0.056**	0.041*
	(0.022)	(0.023)	(0.022)
SP	0.003	0.008	0.020
	(0.024)	(0.024)	(0.022)

Income and Labor Market Participation

SI

	No College			College			Overall		
Hom.	Empl.	CensY	Y	Empl.	CensY	Y	Empl.	CensY	Y
Basic	0.123***	0.085**	0.215***	0.100***	0.102***	0.215***	0.110***	0.095***	0.216***
	(0.039)	(0.043)	(0.061)	(0.026)	(0.026)	(0.041)	(0.014)	(0.022)	(0.034)
Eq Gender	0.096**	0.010	0.118*	0.138***	0.089***	0.248***	0.129***	0.062***	0.209***
	(0.038)	(0.044)	(0.063)	(0.027)	(0.026)	(0.041)	(0.014)	(0.022)	(0.035)
Het. Unobs	0.301***	0.014	0.350***	0.131***	0.011	0.166***	0.194***	-0.010	0.230***
	(0.041)	(0.046)	(0.063)	(0.026)	(0.028)	(0.043)	(0.014)	(0.024)	(0.036)

SP

	No College			College			Overall		
Hom.	Empl.	CensY	Y	Empl.	CensY	Y	Empl.	CensY	Y
Basic	0.115***	0.060	0.187***	0.095***	0.161***	0.261***	0.102***	0.121***	0.230***
	(0.038)	(0.040)	(0.059)	(0.028)	(0.028)	(0.043)	(0.015)	(0.023)	(0.035)
Eq Gender	0.110***	0.000	0.127**	0.118***	0.147***	0.274***	0.115***	0.090***	0.217***
	(0.038)	(0.040)	(0.060)	(0.029)	(0.029)	(0.044)	(0.015)	(0.024)	(0.036)
Het. Unobs	0.117***	0.012	0.147**	0.098***	0.002	0.107**	0.103***	-0.041	0.072*
	(0.038)	(0.041)	(0.060)	(0.027)	(0.032)	(0.049)	(0.013)	(0.026)	(0.039)

Note: *** p<0.01, ** p<0.05, * p<0.1. Standard errors in parentheses. The values presented are result of 40,000 simulations for each *basic* model and 200,000 simulations for each counterfactual model. *Empl.* stands for probability of being employed. *CensY.* is the censored income. Y stands for the uncensored income (i.e., including the zeros of the unemployed). *Het.* presents the fitted values of the model for heterosexuals (i.e., $\Pr\left[D^S\left(\mathbf{X}_D^S, \theta^S\right) = 1\right]$, $\Pr\left[E^S\left(\mathbf{X}_E^S, \theta^S\right) = 1\right]$ and $E\left[Y^S\left(\mathbf{X}_Y^S, \theta^S\right)\right]$). The label *basic* presents the fitted values for the homosexual sample with their own characteristics (i.e., $\Pr\left[D^G\left(\mathbf{X}_D^G, \theta^G\right) = 1\right]$, $\Pr\left[E^G\left(\mathbf{X}_E^G, \theta^G\right) = 1\right]$ and $E\left[Y^G\left(\mathbf{X}_Y^G, \theta^G\right)\right]$). *Eq Gender* presents a counterfactual in which the gender proportion in the homosexuals' samples is equated to the one in the heterosexuals' samples. Label *Het. Unobs* presents the simulated means of homosexuals when the gender distribution and the unobservables are taken from the heterosexual sample (i.e., $\Pr\left[D^G\left(\mathbf{X}_D^G, \theta^S\right) = 1\right]$, $\Pr\left[E^G\left(\mathbf{X}_E^G, \theta^S\right) = 1\right]$ and $E\left[Y^G\left(\mathbf{X}_Y^G, \theta^S\right)\right]$). *Source*: Own calculations.

10 percentage points less likely to be employed, and once employed they earn 9.45 percent (SI) and 12 percent (SP) less.[23] Going into detail, the results show that the income gaps are larger among college-educated workers, who earn 10.2 percent (SI) and 16 percent (SP) less than comparable heterosexuals. Table 5 also shows that the only effect equalizing the gender distribution in the sexual minority groups is the elimination of the earning gaps among workers with no college education.

When I equate skills distributions and gender rations across sexual preferences, income gaps disappear for those employed. The effect is evident regardless of the tertiary education choice. This is due to the high returns skills have among homosexuals, and it happens despite the fact that in this scenario, homosexuals would be 4.1 percent less likely to go to college than heterosexuals. However, homosexuals remain less likely to be employed by 19 and 10 percentage points for the SI and SP definitions, respectively.

The results of the last counterfactual simulation indicate that there is no evidence of income disparities among people that are working. However, becoming employed is the difficult part for homosexuals. The counterfactual simulations show that there is a 10.3–19.4 percentage point difference in the probability of being employed that cannot be explained due to differences of skills and gender composition. That gap is discrimination. And this is where the flexibility of the model with endogenous choices helps interpreting how the mechanisms that create the gaps work. Due to discrimination, the most highly skilled will be able to sort into employment. Hence, the ones who end up employed are a very select group who face high returns to their high stocks of skills. That way, I can reconcile the fact that, in the last counterfactual, homosexuals were provided with

[23]Figures obtained from dividing the gap calculated on the censored income by the mean income of heterosexuals under each definition. For instance, the proportional gap under the SI definition is given by 9.45 percent=26.848/283.99.

a distribution of skills that had greater variance with the elimination of the gap among those who managed to get a job. The barrier homosexuals face is getting a job.

5 Homosexuality as a Latent Continuous Variable

5.1 Identification of the latent homosexuality factor

Homosexuality is not a dichotomous variable, but a continuum of sexual preferences and perceptions. To incorporate that definition, I develop a model in which homosexuality is an additional source of unobserved heterogeneity. Such an approach is particularly well suited in this case as homosexuality is often veiled and only in some cases it is openly manifested or inferred by others. In that sense, homosexuality is a continuous characteristic—unobserved to the econometrician—that influence the answers given to the the sexual preferences/perceptions questionnaire, resembling the relation between latent factors and manifest scores presented in Section 3.2. That is, the model allows for a distinction between the answers given and the underlying homosexuality condition. Latent factor models allows pro- or antigay contexts in which the respondent lives to mediate the *answer*s to the questionnaire, but not shift the true underlying homosexuality. In addition, this approach, as it does not require splitting the sample into majority and minority groups, allows me to inquire about gender-specific disparities against homosexuals. Therefore, I identify gender-specific unobserved homosexual heterogeneity distributions.

I use the reported answers to the questions on sexual preferences, sexual interest, and gender of sexual partners as three manifest scores that are affected by the true latent homosexuality ϖ

and societal, household or personal or contexts $\mathbf{X}_S$

$$
\begin{aligned}
S_{SX} &= \mathbf{X}_S\beta_{SX} + \lambda_{SX}\varpi + \epsilon_{SX} \\
S_{SP} &= \mathbf{X}_S\beta_{SP} + \lambda_{SP}\varpi + \epsilon_{SP} \\
S_{SI} &= \mathbf{X}_S\beta_{SI} + \lambda_{SI}\varpi + \epsilon_{SI}
\end{aligned} \tag{5}
$$

where $\epsilon_{SX} \perp \epsilon_{SP} \perp \epsilon_{SI}$. Then, using the framework presented in Section 3.2, I identify the underlying distribution of the unobserved degree of homosexuality $F_{\varpi}(\cdot)$ clean from observable contexts that could affect the *answers* collected in the sexual preferences/perceptions questionnaire.

In practice, however, a complication arises because answers to the questionnaire are not smoothly spread out across the possible categories. The majority of respondents locate themselves at the heterosexual extreme in each question. In consequence, measurement system (5) needs to be modified to accommodate such skewed distributions of the answers. To do so, I embed the factor model in an index function approach (Heckman and MaCurdy, 1986). Let $S_{\cdot}^{*}$ be the observed value of $S_{\cdot}$. S_{SX}^{*} and S_{SP}^{*} are binary scores that take the value of one if the respondent chose any answer different to the extreme heterosexual option. S_{SI}^{*} is a left-tailed censored variable for which there is a mass at the heterosexual extreme and number of individuals in each of the other categories that report increasing levels of homosexuality. It is important to note that while the manifest sores are not continuous, the latent variables S_{SX}, S_{SI} and S_{SP}, and the latent factor

ϖ are. Therefore, measurement system (5) becomes:

$$S^*_{SX} = \begin{cases} 0 & \text{if } \mathbf{X}_S\beta_{SX} + \lambda_{SX}\varpi + \epsilon_{SX} \leq 0 \\ 1 & \text{if } \mathbf{X}_S\beta_{SX} + \lambda_{SX}\varpi + \epsilon_{SX} > 0 \end{cases} \qquad S^*_{SP} = \begin{cases} 0 & \text{if } \mathbf{X}_S\beta_{SP} + \lambda_{SP}\varpi + \epsilon_{SP} \leq 0 \\ 1 & \text{if } \mathbf{X}_S\beta_{SP} + \lambda_{SP}\varpi + \epsilon_{SP} > 0 \end{cases}$$

$$S^*_{SI} = \begin{cases} 0 & \text{if } \mathbf{X}_S\beta_{SI} + \lambda_{SI}\varpi + \epsilon_{SI} \leq 0 \\ S_{SI} & \text{if } \mathbf{X}_S\beta_{SI} + \lambda_{SI}\varpi + \epsilon_{SI} > 0 \end{cases} \tag{6}$$

Assuming that ϵ_{SX}, ϵ_{SP}, and ϵ_{SI} follow a standard normal, then the estimation entails a system comprised by two probits and a type-I Tobit model—in Amemiya (1985, p. 384) parlance—with unobserved heterogeneity. Identification of all parameters in the system, including the ones that determine $F_{\varpi}(\cdot)$, require the normalization $\lambda_{SI}/\sigma_{\epsilon_{SI}} = 1$.

Table 6 presents the estimated parameters of the measurement system (6) used to identify the latent factor of homosexuality for the complete sample. It shows that the latent factor greatly affects the scores purveyed by the answers to the sexual preferences/perceptions questionnaire. Furthermore, as expected, some contexts matter in the determination of the answers to the questionnaire. The results in Table 6 show that women and people that come from a high SES household tend to report higher homosexuality *scores*, while those that live with both parents as a child tend to do the opposite. These are interesting findings as they epitomize the role contexts play in determining the answers to the sexual preferences/perceptions questionnaire.

Table 6: Identification of Homosexuality as a Latent Factor

	Sex. Partner[a]		Self Perception[b]		Sex. Interest[c]	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
Complete Sample (Obs.=2859)						
Age	-0.355	(0.303)	0.358	(0.375)	0.311	(0.360)
Age Sq.	0.010	(0.010)	-0.016	(0.012)	-0.012	(0.012)
Female	0.279***	(0.091)	0.892***	(0.107)	1.431***	(0.114)
MotherSES: High	0.412***	(0.154)	0.375**	(0.166)	0.657***	(0.179)
LivesBothParents	-0.369***	(0.101)	-0.363***	(0.107)	-0.223*	(0.115)
Latent Homosex.	0.578***	(0.047)	0.774***	(0.070)	1	
Constant	1.449	(2.330)	-3.427	(2.853)	-4.150	(2.766)
Men (Obs.=1291)						
Age	0.048	(0.557)	0.547	(0.747)	1.923	(1.244)
Age Sq.	-0.002	(0.018)	-0.022	(0.025)	-0.066	(0.042)
MotherSES: High	0.660**	(0.265)	1.032***	(0.296)	2.090***	(0.528)
LivesBothParents	-0.400**	(0.165)	-0.597***	(0.192)	-0.328	(0.269)
Latent Homosex.	0.417***	(0.059)	0.578***	(0.084)	1	
Constant	-2.179	(4.261)	-5.434	(5.641)	-18.127**	(9.128)
Women (Obs.=1568)						
Age	-0.510	(0.357)	0.232	(0.385)	0.078	(0.413)
Age Sq.	0.015	(0.011)	-0.011	(0.012)	-0.003	(0.013)
MotherSES: High	0.423**	(0.197)	0.209	(0.192)	0.538**	(0.238)
LivesBothParents	-0.351***	(0.126)	-0.252**	(0.119)	-0.200	(0.149)
Latent Homosex.	0.495***	(0.055)	0.567***	(0.065)	1	
Constant	2.891	(2.751)	-1.766	(2.935)	-1.350	(3.190)

Note: *** p<0.01, ** p<0.05, * p<0.1. *Source*: Own calculations.
[a]Binary Score. Takes the value of one if respondent has ever had at least one sexual partner of the same gender. Estimated as a probit.
[b]Binary Score. Takes the value of one if respondent considers he/she is non-heterosexual in any degree. Estimated as a probit.
[c]Left-tail censored variable. Estimated as a tobit.

5.2 Homosexuality Factor, Education, and Labor Market

Now, I extend model (2)-(3) to incorporate the latent homosexuality heterogeneity. Hence, the labor market variables can be analyzed using the model described by (2)-(3)-(6), where $\mathbf{\Theta} = [\theta, \varpi]$ are the dimensions of unobserved heterogeneity, θ are the latent skills, ϖ is the unobserved homosexuality factor, and $\theta \perp \varpi$.

Table 7: Homosexuality Factor in a Roy Model with Selection Equations

	College†	Income		Employed†	
		NoCollege	*College*	*NoCollege*	*College*
Overall					
Latent Homosex.	0.019	1.998	-4.329*	-0.077**	-0.065***
	(0.020)	(2.929)	(2.391)	(0.037)	(0.024)
Latent Skills	0.380***	23.894	47.461***	0.166	0.659***
	(0.133)	(17.191)	(16.953)	(0.238)	(0.176)
Males					
Latent Homosex.	0.043**	-0.944	-3.098	-0.040	-0.070***
	(0.019)	(2.785)	(2.621)	(0.038)	(0.027)
Latent Skills	0.349**	25.439	71.259***	-0.006	0.593**
	(0.173)	(21.779)	(25.955)	(0.324)	(0.261)
Females					
Latent Homosex.	-0.012	3.417	-4.137	-0.075	-0.038
	(0.026)	(2.976)	(2.788)	(0.050)	(0.030)
Latent Skills	0.532***	10.660	27.472	0.315	0.663***
	(0.181)	(19.457)	(19.879)	(0.305)	(0.221)

Note: *** p<0.01, ** p<0.05, * p<0.1. Controls used in the college choice equation are: age, age squared, gender (in the overall sample), municipality population, mother's and father's education, indicators of skipping school, taken out of classroom and police contact while teenager. Controls in the selection equations are: age, age squared, gender (in the overall sample), whether as a child the person lived with both parents, indicator for practicing sports while teenager and male labor force participation rate in the 25 to 39 age bracket in the municipality. Controls in the income equations are: age, age squared, gender (in the overall sample), father's SES, and mother's education. The College variable takes the value of 1 when the person has undertaken any kind of tertiary education in colleges (*høyskole*) or universities (*universitetet*). This includes three-year, four-year college studies or studies in the university with the same length. Total observations for overall, male and female employed sample: 2,127, 937 and 1,190. *Source*: Own calculations. †Coefficients, not marginal effects.

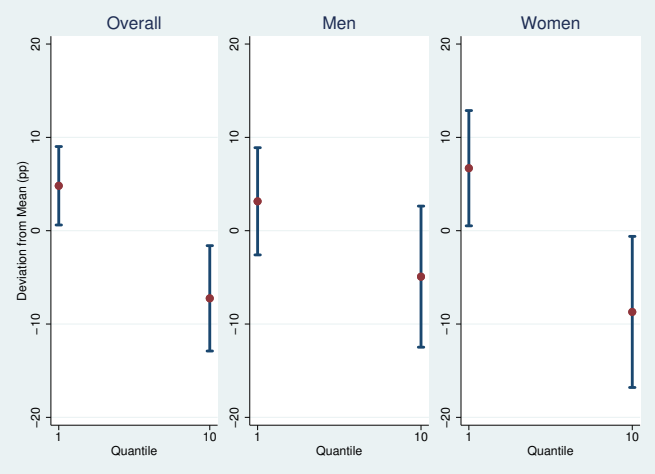
Table 7 presents the main results of the estimation on the complete and the gender-specific samples.[24] Overall, it shows further evidence of the existence of labor market disparities against homosexuals, especially on the employment margin.[25] To analyze the dimensions of the gaps, I use the estimates to predict employment and income for each level of the homosexuality factor

[24]Table 7 reports the results of estimating a full Roy model including endogenous college and employment choice. Section E in the Appendix present the estimates a version of model (2)-(3)-(6) that allows for endogenous choices on employment but without the endogenous choice of education. Overall, the findings are consistent. Namely that the disparities against homosexuals appear at the employment margin and not at the income earned among those who are employed.

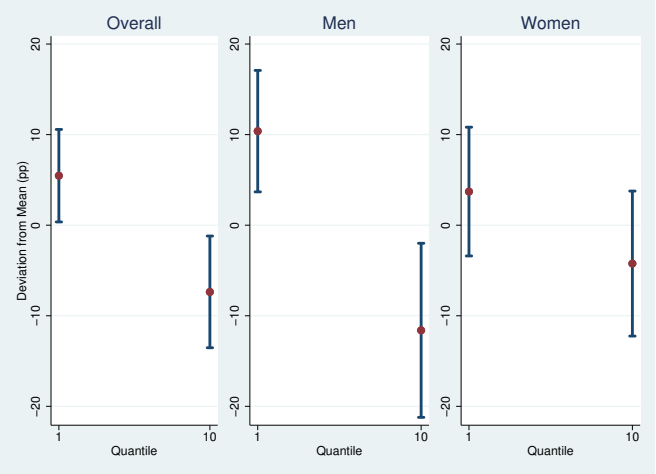
[25]The estimates of the model using full time employment can be found in Section F the Appendix. Furthermore, Table J.2 in the Appendix presents further evidence on the effect of the homosexuality factor on labor-market-related choices. They are the results of estimating models with latent skills on a number of choices like labor market participation, employment, full time employment and working in a white collar occupation. In line with previous results, they indicate that people with a higher degree of homosexuality are less likely to be economically active and employed. Also, it shows no relation between the homosexuality factor and the likelihood of working in a white collar occupation.

Figure 2: Probability of Being Employed

(a) Highest Education Level Achieved: No College



(b) Highest Education Level Achieved: College



Note: Simulated incomes and employment probabilities using the estimated parameters and 40,000 draws from each dimension of unobserved heterogeneity. The capped spikes represent t95% point-wise confidence intervals. *Source: own calculations.*

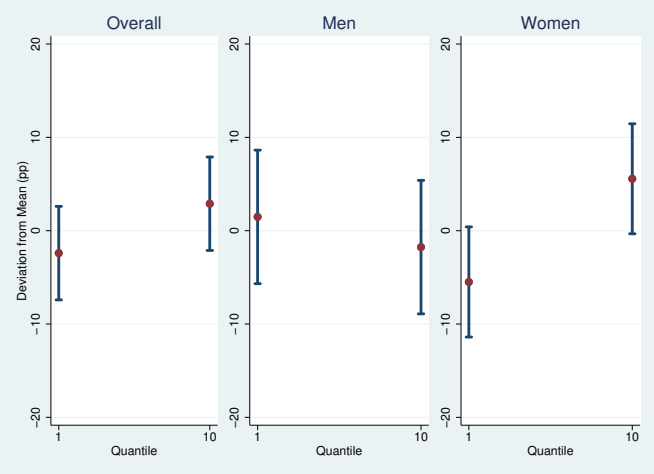
by simulating them with 40,000 draws of each source of unobserved heterogeneity.[26] Figures 2 show the results of the simulations on the employment margin.[27] They indicate that going from the first to the tenth decile of the homosexuality factor distribution—holding everything else constant—diminishes the probability of being employed by about 12.5 percentage points regardless of the education level, a number consistent with the gaps measured in Section 4 based on the dichotomous definitions of homosexuality. These gaps are driven by different gender-specific effects. Namely, the employment disparities against college-educated homosexuals are driven mainly by gaps among men, and the employment disparities against homosexuals with no college education are driven mainly by gaps among women. In fact, Figures 2b suggest that the employment gap between college-educated heterosexual men and college-educated men in the tenth decile of the homosexuality factor is 22 percentage points. According to Figures 2a, the employment gap between

[26]Goodness-of-fit of the models is presented in Tables G.1, G.2 and G.3 in the Appendix. The models perform extremely well in fitting the first and second moments of the variables observed in the data.

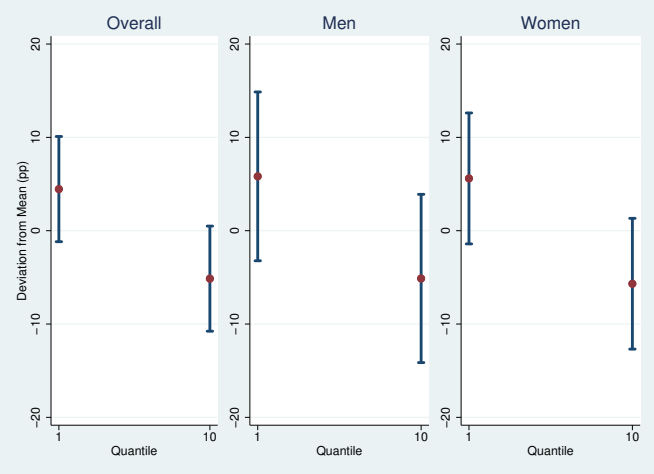
[27]For the sake of brevity, I do not present the simulations to every outcome in each subsample in the main text. They can be found in Appendix I.

Figure 3: Income (Censored)

(a) Highest Education Level Achieved: No College



(b) Highest Education Level Achieved: College



Note: Simulated incomes and employment probabilities using the estimated parameters and 40,000 draws from each dimension of unobserved heterogeneity. The capped spikes represent 95% point-wise confidence intervals. *Source:* own calculations.

heterosexual and homosexual women without college education amounts to 15 percentage points.

Table 7 and Figures 3 show no gradient between income and the homosexuality factor among men, regardless of their education status. Table 7, however, indicates that one standard deviation increase in the overall homosexuality factor ($\sigma_{\varpi} = 1.969$) decreases income by 2.94 percent (($\sigma_{\varpi} * 4.33$)/290, where 290 is the average income for the college-educated) among the college-educated. Figures 3b show that those college-educated that belong to the top decile of the homosexuality factor earn 9.6 percent less than comparable heterosexuals. This overall effect is driven mainly by an income gap that opens up as the homosexuality factor increases among college-educated women. The expected income of college-educated lesbians is 11.2 percent lower than that of college-educated women in the first decile of the homosexuality factor (the difference is significant at the 10% level).

One of the main advantages of the model of endogenous college choice is that it allows me to see any differential effect of homosexuality on income and employment by gender and education level.

Such flexibility is critical to identify a positive gradient between income and the homosexuality factor among women without college education. As opposed to men and college-educated women, women in the top decile of the homosexuality factor without college education earn 11 percent more than comparable heterosexual women.

In sum, the results show that disparities in the labor market against sexual minorities have gender specificities. On one hand, homosexual men are more likely to go to college than their heterosexual counterparts by about 15.7 percentage points—the figure comes from comparing those in the first decile of the homosexuality factor to those in the top decile. And those college-educated homosexual men are the ones who face significant employment disparities. But once employed, homosexual men earn as much as heterosexual men. On the other hand, I present evidence, although less conclusive, that women without college education in higher deciles of the homosexuality factor distribution are less likely to be employed than their heterosexual counterparts, but those with college education face income disparities once employed. These set of results highlight the importance of estimating a model that allows for gender specific estimates and endogenous education and employment choice.

6 Conclusions

This paper incorporates differences in skill distributions and tastes for schooling and employment into the analysis and quantification of the gaps in labor market outcomes observed against homosexual workers. In the process, I extend the Roy model with unobserved heterogeneity to account for endogenous choices of employment. The results from the model indicate that there are in fact differences in the variance distribution of unobserved skills characteristics between homosexual and

heterosexual workers, and that these traits have different effects on education and employment choices, as well as different returns in terms of income across different sexual preferences groups. In particular, skills are more rewarded in terms of income among homosexuals than among heterosexuals. All these results indicate that there are several forces working simultaneously in opening the income gaps observed.

My findings indicate that homosexuals face a probability of finding a job that is 10–20 percentage points lower than the probability faced by a heterosexual. I find that the gap is not explained by differences in the stock of skills or gender composition. I also find that once employed, homosexuals face no income gap because the ones that get a job are highly skilled, and the market rewards those skills profusely. In a second model, I incorporate homosexuality as a latent continuous construct. My findings corroborate the ones obtained from the previous models, attesting to the fact that the biggest hurdle homosexuals face in the labor market is getting a job. Furthermore, the second model allows me to analyze gender-specific outcomes. I find that discrimination in getting a job is greater for homosexual men, but lesbians are also penalized through the income earned, especially the college-educated who manage to get a full-time job.

This is the first paper to incorporate endogenous decisions and unobserved heterogeneity in the analysis of income discrimination against homosexuals. My results show that the overall income gaps observed come entirely from a significant difference in the probability of being employed. Thus, future research is required to improve our understanding of the reasons behind these employability differences. That way, economic research will provide arguments that will help to push forward equality, as barriers against sexual minorities continue to fall in other aspects of life.

References

Allegretto, S. and Arthur, M. (2001). An empirical analysis of homosexual/heterosexual male earnings differentials: unmarried and unequal? *Industrial and labor relations review*, pages 631–646.

Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press.

Anderssen, N. and Hellesund, T. (2009). Heteronormative Consensus in the Norwegian Same-Sex Adoption Debate? *Journal of Homosexuality*, 56(1):102–120.

Antecol, H., Jong, A., and Steinberger, M. (2007). The sexual orientation wage gap: The role of occupational sorting and human capital, and discrimination. *Indus. & Lab. Rel. Rev.*, 61:518.

Antecol, H. and Steinberger, M. D. (2013). Labor Supply Differences Between Married Heterosexual Women and Partnered Lesbians: A Semi-Parametric Decomposition Approach. *Economic inquiry*, 51(1):783–805.

Badgett, M. (1995). The wage effects of sexual orientation discrimination. *Industrial and labor relations review*, pages 726–739.

Bartholomew, D., Knott, M., and Moustaki, I. (2011). Latent variable models and factor analysis: A unified approach. *Wiley Series in Probability and Statistics. West Sussex, United Kingdom*.

Becker, G. (1957). The economics of discrimination. University of Chicago Press: Chicago.

Becker, G. and Chiswick, B. (1966). Education and the distribution of earnings. *The American Economic Review*, 56(1/2):358–369.

Berg, N. and Lien, D. (2002). Measuring the effect of sexual orientation on income: evidence of discrimination? *Contemporary economic policy*, 20(4):394–414.

Bertrand, M., Goldin, C., and Katz, L. F. (2010). Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors. *American Economic Journal: Applied Economics*, 2(3):228–255.

Black, D., Makar, H., Sanders, S., and Taylor, L. (2003). The earnings effects of sexual orientation. *Industrial and labor relations review*, pages 449–469.

Black, D. A., Sanders, S. G., and Taylor, L. J. (2007). The economics of lesbian and gay families. *The Journal of Economic Perspectives*, 21(2):53–70.

Blau, F. D. and Kahn, L. M. (2017). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature*, 55(3):789–865.

Bond, T. N. and Lehmann, J.-Y. K. (2018). Prejudice and racial matches in employment. *Labour Economics*, 51:271–293.

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

Cain, G. G. (1986). The economic analysis of labor market discrimination: A survey. In Ashenfelter, O. and Layard, R., editors, *Handbook of Labor Economics Volume 1*, pages 693–785. Elsevier Masson SAS.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica: Journal of the Econometric Society*, 5:1127–1160.

Carneiro, P., Hansen, K., and Heckman, J. (2003). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review*, 44(2):361–422.

Carpenter, C. (2007). Revisiting the income penalty for behaviorally gay men: Evidence from nhanes iii. *Labour economics*, 14(1):25–34.

Carpenter, C. (2009). Sexual orientation and outcomes in college. *Economics of Education Review*, 28(6):693–703.

Clain, S. and Leppel, K. (2001). An investigation into sexual orientation discrimination as an explanation for wage differences. *Applied economics*, 33(1):37–47.

Espinoza, R., Sarzosa, M., and Urzua, S. (2014). The causal effect of skills: An international study. *University of Maryland*, pages 1–53.

Falcoz, C. (2008). *Homophobie dans l'entreprise*. La Documentation Francaise.

Frank, W. (2014). *Law and the Gay Rights Story*. The Long Search for Equal Justice in a Divided Democracy. Rutgers University Press.

Frohn, D. (2007). *Out im office?!, Sexuelle Identität, (Anti-) Diskriminierung und Diversity am Arbeitsplatz* . Universität zu Köln, Psychologisches Institut.

Hansen, K., Heckman, J., and Mullen, K. (2004). The effect of schooling and ability on achievement test scores. *Journal of Econometrics*, 121(1):39–98.

Harter, S. (1988). Manual for the Self-perception Profile for Adolescents. University of Denver, Denver, CO.

Harter, S. (2012). *Self-perception profile for adolescents: Manual and Questionnaires*. University of Denver, Denver, CO.

Heckman, J. and Honore, B. (1990). The empirical content of the roy model. *Econometrica: Journal of the Econometric Society*, pages 1121–1149.

Heckman, J. and Masterov, D. (2007). The productivity argument for investing in young children. *Applied Economic Perspectives and Policy*, 29(3):446–493.

Heckman, J., Stixrud, J., and Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3):411–482.

Heckman, J. J. and MaCurdy, T. E. (1986). Chapter 32 Labor econometrics. *Handbook of econometrics*, 3:1917–1977.

Hennum, R. (2001). Lesbiske og homofiles rettsstilling [Law status for gays and lesbians]. In Brantsaeter, M., Eikvam, T., Kjaer, R., and Amas, K. O., editors, *Norsk homoforskning Gay Science in Norway*, pages 85–103. Oslo.

ILGA Europe (2019). *Annual Review of the Human Rights Situation of Lesbian, Gay, Bisexual, Trans and Intersex people* . ILGA Europe, Brussels, Belgium.

Judd, K. L. (1998). *Numerical Methods in Economics*. MIT Press, Cambridge, Massachusetts.

Kotlarski, I. (1967). On characterizing the gamma and the normal distribution. *Pacific Journal of Mathematics*, 20(1):69–76.

Lang, K. and Lehmann, J.-Y. K. (2012). Racial Discrimination in the Labor Market: Theory and Empirics. *Journal of Economic Literature*, 50(4):959–1006.

Laumann, E., Gagnon, J., Michael, R., and Michaels, S. (1994). The social organization of sexuality: Sexual practices in the united states. Chicago: University of Chicago Press.

Manning, A. and Swaffield, J. (2008). The gender gap in early-career wage growth. *The Economic Journal*, 118(530):983–1024.

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy*, 66(4):281–302.

Mincer, J. and Polachek, S. (1974). Family Investments in Human Capital: Earnings of Women. *Journal of Political Economy*, 82(2, Part 2):S76–S108.

Mroz, T. A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica*, 55(4):765.

Mucciaroni, G. (2009). *Same Sex, Different Politics*. Success and Failure in the Struggles over Gay Rights. University of Chicago Press.

Murnane, R., Willett, J., and Levy, F. (1995). The growing importance of cognitive skills in wage determination. *The Review of Economics and Statistics*, 77(2):251–266.

Neal, D. A. and Johnson, W. R. (1996). The role of premarket factors in black-white wage differences. *Journal of Political Economy*, 104(5):869–895.

Ritter, J. A. and Taylor, L. J. (2011). Racial disparity in unemployment. *The Review of Economics and Statistics*, 93(1):30–42.

Roy, A. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers*, 3(2):135–146.

Sarzosa, M. (2015). The dynamic consequences of bullying on skill accumulation. Purdue University.

Sarzosa, M. and Urzua, S. (2016). Implemeting factor models for unobserved heterogeneity in stata. *The Stata Journal*, 16(1):197–228.

Strand, N. P. and von Soest, T. (2008). Young in norway - longitudinal. documentation of design, variables and scales. NOVA.

Tebaldi, E. and Elmslie, B. (2006). Sexual orientation and labour supply. *Applied Economics*, 38(5):549–562.

Urzua, S. (2008). Racial labor market gaps. *Journal of Human Resources*, 43(4):919.

van Balen, B., Barry, U., Holzhacker, R., Villagomez, E., and Wladasch, K. (2010). Part I - The Situation of LGBT Groups in the Labour Market in European Member States. Technical Report 2008-007, Human European Consultancy - OSB Consulting GmbH.

Weinberger, C. J. and Kuhn, P. J. (2010). Changing Levels or Changing Slopes? The Narrowing of the Gender Earnings Gap 1959 – 1999. *ILR Review*, 63(3):384–406.

Wichstraum, L. (1995). Harter's Self-Perception Profile for Adolescents: Reliability, Validity, and Evaluation of the Question Format. *Journal of Personality Assessment*, 65(1):100–116.

Appendix

A YiN Questions and Answers Regarding Sexual Preferences

1. **Sexual Interest (SI)**: Are you sexually interested in men or women (sexually attracted to, sexual fantasies about)?

   - Women only
   - Mainly women, but exceptionally men
   - Mainly women, but sometimes men
   - About the same amount women and men
   - Mainly men, but sometimes women
   - Mainly men, but exceptionally women
   - Men only

2. **Self-perception (SP)**: How would you rate yourself on a scale from absolutely heterosexual to absolutely homosexual/lesbian?

   - Only heterosexual
   - Mainly heterosexual, to a very small extent homosexual/lesbian
   - Mainly heterosexual, to some extent homosexual/lesbian
   - About as much homosexual/lesbian as heterosexual
   - Mainly homosexual/lesbian, to some extent heterosexual
   - Mainly homosexual/lesbian, to a very small extent heterosexual
   - Only homosexual/lesbian

3. **Sexual Relations (SX)**: Have you had any kind of sexual relations with persons of the same gender as yourself?

B General Distribution of Skills

Let me introduce the unobserved heterogeneity and the Roy model of potential outcomes in a context where one general distribution of skills is estimated. Putting it in terms of (2), I estimate the following system:

$$\mathbf{Y} = \begin{bmatrix} D \\ E_1 \\ Y_1 \\ E_0 \\ Y_0 \end{bmatrix} = \begin{bmatrix} \mathbb{1}\left[\gamma_D Gay + \mathbf{X}_D\beta^{Y_D} + \alpha^{Y_D}\theta + e^D > 0\right] \\ D\mathbb{1}\left[\gamma_{E_1} Gay + \mathbf{X}_E\beta^{E_1} + \alpha^{E_1}\theta + e^{E_1} > 0\right] \\ DE_1\left(\gamma_{Y_1} Gay + \mathbf{X}_Y\beta^{Y_1} + \alpha^{Y_1}\theta + e^{Y_1}\right) \\ (1-D)\,\mathbb{1}\left[\gamma_{E_0} Gay + \mathbf{X}_E\beta^{E_0} + \alpha^{E_0}\theta + e^{E_0} > 0\right] \\ (1-D)\,E_0\left(\gamma_{Y_0} Gay + \mathbf{X}_Y\beta^{Y_0} + \alpha^{Y_0}\theta + e^{Y_0}\right) \end{bmatrix}$$

$$\mathbf{T} = \mathbf{X}_T\beta^T + \alpha^{\mathbf{T}}\mathbf{\Theta} + \mathbf{e}^{\mathbf{T}}$$

for the complete sample.[28] In this case, the parameters of interest are γ_D, γ_{E_0}, γ_{E_1}, γ_{Y_0} and γ_{Y_1}.[29] I estimate two sets of models. One considering the employment decision of whether to have a job or not (Table B.1) and another one where the employment decision is to have a full time job or not (Table B.2). In both Tables and for each definition, I present the estimated coefficients in the top panel and the gender-specific effects in terms of each outcome variable (i.e., $^{\Delta\mathbf{Y}}/_{\Delta Gay}$) in the bottom panel. The calculatoin of the gender-specific effects incorporate the fact that the model is non-linear in the college and employment decisions.

Tables B.1 and B.2 show that, even after controlling for skills, college attendance does not differ across sexual preferences, with the exception of results according to the SX definition. I find that SX homosexual men are 13 percentage points more likely to have gone to college than comparable heterosexual men, and that lesbians are 14.6 percentage points less likely to have gone to college than heterosexual women.

Tables B.1 and B.2 also shows that the disparities take place in accessing jobs and not in the pay itself, and that these gaps are larger among college educated than among those with no college education. College-educated homosexual men according to the SI definition are 20 percentage points less likely to be employed and 30 percentage points less likely to have a full time job than comparable heterosexuals. SI homosexual men without college education are 10 percentage points less likely to be employed and 12 percentage points less likely to have a full-time job than comparable heterosexuals. Figures are similar under the SP definition. College-educated homosexual men according to the SP definition are 12.7 percentage points less likely to be employed and 23.7 percentage points less likely to have a full time job than comparable heterosexuals. SP homosexual men without college education have no significant differences with comparable heterosexuals in terms of the probability of having a job. Among women, Table B.1 and B.2 show no significant an employability gap against lesbians.

The results of Table B.1 and B.2 attest to the importance of a model that allows for endogenous choices and separate estimates for two different sectors. This way, I find particular sectors where discrimination is stronger, namely among the higher educated male workers.

[28] See the estimation of the measurement system of test scores for the whole sample from which the distribution of the latent skills θ are identified in Table H.1 and the estimated distribution of skills θ in Figure H.1 in Appendix H.

[29] For completeness, I estimate a model without an education choice equation—which is the closest to the reduced form estimates in the existing literature—and present its results in Appendix C.

Table B.1: Effect of Homosexuality on Annual Income, Employment and College, Roy Model With Overall Distribution

	College	Income		Employed	
		No College	*College*	*No College*	*College*
SI (*N=2274 / N Uncens=1821*)					
Gay	0.089	9.320	-5.995	-0.457*	-0.640***
	(0.203)	(23.098)	(24.034)	(0.250)	(0.157)
GayXFemale	-0.017	-2.829	0.418	0.186	0.487***
	(0.240)	(25.114)	(23.615)	(0.304)	(0.185)
Effect for Men	0.034	9.320	-5.995	-0.098*	-0.203***
	(0.060)	(23.098)	(24.034)	(0.054)	(0.045)
Effect for Women	0.021	6.491	-5.577	-0.071	-0.047*
	(0.034)	(34.113)	(33.686)	(0.044)	(0.026)
SP (*N=2280 / N Uncens=1828*)					
Gay	0.105	9.879	-32.440	-0.296	-0.404**
	(0.192)	(19.766)	(19.811)	(0.240)	(0.161)
GayXFemale	-0.152	4.223	17.293	0.085	0.291
	(0.234)	(23.204)	(21.409)	(0.296)	(0.194)
Effect for Men	0.040	9.879	-32.440	-0.059	-0.127***
	(0.059)	(19.766)	(19.811)	(0.047)	(0.045)
Effect for Women	-0.014	14.102	-15.147	-0.054	-0.035
	(0.038)	(30.476)	(29.162)	(0.042)	(0.030)
SX (*N=2283 / N Uncens=1827*)					
Gay	0.346*	-4.722	1.765	-0.287	-0.130
	(0.205)	(21.987)	(17.996)	(0.271)	(0.173)
GayXFemale	-0.784***	22.883	-27.863	0.038	-0.017
	(0.272)	(26.865)	(22.636)	(0.334)	(0.223)
Effect for Men	0.128**	-4.722	1.765	-0.059	-0.040
	(0.063)	(21.987)	(17.996)	(0.058)	(0.048)
Effect for Women	-0.146**	18.161	-26.097	-0.066	-0.046
	(0.059)	(34.709)	(28.912)	(0.054)	(0.044)

Note: *** p<0.01, ** p<0.05, * p<0.1. The coefficients presented are the ones associated with the homosexuality indicators. The College variable takes the value of 1 when the person has undertaken any kind of tertiary education in colleges (*høyskole*) or universities (*universitetet*). This includes three-year, four-year college studies or studies in the university with the same length. The controls used in the College choice equation and not presented in the Table are: age, age squared, gender, dummies for father's education, dummies for mother's education, whether as a child the person lived with one parent, a remarried parent or with no parents, number of books available at home as a child, and the unobservable factor of skills. The controls used in the income equations and not presented in the Table are age, age squared, gender, dummies for Father SES, population of the municipality and the unobservable factor of skills. For every definition, the top panel presents the coefficients and the bottom panels present the effects in terms of changes in the outcome variable for each gender. *Source*: Own calculations.

Table B.2: Effect of Homosexuality on Annual Full Time Income, Full Time Employment and College, Roy Model With Overall Distribution

	College	Income		Employed	
		No College	*College*	*No College*	*College*
SI (*N=2335 / N Uncens=1409*)					
Gay	0.014	14.468	13.425	-0.432*	-0.772***
	(0.222)	(23.809)	(30.958)	(0.239)	(0.152)
GayXFemale	-0.009	-38.460	-29.256	0.632**	0.807***
	(0.272)	(32.498)	(33.421)	(0.282)	(0.176)
Effect for Men	0.005	14.468	13.425	-0.117	-0.293***
	(0.082)	(23.809)	(30.958)	(0.073)	(0.055)
Effect for Women	0.001	-23.992	-15.831	0.080	0.014
	(0.039)	(40.276)	(45.540)	(0.052)	(0.030)
SP (*N=2340 / N Uncens=1415*)					
Gay	-0.044	10.439	-3.184	-0.251	-0.625***
	(0.209)	(20.205)	(27.221)	(0.230)	(0.154)
GayXFemale	-0.080	-1.006	-15.434	0.329	0.600***
	(0.265)	(27.802)	(28.976)	(0.274)	(0.182)
Effect for Men	-0.017	10.439	-3.184	-0.062	-0.237***
	(0.078)	(20.205)	(27.221)	(0.060)	(0.056)
Effect for Women	-0.036	9.433	-18.618	0.031	-0.010
	(0.045)	(34.360)	(39.745)	(0.052)	(0.034)
SX (*N=2345 / N Uncens=1414*)					
Gay	0.230	-22.528	11.515	-0.055	-0.260
	(0.216)	(21.501)	(20.460)	(0.269)	(0.162)
GayXFemale	-0.664**	18.503	-41.058	0.232	0.351*
	(0.301)	(31.324)	(27.361)	(0.323)	(0.208)
Effect for Men	0.086	-22.528	11.515	-0.013	-0.096*
	(0.075)	(21.501)	(20.460)	(0.063)	(0.058)
Effect for Women	-0.137**	-4.025	-29.542	0.071	0.036
	(0.068)	(37.985)	(34.156)	(0.067)	(0.049)

Note: *** p<0.01, ** p<0.05, * p<0.1. The coefficients presented are the ones associated with the homosexuality indicators. The College variable takes the value of 1 when the person has undertaken any kind of tertiary education in colleges (*høyskole*) or universities (*universitetet*). This includes three-year, four-year college studies or studies in the university with the same length. The controls used in the College choice equation and not presented in the Table are: age, age squared, gender, dummies for father's education, dummies for mother's education, whether as a child the person lived with one parent, a remarried parent or with no parents, number of books available at home as a child, and the unobservable factor of skills. The controls used in the income equations and not presented in the Table are age, age squared, gender, dummies for Father SES, population of the municipality, height and the unobservable factor of skills. For every definition, the top panel presents the coefficients and the bottom panels present the effects in terms of changes in the outcome variable for each gender. *Source*: Own calculations.

C General Distribution of Skills: Income Gap Controlling for Unobservables

In this appendix, I introduce the unobserved skills in a labor market outcome model similar to those of the existing literature. The idea is to expand the reduced-form results to incorporate unobserved heterogeneity. Hence the equation I estimate is

$$Y = \alpha + \gamma Gay + \beta \mathbf{X} + \psi IMR + \lambda\theta + \varepsilon$$
$$E = \alpha_E + \gamma_E Gay + \beta_E \mathbf{X} + \lambda_E \theta + \epsilon$$
$$\mathbf{T} = \mathbf{X}_T \beta^T + \alpha^{\mathbf{T}} \mathbf{\Theta} + \mathbf{e}^{\mathbf{T}}$$

where θ is the unobserved skills that are drawn from a general distribution estimated using the complete sample. Like with the reduced-form results, I am interested in the estimates of γ and γ_E. The results are presented in Table C.1 for each definition of homosexuality and each measure of yearly income. They suggest that the greatest differences between heterosexuals' and homosexuals' job market rewards are in terms of the access to a job rather than in the amount payed once they have a job. Interestingly, once I control for skills, I find that, unlike homosexual males, lesbians face no difference in the probability of being employed compared to their heterosexual counterparts.

Table C.1: Income Gap Controlling for Skills with Selection Equations

	SI		**SP**		**SX**	
	Income	Employed	Income	Employed	Income	Employed
IncomeEmpl						
Gay	-14.272	-0.623***	-24.379	-0.374***	-14.154	-0.231
	(19.590)	(0.130)	(15.335)	(0.130)	(14.324)	(0.142)
Gay × Female	9.963	0.431***	17.429	0.236	5.909	0.060
	(18.366)	(0.155)	(16.021)	(0.159)	(17.214)	(0.182)
Obs.(N|Cnsrd)	2259	1806	2266	1814	2270	1814
IncomeFullTime4						
Gay	4.083	-0.699***	-4.790	-0.526***	9.011	-0.266**
	(46.164)	(0.124)	(38.594)	(0.123)	(31.953)	(0.134)
Gay × Female	-46.725	0.768***	-30.537	0.520***	-52.915	0.367**
	(52.697)	(0.145)	(43.168)	(0.147)	(43.874)	(0.169)
Obs.(N|Cnsrd)	2368	1442	2374	1449	2378	1447

Note: *** p<0.01, ** p<0.05, * p<0.1. The controls used in all the equation and not presented in the Table are: age, age squared, gender, dummies for father's education, whether as a child the person lived with one parent, a remarried parent or with no parents. Constants not shown. *Source*: Own calculations.

D Models With Dichotomous Definitions of Homosexuality

D.1 Loadings of Latent Factors

Table D.1: Skill Loadings: Income and Employment, Roy Model

	Heterosexuals		Homosexuals	
	SI	SP	SI	SP
College†	2.078***	1.992***	6.717	6.922
	(0.350)	(0.341)	(4.905)	(5.268)
Employed				
No College	0.043	0.483	5.537	-0.112
	(0.329)	(0.325)	(4.412)	(1.268)
College	0.624*	0.516	0.502	0.952
	(0.352)	(0.341)	(2.686)	(1.798)
Income				
No College	35.540	29.974	124.351	105.036
	(22.786)	(25.782)	(255.394)	(91.294)
College	102.116***	76.439**	300.241	428.380***
	(32.114)	(31.290)	(250.011)	(157.841)
Obs.(N)	1746	1825	518	446
Obs.(Cnsrd)	1438	1493	373	326

Note: *** p<0.01, ** p<0.05, * p<0.1. Controls used in the college choice equation are: age, age squared, gender, municipality population, mother's and father's education, indicators of skipping school, taken out of classroom and police contact while teenager. Controls in the selection equations are: age, age squared, gender, whether as a child the person lived with both parents, indicator for practicing sports while teenager and male labor force participation rate in the 25 to 39 age bracket in the municipality. Controls in the income equations are: age, age squared, gender, father's SES, and mother's education. The College variable takes the value of 1 when the person has undertaken any kind of tertiary education in colleges (*høyskole*) or universities (*universitetet*). This includes three-year, four-year college studies or studies in the university with the same length. *Source*: Own calculations. †Coefficients, not marginal effects.

D.2 Goodness-of-Fit

Table D.2: Goodness of Fit - Annual Income by Schooling level and Sexual Preference Definition Roy Model, Heterosexuals Sample

	College	**Employment**		**Income**	
	Means	Means	Std.Dev.	Means	Std.Dev.
SI					
1. Model Without Endogenous Education Choice					
Actual		0.824	0.381	235.909	154.010
Model		0.821	0.383	235.954	155.524
2. Model With Endogenous Education Choice					
Actual	0.649				
Model	0.654				
No College					
Actual		0.863	0.344	242.878	142.006
Model		0.854	0.352	240.357	145.003
College					
Actual		0.805	0.396	230.578	161.850
Model		0.793	0.405	226.475	163.426
SP					
1. Model Without Endogenous Education Choice					
Actual		0.817	0.386	234.355	155.464
Model		0.815	0.388	234.145	156.276
2. Model With Endogenous Education Choice					
Actual	0.665				
Model	0.665				
No College					
Actual		0.864	0.343	242.863	143.266
Model		0.860	0.346	241.409	145.566
College					
Actual		0.796	0.403	228.534	162.795
Model		0.789	0.408	226.563	163.909

Note: *** p<0.01, ** p<0.05, * p<0.1 indicate the probability of rejecting the null hypothesis of the model predictions being different to the actual moments provided by the data. The model predictions are the result of 40,000 simulations of the estimated models. *Source*: Own calculations.

Table D.3: Goodness of Fit - Annual Income by Schooling level and Sexual Preference Definition Roy Model, Homosexuals Sample

	College	**Employment**		**Income**	
	Means	Means	Std.Dev.	Means	Std.Dev.
SI					
1. Model Without Endogenous Education Choice					
Actual		0.722	0.448	187.037	153.482
Model		0.705	0.456	183.454	154.725
2. Model With Endogenous Education Choice					
Actual	0.739				
Model	0.719				
No College					
Actual		0.771	0.421	197.709	153.842
Model		0.731	0.430	188.752	159.243
College					
Actual		0.707	0.456	184.355	157.720
Model		0.694	0.460	177.723	157.386
SP					
1. Model Without Endogenous Education Choice					
Actual		0.740	0.439	188.186	148.691
Model		0.723	0.447	184.378	150.918
2. Model With Endogenous Education Choice					
Actual	0.678				
Model	0.662				
No College					
Actual		0.788	0.410	204.260	148.560
Model		0.745	0.431	196.365	157.397
College					
Actual		0.720	0.450	181.901	154.018
Model		0.694	0.458	167.424*	154.405

Note: *** p<0.01, ** p<0.05, * p<0.1 indicate the probability of rejecting the null hypothesis of the model predictions being different to the actual moments provided by the data. The model predictions are the result of 40,000 simulations of the estimated models. *Source*: Own calculations.

D.3 Results from Counterfactuals

Table D.4: Heterosexual and Homosexual Annual Income and College Enrollment Under Different Assumptions

	College Enrollment			
	Heter.	Homosexuals		
	Basic	Basic	Eq Gender	Het. Unobs
SI	0.654	0.719	0.710	0.613
SP	0.665	0.662	0.657	0.645

Income and Labor Market Participation

SI

	No College			College			Overall		
	Empl.	CensY	Y	Empl.	CensY	Y	Empl.	CensY	Y
Het.	0.854	281.422	240.357	0.793	285.450	226.475	0.814	283.990	231.276
Hom.	0.731	257.434	188.752	0.694	256.336	177.723	0.704	257.142	181.236
Hom. (Eq Gender)	0.758	278.577	211.919	0.656	260.026	170.396	0.685	266.456	182.852
Hom. (Het. Unobs)	0.553	277.621	156.301	0.662	282.267	188.795	0.620	286.945	177.971

SP

	No College			College			Overall		
	Empl.	CensY	Y	Empl.	CensY	Y	Empl.	CensY	Y
Het.	0.860	280.631	241.409	0.789	287.072	226.563	0.813	284.782	231.543
Hom.	0.745	263.791	196.365	0.694	240.754	167.424	0.712	250.464	178.354
Hom. (Eq Gender)	0.750	280.686	210.683	0.671	244.829	164.535	0.699	259.086	181.223
Hom. (Het. Unobs)	0.744	277.236	205.895	0.692	286.362	202.311	0.710	296.529	214.757

Note: The values presented are result of 40,000 simulations for each *basic* model and 200,000 simulations for each counterfactual model. *Empl.* stands for probability of being employed. *CensY.* is the censored income. *Y* stands for the uncensored income (i.e., including the zeros of the unemployed). *Het.* presents the fitted values of the model for heterosexuals (i.e., $\Pr\left[D^S\left(\mathbf{X}_D^S, \theta^S\right)=1\right]$, $\Pr\left[E^S\left(\mathbf{X}_E^S, \theta^S\right)=1\right]$ and $E\left[Y^S\left(\mathbf{X}_Y^S, \theta^S\right)\right]$). The label *basic* presents the fitted values for the homosexual sample with their own characteristics (i.e., $\Pr\left[D^G\left(\mathbf{X}_D^G, \theta^G\right)=1\right]$, $\Pr\left[E^G\left(\mathbf{X}_E^G, \theta^G\right)=1\right]$ and $E\left[Y^G\left(\mathbf{X}_Y^G, \theta^G\right)\right]$). *Eq Gender* presents a counterfactual in which the gender proportion in the homosexuals' samples is equated to the one in the heterosexuals' samples. Label *Het. Unobs* presents the simulated means of homosexuals when the gender distribution and the unobservables are taken from the heterosexual sample (i.e., $\Pr\left[D^G\left(\mathbf{X}_D^G, \theta^S\right)=1\right]$, $\Pr\left[E^G\left(\mathbf{X}_E^G, \theta^S\right)=1\right]$ and $E\left[Y^G\left(\mathbf{X}_Y^G, \theta^S\right)\right]$). *Source*: Own calculations.

E Models with homosexuality factor but without Endogenous Choice of Schooling

Table E.1: Homosexuality Factor in Model with Selection Equations

	Overall Sample		Males		Females	
	Income	Employed	Income	Employed	Income	Employed
IncomeEmpl						
Latent Homosex.	-2.325	-0.080***	-3.602	-0.065***	-2.609	-0.061**
	(2.776)	(0.019)	(2.642)	(0.021)	(2.807)	(0.026)
Latent Skills	29.734*	0.415***	39.490**	0.272	19.398	0.502***
	(15.663)	(0.135)	(18.402)	(0.189)	(19.711)	(0.173)
Obs.(N|Censored)	2214	1781	971	803	1243	978
IncomeFullTime4						
Latent Homosex.	-3.695	-0.075***	-2.401	-0.075***	-6.764**	0.022
	(2.890)	(0.020)	(2.995)	(0.020)	(3.036)	(0.022)
Latent Skills	24.375	0.447***	29.363	0.287	9.903	0.282*
	(17.671)	(0.145)	(19.906)	(0.175)	(20.836)	(0.154)
Obs.(N|Censored)	1816	1383	982	738	1293	645

Note: *** p<0.01, ** p<0.05, * p<0.1. The controls used in all the equation and not presented in the Table are: age, age squared, gender (complete sample only), dummies for father's education, whether as a child the person lived with one parent, a remarried parent or with no parents, and the Inverse Mills Ratio. *Source*: Own calculations.

In this Appendix, I present the results of the analysis of the effect of homosexuality as unobserved heterogeneity using a version of model (2)-(3)-(6) that allows for endogenous choices on employment but without the endogenous choice of education.

Table E.1 presents the results for the complete sample and for the gender specific samples. In particular, it shows that discrimination against homosexuals happens at the time of obtaining a job and not at the income earned itself. It is evident that the homosexuality factor is negatively related with the probability of being employed, while that is not the case for the income earned. To analyze the dimensions of those gaps, I use the estimates to predict employment and income for each level of the homosexuality factor by simulating them with 40,000 draws of each source of unobserved heterogeneity.[30]

Figure E.1 shows the results of the simulations on the employment margin—those on the income part of the model can be found in Figure I.1 in Appendix I. They indicate that going from the heterosexual extreme to the homosexual extreme of the homosexuality factor distribution—holding

[30]Tables G.1 and G.2 in the Appendix show that the models do a good job fitting the means and standard errors of the actual labor market variables.

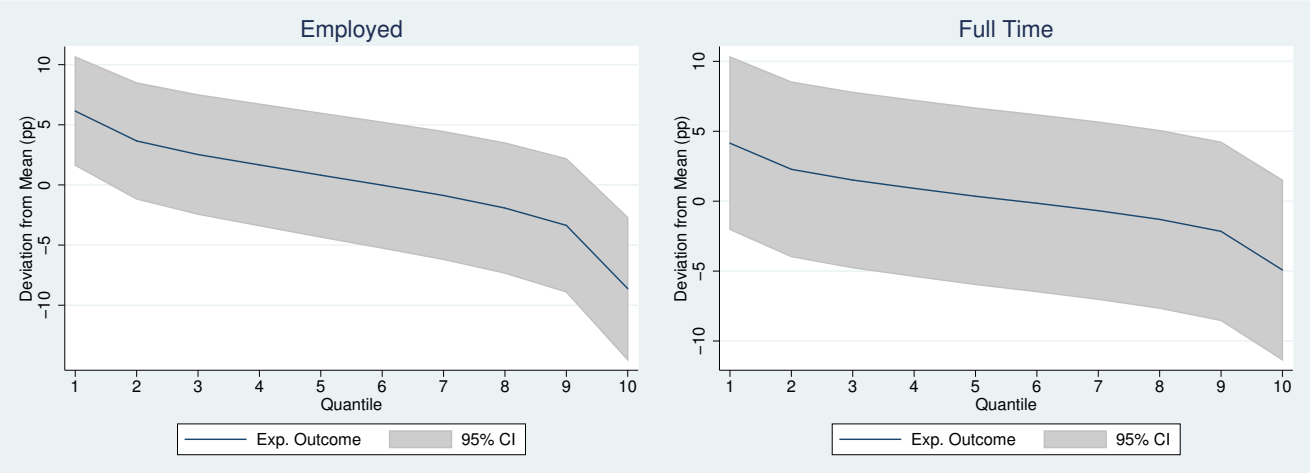
everything else constant—decreases the probability of being employed by 14.76 percentage points, a number consistent with the gaps measured in Section 4 based on the dichotomous definitions of homosexuality. The same pattern is observed when analyzing full-time workers. High levels of the latent factor of homosexuality face a probability of being full-time employed 9 percentage points lower.

As mentioned above, one of the advantages of analyzing homosexuality as a source of unobserved heterogeneity is that I can estimate gender-specific homosexuality factors and analyze discrimination against gay men separately from discrimination against lesbians. Using the gender specific model whose results I also present in Table E.1, and whose goodness-of-fit I also check in Tables G.1 and G.2 in the appendix, I predict the probability of employment and income for each level of the homosexuality factor.

Gender-Specific Model, Men: Figure E.2 analyzes labor market discrimination against gay men. They show a steep negative gradient between the level of latent homosexuality and the chances of getting a job. In fact, men in the top decile of the homosexuality factor are 18.11 percentage points less likely to be employed and 25.54 percentage points less likely to hold a full-time job than extreme heterosexual men.

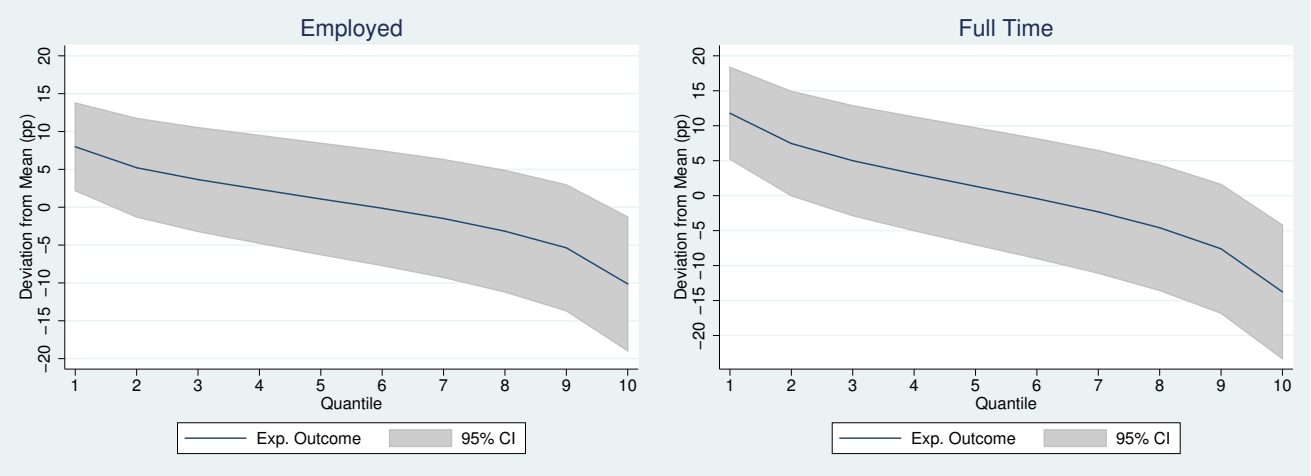
Gender-Specific Model, Women: Discrimination against homosexual women is different. Figure E.3 shows that the employment situation of women depends on whether they are full time employed or not. Homosexual women face a 13.57 percentage points lower probability of finding *a* job than extreme heterosexual women. However, lesbians have the same probability of finding a full-time job as heterosexual women. Due to this lack of selection and contrary to the men's situation, homosexual women do face discrimination in terms of income in full-time jobs. Women with high values of the homosexuality factor receive, on average, an income from their full-time jobs that is 16.39 percent shorter that what comparable heterosexual women receive.

Figure E.1: Probability of Employment at Levels of Homosexuality
Complete Sample



Note: Simulated incomes and employment probabilities using the estimated parameters and 40,000 draws from each dimension of unobserved heterogeneity. *Source*: own calculations.

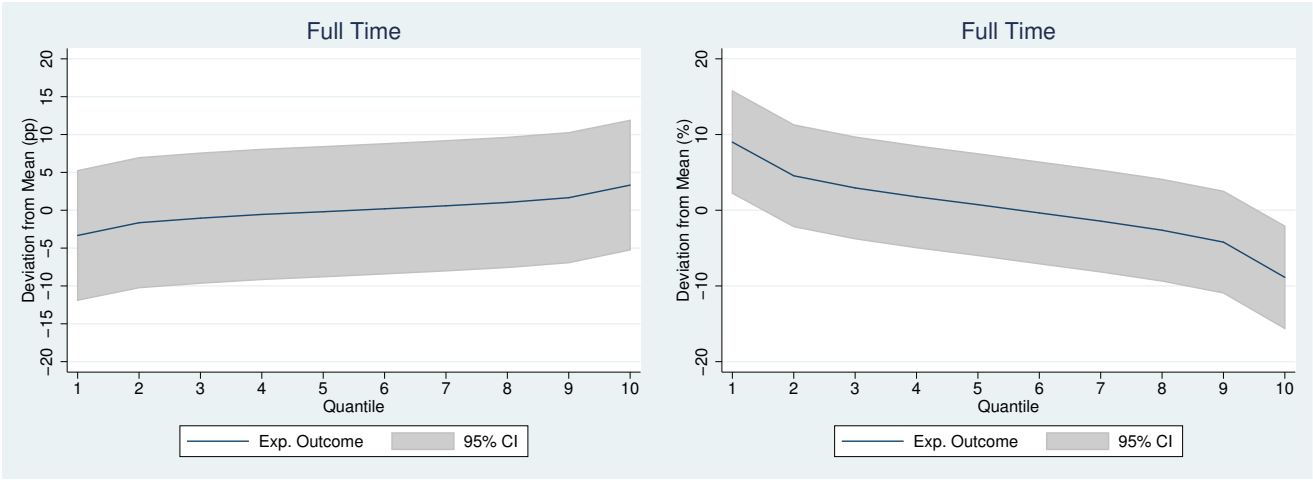
Figure E.2: Probability of Employment at Levels of Homosexuality
Men



Note: Simulated incomes and employment probabilities using the estimated parameters and 40,000 draws from each dimension of unobserved heterogeneity. *Source*: own calculations.

Figure E.3: Women

(a) Women: Probability of Being Employed (b) Women: Income (Censored)



Note: Simulated incomes and employment probabilities using the estimated parameters and 40,000 draws from each dimension of unobserved heterogeneity. *Source*: own calculations.

F Estimations Using Full Time Employment

Table F.1: Homosexuality Factor in a Roy Model with Selection Equations

	College†	Income		Employed†	
		NoCollege	*College*	*NoCollege*	*College*
IncomeFullTime4					
Overall					
Latent Homosex.	0.021	0.273	-3.077	0.000	-0.035
	(0.020)	(3.397)	(2.589)	(0.034)	(0.022)
Latent Skills	0.350***	33.182	36.756*	0.147	0.415***
	(0.129)	(20.183)	(19.080)	(0.216)	(0.155)
Males					
Latent Homosex.	0.047**	-1.047	-1.238	-0.032	-0.084***
	(0.019)	(2.996)	(2.793)	(0.035)	(0.025)
Latent Skills	0.331*	34.831	41.267	-0.116	0.660***
	(0.170)	(22.595)	(27.263)	(0.301)	(0.244)
Females					
Latent Homosex.	-0.012	0.372	-4.553	0.046	0.021
	(0.025)	(3.345)	(3.103)	(0.044)	(0.026)
Latent Skills	0.477***	8.148	29.693	0.316	0.247
	(0.173)	(23.676)	(22.790)	(0.279)	(0.194)

Note: *** p<0.01, ** p<0.05, * p<0.1. Controls used in the college choice equation are: age, age squared, gender (in the overall sample), municipality population, mother's and father's education, indicators of skipping school, taken out of classroom and police contact while teenager. Controls in the selection equations are: age, age squared, gender (in the overall sample), whether as a child the person lived with both parents, indicator for practicing sports while teenager and male labor force participation rate in the 25 to 39 age bracket in the municipality. Controls in the income equations are: age, age squared, gender (in the overall sample), father's SES, and mother's education. The College variable takes the value of 1 when the person has undertaken any kind of tertiary education in colleges (*høyskole*) or universities (*universitetet*). This includes three-year, four-year college studies or studies in the university with the same length. Total observations for overall, male and female full time sample: 2,218, 951 and 1,267. *Source*: Own calculations. †Coefficients, not marginal effects.

Although Table E.1 shows that homosexuality affects the chances of getting both *a* job and a full-time job, Table 7 reveals that the selection channels in each margin differ: while , it does so only to the college educated in the full-time job margin (by about 10 percentage points).

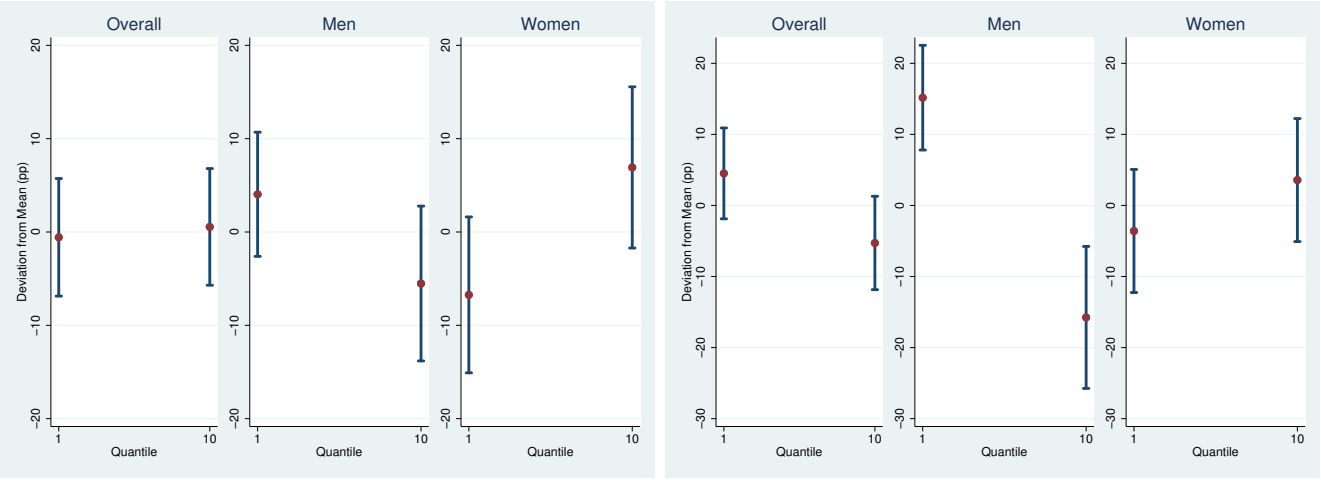
Table 7 also documents gender-specific effects. In fact, in Appendix E, I already showed that labor market discrimination for homosexual women works differently than the discrimination against homosexual men. The Roy model with endogenous education choices allows me to check the extent to which discrimination varies across levels of educational attainment at each gender. I find that the employment discrimination against homosexual men documented in Table E.1 is driven mainly by the employment discrimination against college-educated homosexual men. The gap balloons to 31 percentage points if we consider full-time employment. Consistent with previous results, I find no effect of latent homosexuality on the income earned on the job, regardless of the education choice.

The results in Table 7 indicate that the gap in the probability of employment against homosexual women documented in Table E.1 is due mainly to the discrimination faced by homosexuals without college education. It also shows that college-educated lesbians are the ones who face the

income penalty documented above. In fact, the simulated income earned once employed for each level of homosexuality factor presented in Figures F.2 show that employed homosexual women earn 10.7 percent less than the extreme heterosexual ones, and 11.37 percent if they are employed full time.

Figure F.1: Probability of Being Employed

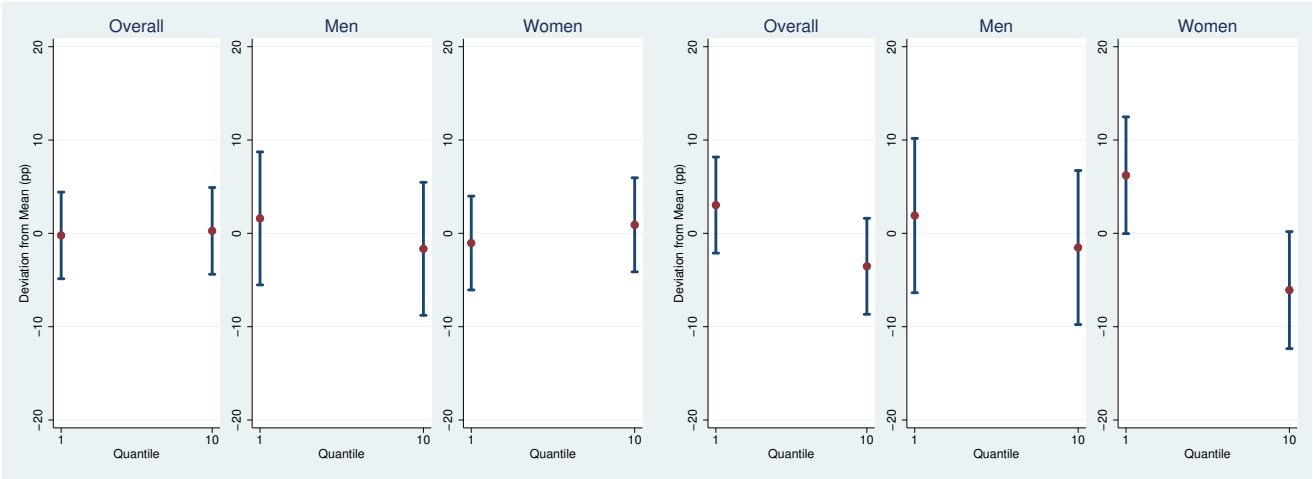
(a) Highest Education Level Achieved: No College (b) Highest Education Level Achieved: College



Note: Simulated incomes and employment probabilities using the estimated parameters and 40,000 draws from each dimension of unobserved heterogeneity. *Source: own calculations.*

Figure F.2: Income (Censored)

(a) Highest Education Level Achieved: No College (b) Highest Education Level Achieved: College



Note: Simulated incomes and employment probabilities using the estimated parameters and 40,000 draws from each dimension of unobserved heterogeneity. *Source:* own calculations.

G Goodness-of-Fit of Models with Homosexuality as a Factor

Table G.1: Goodness of Fit - Annual Income by Schooling level
Overall Sample

	College	**Employment**		**Income**	
	Means	Means	Std.Dev.	Means	Std.Dev.
Employed					
1. Model Without Endogenous Education Choice					
Actual		0.804	0.397	223.187	156.443
Model		0.802	0.398	222.466	156.761
2. Model With Endogenous Education Choice					
Actual	0.688				
Model	0.676				
No College					
Actual		0.840	0.367	244.936	145.460
Model		0.840	0.366	244.512	145.625
College					
Actual		0.771	0.420	223.312	161.473
Model		0.770	0.421	223.250	162.155
Full Time					
1. Model Without Endogenous Education Choice					
Actual		0.608	0.488	181.083	174.761
Model		0.609	0.488	180.793	174.368
2. Model With Endogenous Education Choice					
Actual	0.682				
Model	0.671				
No College					
Actual		0.641	0.480	200.978	174.080
Model		0.645	0.478	202.051	174.308
College					
Actual		0.573	0.495	176.562	177.097
Model		0.571	0.495	175.004	176.686

Note: *** p<0.01, ** p<0.05, * p<0.1 indicate the probability of rejecting the null hypothesis of the model predictions being different to the actual moments provided by the data. The model predictions are the result of 40,000 simulations of the estimated models. *Source*: Own calculations.

Table G.2: Goodness of Fit - Annual Income Gender Specific Models
Males

	College	**Employment**		**Income**	
	Means	Means	Std.Dev.	Means	Std.Dev.
Employed					
1. Model Without Endogenous Education Choice					
Actual		0.827	0.379	251.367	167.840
Model		0.823	0.381	249.887	168.638
2. Model With Endogenous Education Choice					
Actual	0.607				
Model	0.600				
No College					
Actual		0.882	0.323	287.180	149.636
Model		0.874	0.331	284.946	152.562
College					
Actual		0.778	0.416	243.243	177.536
Model		0.767	0.422	237.917	178.456
Full Time					
1. Model Without Endogenous Education Choice					
Actual		0.752	0.432	239.183	180.509
Model		0.750	0.433	238.542	181.420
2. Model With Endogenous Education Choice					
Actual	0.601				
Model	0.600				
No College					
Actual		0.831	0.375	278.003	164.301
Model		0.828	0.377	276.680	166.016
College					
Actual		0.684	0.465	225.199	189.382
Model		0.682	0.465	224.043	189.293

Note: *** p<0.01, ** p<0.05, * p<0.1 indicate the probability of rejecting the null hypothesis of the model predictions being different to the actual moments provided by the data. The model predictions are the result of 40,000 simulations of the estimated models. *Source*: Own calculations.

Table G.3: Goodness of Fit - Annual Income Gender Specific Models
Females

	College	**Employment**		**Income**	
	Means	Means	Std.Dev.	Means	Std.Dev.
Employed					
1. Model Without Endogenous Education Choice					
Actual		0.787	0.410	200.971	143.040
Model		0.783	0.412	200.329	143.820
2. Model With Endogenous Education Choice					
Actual	0.736				
Model	0.733				
No College					
Actual		0.794	0.405	196.027	122.846
Model		0.791	0.406	195.460	123.727
College					
Actual		0.769	0.421	211.830	149.295
Model		0.766	0.423	211.018	149.794
Full Time					
1. Model Without Endogenous Education Choice					
Actual		0.498	0.500	136.583	156.306
Model		0.499	0.500	135.705	155.606
2. Model With Endogenous Education Choice					
Actual	0.725				
Model	0.721				
No College					
Actual		0.438	0.497	118.913	143.839
Model		0.442	0.496	118.492	142.733
College					
Actual		0.504	0.500	145.991	162.094
Model		0.504	0.500	145.257	161.768

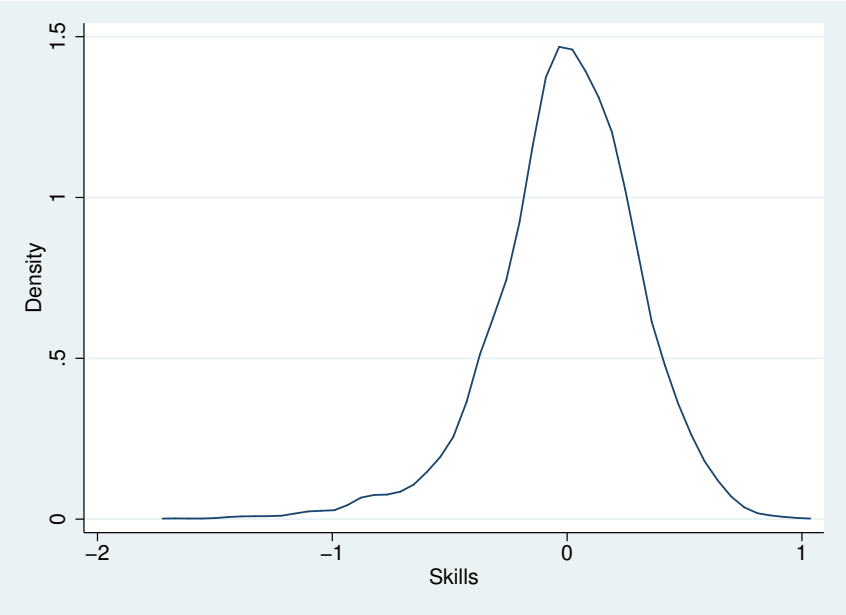
Note: *** p<0.01, ** p<0.05, * p<0.1 indicate the probability of rejecting the null hypothesis of the model predictions being different to the actual moments provided by the data. The model predictions are the result of 40,000 simulations of the estimated models. *Source*: Own calculations.

H Estimates of Skill Distributions and Loadings

This appendix presents the details of the results of estimating model (3) to identify the skill factors in different subsamples. Namely, the whole sample and homosexual and heterosexual subsamples according to the SI and the SP definitions. The estimations on the whole sample are the ones that allow the results presented in Section B. The estimations on the sexual-preference-specific subsamples according to each definition relate to the results presented in Sections 4.1 and 4.2.

H.1 Whole Sample

Figure H.1: Latent Skills Distribution (Whole Sample)



Note: Estimated from measurement systems presented in Table H.1. Estimations use the mixing of two normals. Distributional parameter are the following: $\mu_1 = 0.555$, $\sigma_1 = 0.393$, $\mu_2 = -3.264$, $\sigma_2 = 0.030$, and the mixing probability is 0.058. The mixing probability refers to the weight given to the second normal (i.e., the one described by parameters μ_2 and σ_2) in the mixture. The weight given to the first normal (i.e., the one described by parameters μ_1 and σ_1) in the mixture is given by one minus the mixing probability. Author's calculations.

Table H.1: Skills Measurement System Estimation (Whole Sample)

	Soc. Accept.	Sympathetic	Self-esteem	Leadership	Grades	Scholastic
Female	0.007	0.309***	-0.267***	-0.066***	0.339***	-0.136***
	(0.027)	(0.028)	(0.028)	(0.022)	(0.069)	(0.026)
Age	0.039***	0.032***	0.005	0.033***	0.178***	-0.027***
	(0.008)	(0.008)	(0.008)	(0.006)	(0.019)	(0.007)
Father: Univ	0.048	0.047	0.083*	0.107***	0.793***	0.216***
	(0.045)	(0.045)	(0.045)	(0.036)	(0.113)	(0.042)
Father: 3yrColl	0.029	0.071	0.017	0.063	0.922***	0.227***
	(0.048)	(0.049)	(0.049)	(0.039)	(0.123)	(0.045)
Father: Vocat	0.023	0.063	-0.044	0.100***	0.129	0.025
	(0.040)	(0.041)	(0.041)	(0.033)	(0.102)	(0.038)
Mother: Univ	0.036	0.040	0.017	0.035	0.296**	0.094*
	(0.059)	(0.059)	(0.059)	(0.048)	(0.148)	(0.055)
Num. Siblings	-0.014	-0.012	-0.006	-0.024**	-0.015	-0.020*
	(0.013)	(0.013)	(0.013)	(0.010)	(0.032)	(0.012)
LivesBothParents	0.081**	-0.012	0.107***	-0.028	0.413***	0.109***
	(0.035)	(0.034)	(0.034)	(0.027)	(0.085)	(0.031)
Father: HighSES	0.001	-0.019	0.032	0.053*	0.512***	0.090***
	(0.034)	(0.034)	(0.034)	(0.027)	(0.085)	(0.031)
Mother: HighSES	0.097*	0.117**	0.060	0.107**	0.079	0.119**
	(0.057)	(0.057)	(0.057)	(0.046)	(0.142)	(0.052)
Constant	-0.659***	-0.657***	-0.021	-0.453***	-3.454***	0.336***
	(0.128)	(0.126)	(0.127)	(0.101)	(0.312)	(0.116)
Skills	1.865***	1.367***	1.419***	0.742***	0.774***	1
	(0.113)	(0.084)	(0.075)	(0.050)	(0.131)	.

Note: *Univ* (*universitetet*) includes program of four or more years. *3yrColl* corresponds to the *høyskole*. *High SES* describes parents who have a high administrative or academic job. *Soc. Accept.* (social acceptance), *Sympathetic* (capable of building close friendships), *Self-esteem*, *Leadership* and *Scholastic Competence* are measures that come from SPPA scales (Harter, 1988). *Grades* correspond to a demeaned aggregation of school grades (Math, Norwegian and English). Observations: 2,595. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. *Source*: Own calculations.

H.2 By Sexual Preference

Table H.2: Skills Measurement System Estimation (Homosexual SI)

	Soc. Accept.	Sympathetic	Self-esteem	Leadership	Grades	Scholastic
Female	0.119**	0.230***	-0.156**	-0.108*	0.244	-0.171***
	(0.050)	(0.065)	(0.068)	(0.056)	(0.163)	(0.062)
Age	0.005	0.015	-0.018	0.020	0.149***	-0.035**
	(0.009)	(0.015)	(0.016)	(0.013)	(0.038)	(0.015)
Father: Univ	0.009	-0.013	0.066	0.187***	0.634***	0.270***
	(0.061)	(0.084)	(0.088)	(0.072)	(0.212)	(0.081)
Father: 3yrColl	0.095*	0.131	0.078	0.047	0.786***	0.170*
	(0.054)	(0.088)	(0.093)	(0.077)	(0.228)	(0.087)
Father: Vocat	0.119*	0.157*	-0.122	0.206***	0.017	-0.022
	(0.069)	(0.083)	(0.085)	(0.070)	(0.202)	(0.078)
Mother: Univ	-0.018	-0.110	-0.005	0.178**	0.241	0.023
	(0.052)	(0.101)	(0.109)	(0.090)	(0.268)	(0.102)
Num. Siblings	-0.036***	-0.045*	-0.006	-0.024	-0.032	-0.016
	(0.014)	(0.023)	(0.024)	(0.020)	(0.060)	(0.023)
LivesBothParents	-0.095**	-0.178***	0.019	-0.148***	0.658***	0.106*
	(0.037)	(0.061)	(0.065)	(0.054)	(0.160)	(0.061)
Father: HighSES	0.024	0.069	0.074	-0.015	0.366**	0.083
	(0.043)	(0.064)	(0.068)	(0.056)	(0.165)	(0.063)
Mother: HighSES	0.013	0.098	0.074	-0.048	0.189	0.086
	(0.061)	(0.098)	(0.103)	(0.086)	(0.253)	(0.096)
Constant	-0.098	-0.207	0.248	-0.111	-2.902***	0.532**
	(0.140)	(0.236)	(0.252)	(0.209)	(0.617)	(0.235)
Skills	4.541***	2.549***	1.729***	1.036***	-0.243	1
	(0.824)	(0.494)	(0.364)	(0.244)	(0.481)	.

Note: *Univ* (*universitetet*) includes program of four or more years. *3yrColl* corresponds to the *høyskole*. *High SES* describes parents who have a high administrative or academic job. *Soc. Accept.* (social acceptance), *Sympathetic* (capable of building close friendships), *Self-esteem*, *Leadership* and *Scholastic Competence* are measures that come from SPPA scales (Harter, 1988). *Grades* correspond to a demeaned aggregation of school grades (Math, Norwegian and English). Observations: 568. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. *Source*: Own calculations.

Table H.3: Skills Measurement System Estimation (Heterosexual SI)

	Soc. Accept.	Sympathetic	Self-esteem	Leadership	Grades	Scholastic
Female	-0.006	0.324***	-0.274***	-0.066***	0.322***	-0.144***
	(0.031)	(0.031)	(0.031)	(0.025)	(0.080)	(0.029)
Age	0.046***	0.036***	0.011	0.038***	0.191***	-0.024***
	(0.009)	(0.009)	(0.009)	(0.007)	(0.023)	(0.008)
Father: Univ	0.057	0.049	0.100*	0.082*	0.825***	0.197***
	(0.052)	(0.052)	(0.052)	(0.042)	(0.134)	(0.048)
Father: 3yrColl	0.047	0.081	0.024	0.076*	0.978***	0.271***
	(0.057)	(0.057)	(0.057)	(0.046)	(0.147)	(0.053)
Father: Vocat	0.009	0.037	-0.010	0.064*	0.159	0.044
	(0.046)	(0.046)	(0.046)	(0.037)	(0.119)	(0.043)
Mother: Univ	0.055	0.102	0.015	-0.013	0.316*	0.106*
	(0.069)	(0.069)	(0.069)	(0.055)	(0.177)	(0.064)
Num. Siblings	-0.011	-0.004	-0.007	-0.023*	-0.007	-0.026*
	(0.015)	(0.015)	(0.015)	(0.012)	(0.038)	(0.014)
LivesBothParents	0.118***	0.026	0.121***	0.015	0.332***	0.102***
	(0.040)	(0.040)	(0.040)	(0.032)	(0.101)	(0.037)
Father: HighSES	-0.004	-0.030	0.017	0.073**	0.577***	0.095***
	(0.039)	(0.039)	(0.039)	(0.031)	(0.100)	(0.036)
Mother: HighSES	0.118*	0.110*	0.060	0.145***	0.024	0.129**
	(0.067)	(0.067)	(0.067)	(0.054)	(0.171)	(0.062)
Constant	-0.781***	-0.750***	-0.101	-0.575***	-3.651***	0.303**
	(0.146)	(0.144)	(0.145)	(0.115)	(0.365)	(0.133)
Skills	1.667***	1.238***	1.373***	0.686***	0.902***	1
	(0.099)	(0.078)	(0.074)	(0.049)	(0.136)	.

Note: *Univ* (*universitetet*) includes program of four or more years. *3yrColl* corresponds to the *høyskole*. *High SES* describes parents who have a high administrative or academic job. *Soc. Accept.* (social acceptance), *Sympathetic* (capable of building close friendships), *Self-esteem*, *Leadership* and *Scholastic Competence* are measures that come from SPPA scales (Harter, 1988). *Grades* correspond to a demeaned aggregation of school grades (Math, Norwegian and English). Observations: 2,001. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. *Source*: Own calculations.

Table H.4: Skills Measurement System Estimation (Homosexual SP)

	Soc. Accept.	Sympathetic	Self-esteem	Leadership	Grades	Scholastic
Female	0.132*	0.353***	-0.099	-0.053	0.126	-0.144**
	(0.072)	(0.071)	(0.069)	(0.059)	(0.164)	(0.065)
Age	0.009	-0.003	-0.050***	0.005	0.162***	-0.048***
	(0.019)	(0.019)	(0.018)	(0.016)	(0.043)	(0.017)
Father: Univ	0.025	0.011	0.169	0.266***	0.323	0.358***
	(0.112)	(0.109)	(0.106)	(0.090)	(0.250)	(0.099)
Father: 3yrColl	0.040	0.094	0.053	0.131	0.656**	0.144
	(0.116)	(0.113)	(0.111)	(0.095)	(0.265)	(0.104)
Father: Vocat	0.160	0.211**	0.036	0.222***	0.027	0.016
	(0.104)	(0.101)	(0.097)	(0.082)	(0.225)	(0.089)
Mother: Univ	0.084	-0.064	0.147	0.125	0.477	0.038
	(0.126)	(0.125)	(0.123)	(0.106)	(0.295)	(0.116)
Num. Siblings	0.009	-0.022	0.026	-0.035	-0.005	0.003
	(0.025)	(0.025)	(0.025)	(0.021)	(0.060)	(0.023)
LivesBothParents	-0.006	-0.076	0.047	-0.139**	0.803***	0.267***
	(0.072)	(0.071)	(0.069)	(0.060)	(0.167)	(0.065)
Father: HighSES	-0.010	0.039	-0.080	-0.042	0.450**	0.024
	(0.090)	(0.087)	(0.084)	(0.071)	(0.195)	(0.077)
Mother: HighSES	0.239**	0.187	0.133	0.062	0.256	0.082
	(0.120)	(0.117)	(0.114)	(0.097)	(0.269)	(0.106)
Constant	-0.377	-0.207	0.583**	0.020	-3.430***	0.459*
	(0.299)	(0.291)	(0.284)	(0.244)	(0.677)	(0.266)
Skills	2.777***	2.198***	1.762***	0.890***	0.448	1
	(0.449)	(0.359)	(0.281)	(0.173)	(0.373)	.

Note: *Univ* (*universitetet*) includes program of four or more years. *3yrColl* corresponds to the *høyskole*. *High SES* describes parents who have a high administrative or academic job. *Soc. Accept.* (social acceptance), *Sympathetic* (capable of building close friendships), *Self-esteem*, *Leadership* and *Scholastic Competence* are measures that come from SPPA scales (Harter, 1988). *Grades* correspond to a demeaned aggregation of school grades (Math, Norwegian and English). Observations: 507. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. *Source*: Own calculations.

Table H.5: Skills Measurement System Estimation (Heterosexual SP)

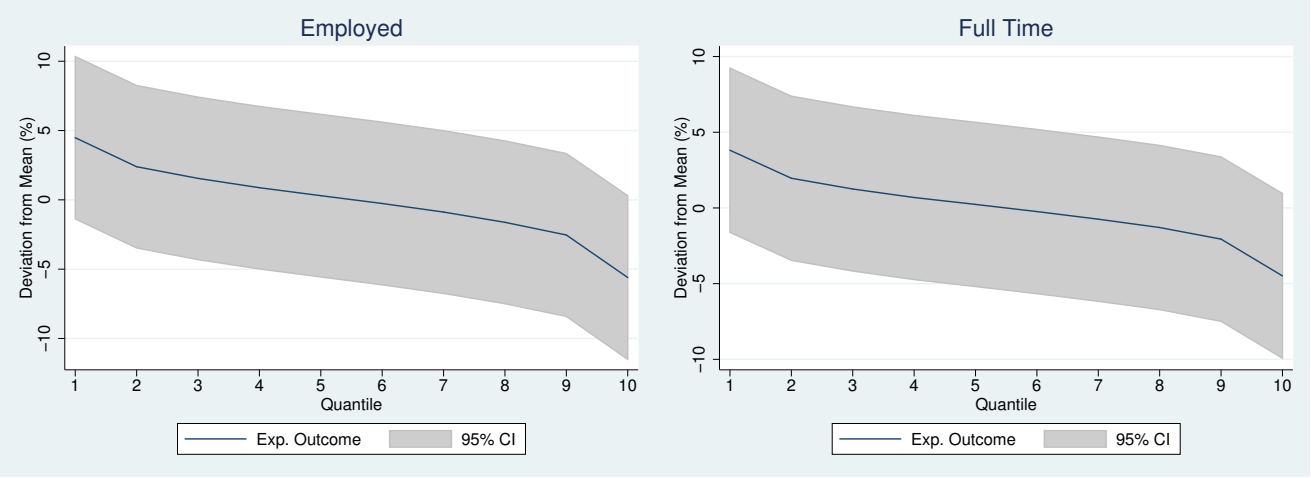
	Soc. Accept.	Sympathetic	Self-esteem	Leadership	Grades	Scholastic
Female	0.000	0.311***	-0.286***	-0.067***	0.409***	-0.129***
	(0.030)	(0.030)	(0.031)	(0.024)	(0.078)	(0.028)
Age	0.045***	0.039***	0.018**	0.039***	0.178***	-0.022***
	(0.008)	(0.009)	(0.009)	(0.007)	(0.022)	(0.008)
Father: Univ	0.064	0.047	0.081	0.078**	0.890***	0.192***
	(0.049)	(0.050)	(0.050)	(0.040)	(0.127)	(0.046)
Father: 3yrColl	0.040	0.069	0.023	0.049	0.986***	0.266***
	(0.052)	(0.054)	(0.054)	(0.043)	(0.138)	(0.050)
Father: Vocat	0.003	0.034	-0.054	0.071**	0.153	0.043
	(0.044)	(0.045)	(0.045)	(0.036)	(0.114)	(0.041)
Mother: Univ	0.029	0.086	-0.014	0.011	0.270	0.104*
	(0.065)	(0.066)	(0.067)	(0.053)	(0.170)	(0.062)
Num. Siblings	-0.023	-0.008	-0.020	-0.021*	-0.013	-0.031**
	(0.014)	(0.015)	(0.015)	(0.012)	(0.038)	(0.014)
LivesBothParents	0.112***	0.009	0.127***	0.010	0.290***	0.060*
	(0.038)	(0.039)	(0.039)	(0.031)	(0.098)	(0.036)
Father: HighSES	-0.007	-0.027	0.045	0.067**	0.528***	0.103***
	(0.037)	(0.037)	(0.038)	(0.030)	(0.095)	(0.035)
Mother: HighSES	0.087	0.118*	0.073	0.126**	0.020	0.143**
	(0.063)	(0.065)	(0.065)	(0.052)	(0.165)	(0.060)
Constant	-0.732***	-0.759***	-0.187	-0.579***	-3.380***	0.331**
	(0.138)	(0.139)	(0.141)	(0.111)	(0.352)	(0.129)
Skills	1.688***	1.235***	1.351***	0.713***	0.821***	1
	(0.105)	(0.081)	(0.075)	(0.051)	(0.138)	.

Note: *Univ* (*universitetet*) includes program of four or more years. *3yrColl* corresponds to the *høyskole*. *High SES* describes parents who have a high administrative or academic job. *Soc. Accept.* (social acceptance), *Sympathetic* (capable of building close friendships), *Self-esteem*, *Leadership* and *Scholastic Competence* are measures that come from SPPA scales (Harter, 1988). *Grades* correspond to a demeaned aggregation of school grades (Math, Norwegian and English). Observations: 2,066. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. *Source*: Own calculations.

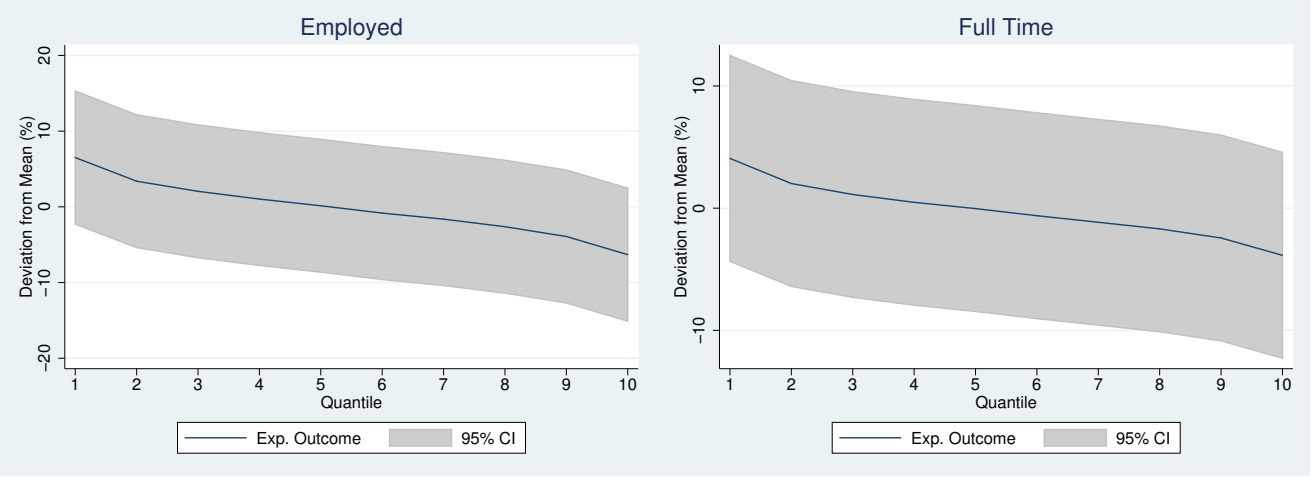
I Additional Simulations of the Models with Homosexuality Factor

Figure I.1: Expected Income (Censored) at Levels of Homosexuality

(a) Complete Sample



(b) Males



Note: Simulated incomes and employment probabilities using the estimated parameters and 40,000 draws from each dimension of unobserved heterogeneity. *Source*: Own calculations.

J Additional Tables and Graphs

Table J.1: Gap in College Attendance

	(1)		(2)		(3)		(4)	
Defini.	Coeff.	Std.Err.	Coeff.	Std.Err.	Coeff.	Std.Err.	Coeff.	Std.Err.
SI	0.068***	(0.025)	0.032	(0.025)	0.041	(0.048)	0.024	(0.048)
SI×*female*					-0.013	(0.056)	0.001	(0.056)
SP	-0.008	(0.026)	-0.028	(0.026)	0.039	(0.046)	0.023	(0.046)
SP×*female*					-0.099*	(0.055)	-0.051	(0.055)
SX	-0.016	(0.032)	-0.016	(0.031)	0.067	(0.049)	0.072	(0.049)
SX×*female*					-0.137**	(0.063)	-0.139**	(0.063)
Age & gender			X		X		X	
Test Scores							X	

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The College variable takes the value of 1 when the person has undertaken any kind of tertiary education in colleges (*høyskole*) or universities (*universitetet*). This includes three-year, four-year college studies or studies in the university with the same length. Age enters in a quadratic polynomial. The estimations with test scores include six scores, one related to school grades (Math, Norwegian and English), two of them relate to agreeableness and the ease to make friends, the others relate to positivism, leadership and scholastic competence. All of them collected in the first wave of the survey. Sample comprises the economically active by wave 4 to be comparable to structural model results. *Source*: Own calculations.

Table J.2: Homosexuality, Education and Labor Market Choices

	Active	Employed	Full Time	White Collar	Has ISCO
Complete Sample					
Latent Homosex.	-0.077***	-0.072***	-0.027	-0.011	0.001
	(0.020)	(0.019)	(0.017)	(0.021)	(0.019)
Latent Skills	0.318**	0.427***	0.379***	0.914***	0.063
	(0.140)	(0.131)	(0.119)	(0.142)	(0.125)
Obs.	2417	2417	2417	1900	2687
Males					
Latent Homosex.	-0.079***	-0.068***	-0.084***		
	(0.023)	(0.020)	(0.019)		
Latent Skills	0.141	0.295	0.360**		
	(0.207)	(0.183)	(0.172)		
Obs.	1027	1047	1047		
Females					
Latent Homosex.	-0.043*	-0.046*	0.036*		
	(0.026)	(0.025)	(0.022)		
Latent Skills	0.358**	0.485***	0.343**		
	(0.176)	(0.168)	(0.152)		
Obs.	1370	1370	1370		

Note: *** p<0.01, ** p<0.05, * p<0.1. The controls used in all the equation and not presented in the Table are: age, age squared, gender, dummies for father's education, whether as a child the person lived with one parent, a remarried parent or with no parents. Standard error in parentheses. *Source*: Own calculations.

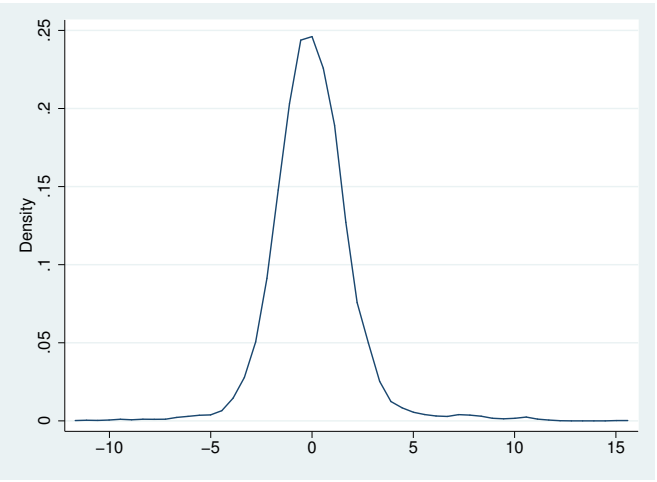
Table J.3: Difference in Observable Characteristics Between Heterosexuals and Homosexuals

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
			Father's Education			Father's SES			When Child Lived With:		
	Age	Gender	Univers	3YrColl	Vocat.	High	Skilled	Sibling	OnePar	RemPar	NoPar
SI	-0.050	0.281***	0.042**	0.015	-0.010	0.053**	-0.022	-0.032	0.027*	0.024**	0.001
	(0.095)	(0.025)	(0.018)	(0.015)	(0.018)	(0.023)	(0.018)	(0.058)	(0.016)	(0.011)	(0.011)
SP	-0.317***	0.200***	0.014	-0.022	-0.019	-0.021	-0.019	0.078	0.042**	0.023*	0.004
	(0.099)	(0.026)	(0.019)	(0.015)	(0.019)	(0.024)	(0.019)	(0.060)	(0.017)	(0.012)	(0.011)
SX	-0.121	0.071**	0.067***	-0.014	-0.027	0.039	-0.050**	0.031	0.060***	0.027*	0.024*
	(0.121)	(0.032)	(0.023)	(0.019)	(0.023)	(0.029)	(0.023)	(0.073)	(0.021)	(0.014)	(0.014)
SX+	-0.266***	0.165***	0.014	-0.023	-0.029	-0.020	-0.019	0.071	0.057***	0.021*	0.012
	(0.102)	(0.027)	(0.020)	(0.016)	(0.019)	(0.025)	(0.019)	(0.062)	(0.018)	(0.012)	(0.012)

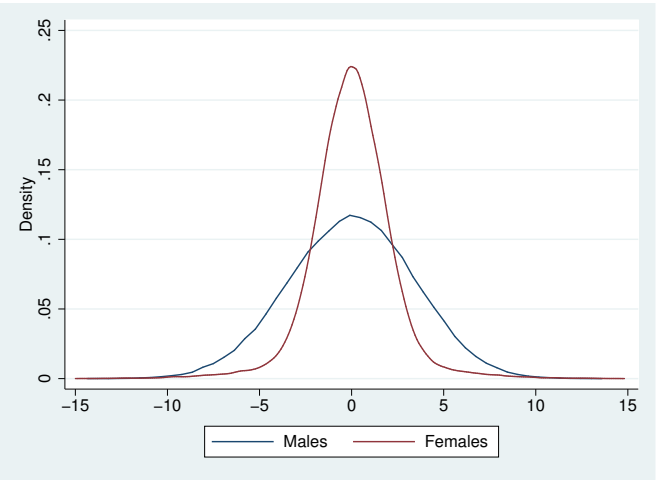
Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. All estimations were drawn from different regressions with the dependent variable being the homosexuality indicator. *Univers* estimates whether the father went to the university. *3YRColl* estimates whether the father went to a three-year college. *Vocat* estimates whether the father undertook vocational training. *OnePat* estimates if the respondent lived with only one of the two parents. *RemPar* estimates if the respondent lived with one father that remarried. *NoPar* estimates whether the respondent lived without her parents during childhood. *Source*: Own calculations.

Figure J.1: Homosexuality Factor

(a) Full Sample



(b) Gender Specific Samples



Note: Estimated from measurement systems presented in Table 6. Estimations use the mixing of two normals. Left density is estimated for the complete sample. $\mu_1 = 0.969$, $\sigma_1 = 4.318$, $\mu_2 = -0.112$, $\sigma_2 = 1.429$, and the mixing probability is 0.104. The right figure presents the gender-specific densities. For males: $\mu_1 = 0.825$, $\sigma_1 = 3.115$, $\mu_2 = -2.274$, $\sigma_2 = 3.089$, and the mixing probability is 0.734. For women: $\mu_1 = -0.033$, $\sigma_1 = 4.441$, $\mu_2 = 0.005$, $\sigma_2 = 1.617$, and the mixing probability is 0.133. The mixing probability refers to the weight given to the second normal (i.e., the one described by parameters μ_2 and σ_2) in every mixture. The weight given to the first normal (i.e., the one described by parameters μ_1 and σ_1) in every mixture is given by one minus the mixing probability. *Source*: Own calculations.