Does it really get better?

A response to Carpenter and Eppink (2017)

Duc Hien Nguyen
Econ 781-UMass Amherst

September 20, 2020

Original paper

- Carpenter, C.S and Eppink S.T, "Does it get better? Recent estimates of sexual orientation and earnings in the United States", Southern Economic Journal 2017, 84(2), 416-441.
- 2013-2015 National Health Interivew Survey (NHIS) data
- "These data also show—**for the first time in the literature**—that self-identified gay men also earn significantly **more** than comparable heterosexual men, a difference on the order of **10% of annual earnings.**"
- "We...suggest that reduced discrimination and changing patterns of household specialization are unlikely to be the primary mechanisms."

Literature on SOGI discrimination

- Klawitter (2015) meta-analysis: 34 studies of gay men and 29 studies of lesbian women
- Gay penalty of -11% (range -30% to 0%), standard error 2%
- Lesbian premium of 9% (range -25% to +45%), standard error 2%

Recent findings

Table: Recent findings of annual gay and lesbian earnings gap

Study	Data sources	Log annual earnings/income	
		Men	Women
Waite and Denier (2015)[ii]	Canadian Census 2006	-0.074***	0.091***
		(0.011)	(0.011)
Cerf (2016)[i]	Canadian CHS 2003-9	-0.090	0.027
		(0.057)	(0.044)
Cerf (2016)[ii]	Canadian CHS 2003-9	-0.130**	0.079*
		(0.066)	(0.061)
Carpenter and Eppink (2017)	NHIS 2013-2015	0.097**	0.086**
		(0.038)	(0.043)
Jepsen and Jepsen (2017)[ii]	ACS 2007-2011	-0.111***	0.107***
		(0.004)	(0.004)
Martel and Hansen (2017)	GSS 2008-2012		-0.339
			(-0.305)
Martel (2018)[ii]	ACS 2013-2015	-0.115***	
		(0.009)	
Dilmaghani (2018)	CADUMS 2008-2012	0.045	0.116*
		(0.066)	(0.058)

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors in bracket.
(i) Single/never married sample, (ii) married/partnered sample

Replication process: Summary stats

- Getting the data and clean: NHIS
- Descriptive statistics
- Charts, charts, and more charts

Descriptive statistics

Table: Descriptive Statistics, main SOGI categories

Variable	Heterosexual women	Bisexual women	Lesbian women	Heterosexual men	Bisexual men	Gay men
Age	43.5	36.7	43.1	43.2	38.7	42.0
	(11.1)	(10.1)	(10.4)	(11.1)	(11.4)	(10.7)
BA degree or more	0.409	0.468	0.518	0.364	0.436	0.495
	(0.492)	(0.499)	(0.500)	(0.481)	(0.496)	(0.500)
Associate degree	0.144	0.107	0.143	0.122	0.130	0.086
	(0.352)	(0.309)	(0.350)	(0.327)	(0.336)	(0.281)
Some college	0.182	0.260	0.154	0.167	0.218	0.206
	(0.386)	(0.439)	(0.361)	(0.373)	(0.413)	(0.404)
High school degree	0.191	0.114	0.161	0.241	0.138	0.177
	(0.393)	(0.317)	(0.367)	(0.428)	(0.345)	(0.381)
Less than HS	0.072	0.051	0.024	0.104	0.078	0.036
	(0.259)	(0.220)	(0.154)	(0.305)	(0.268)	(0.187)
Less than HS (2)	0.059	0.036	0.017	0.084	0.028	0.031
	(0.236)	(0.186)	(0.129)	(0.277)	(0.164)	(0.174)
White	0.784	0.829	0.812	0.812	0.837	0.855
	(0.412)	(0.376)	(0.391)	(0.390)	(0.370)	(0.352)
Sample size	22,337	252	426	21,444	118	540

Standard deviation in bracket

Descriptive statistics (cont.)

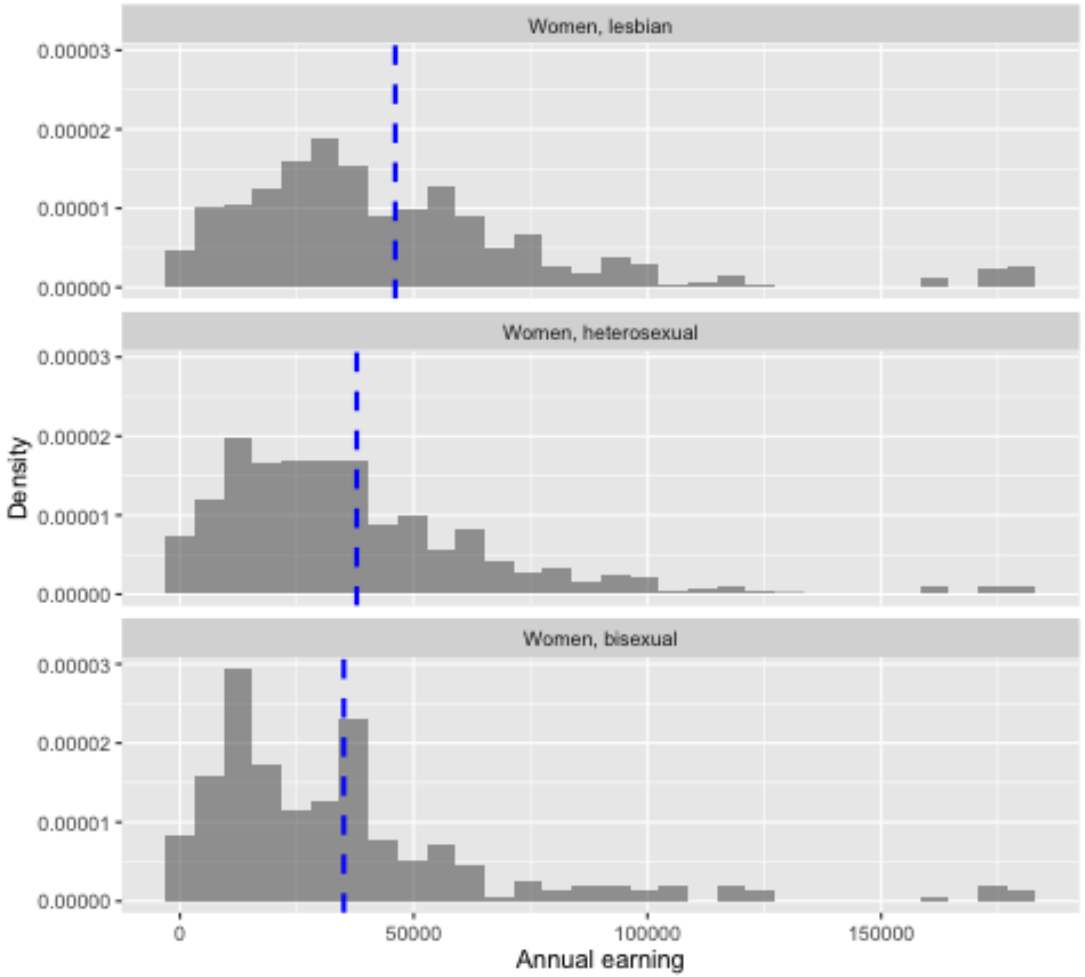
Table: Descriptive Statistics, main SOGI categories

Variable	Heterosexual women	Bisexual women	Lesbian women	Heterosexual men	Bisexual men	Gay men
Partnered	0.668	0.462	0.661	0.744	0.406	0.474
	(0.471)	(0.499)	(0.473)	(0.436)	(0.491)	(0.499)
Having children	0.471	0.339	0.275	0.443	0.302	0.078
	(0.499)	(0.473)	(0.447)	(0.497)	(0.459)	(0.268)
Northeast	0.174	0.113	0.184	0.164	0.126	0.189
	(0.379)	(0.317)	(0.387)	(0.371)	(0.332)	(0.391)
Midwest	0.230	0.235	0.202	0.237	0.247	0.152
	(0.421)	(0.424)	(0.401)	(0.425)	(0.431)	(0.359)
South	0.369	0.399	0.352	0.362	0.310	0.367
	(0.482)	(0.490)	(0.478)	(0.481)	(0.463)	(0.482)
West	0.227	0.253	0.262	0.237	0.317	0.292
	(0.419)	(0.435)	(0.440)	(0.425)	(0.465)	(0.455)
Annual Earnings	39,909.86	38,821.83	47,038.38	57,054.93	49,802.38	59,636.94
	(32,919.68)	(36,472.95)	(36,481.62)	(42,904.07)	(42,571.15)	(45,041.91)
Full-time worker	0.732	0.737	0.812	0.871	0.731	0.843
	(0.443)	(0.440)	(0.390)	(0.335)	(0.443)	(0.364)
Sample size	22,337	252	426	21,444	118	540

Standard deviation in bracket

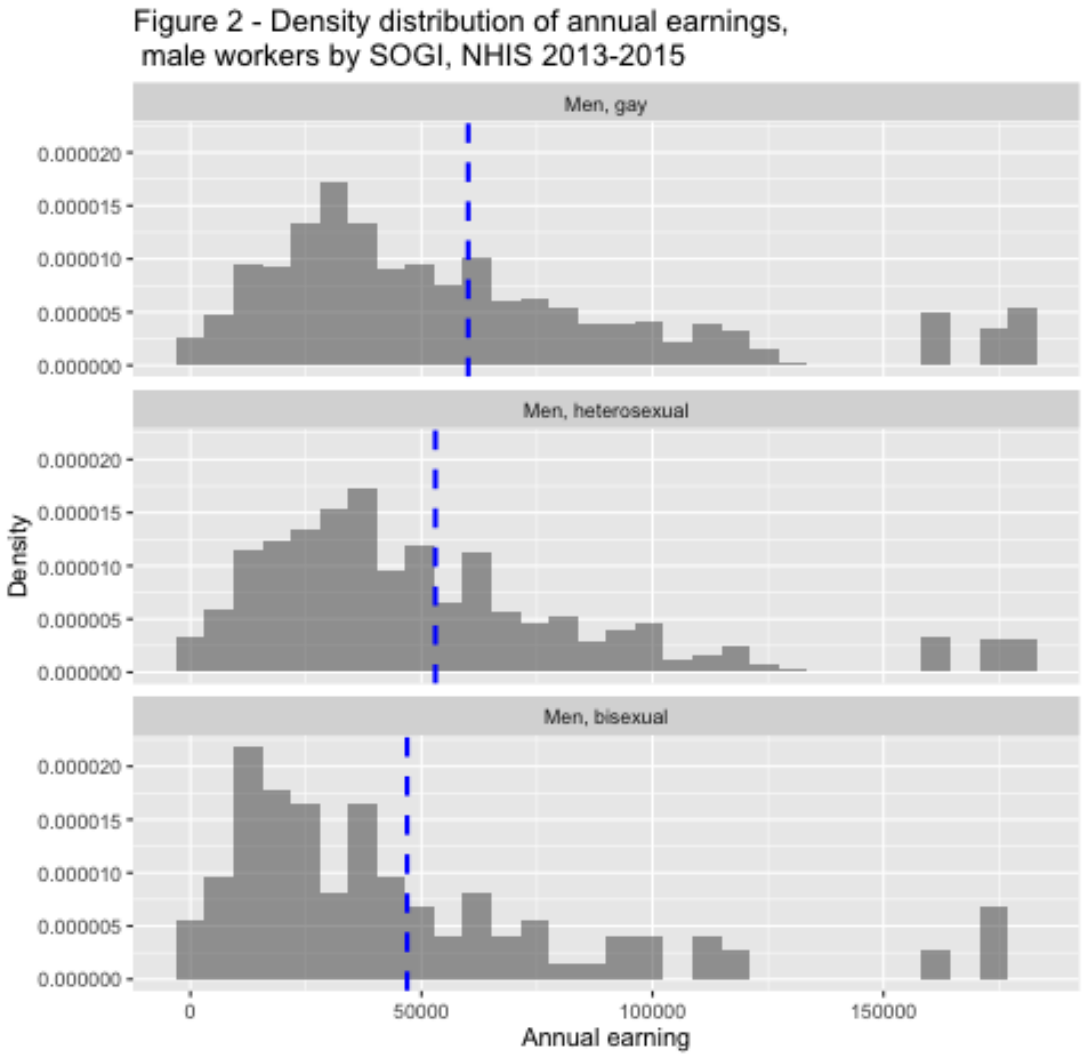
Earnings distribution: Female workers

Figure 1 - Density distribution of annual earnings, female workers by SOGI, NHIS 2013-2015



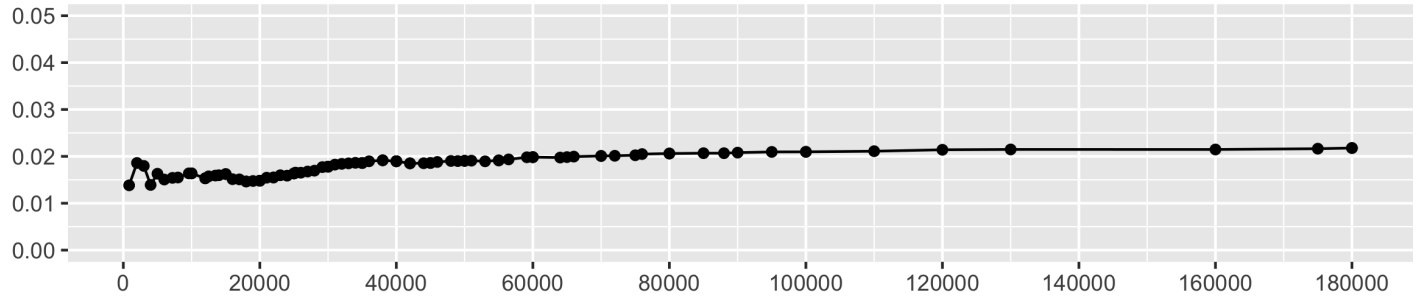
NB:The top 5% eamings are topcoded. Vertical lines correpond to mean earnings

Earnings distribution: Male workers

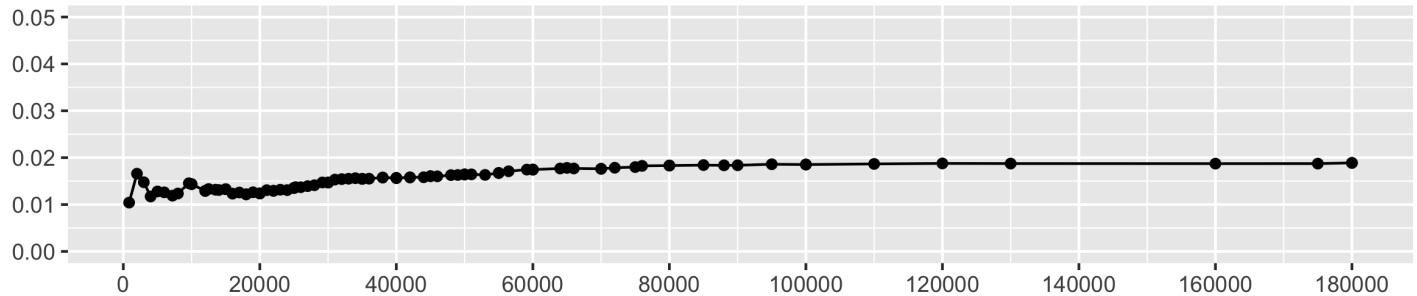


Sexual minorities along earnings distribution

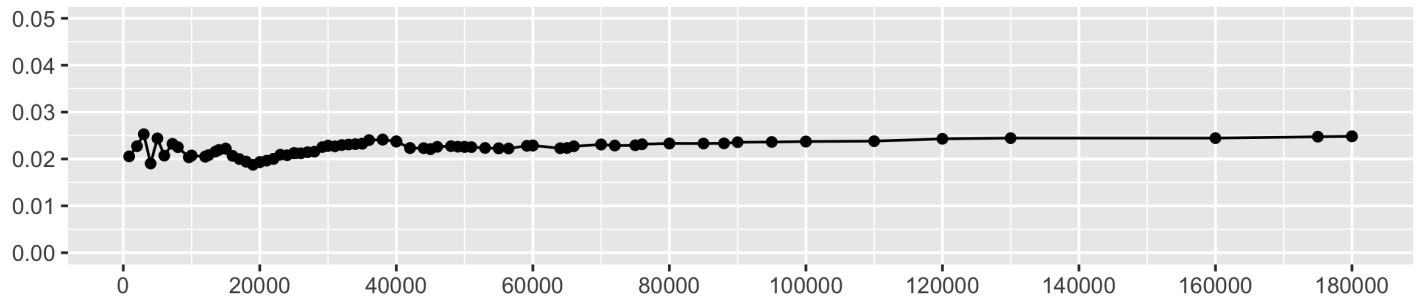
Figure 4 - Ratio of homosexual to heterosexual workers by earning percentile



(a) Both male and female

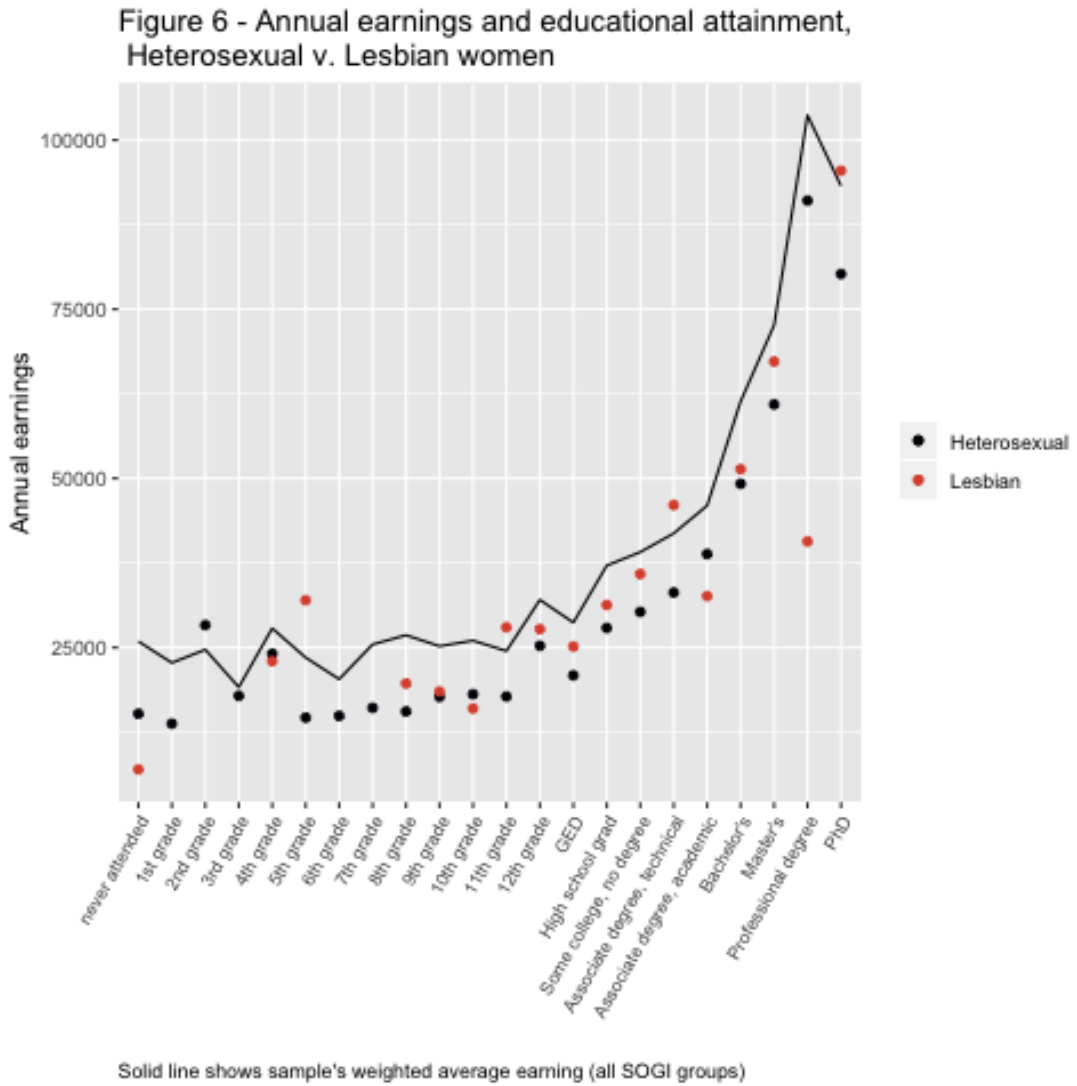


(b) Female only

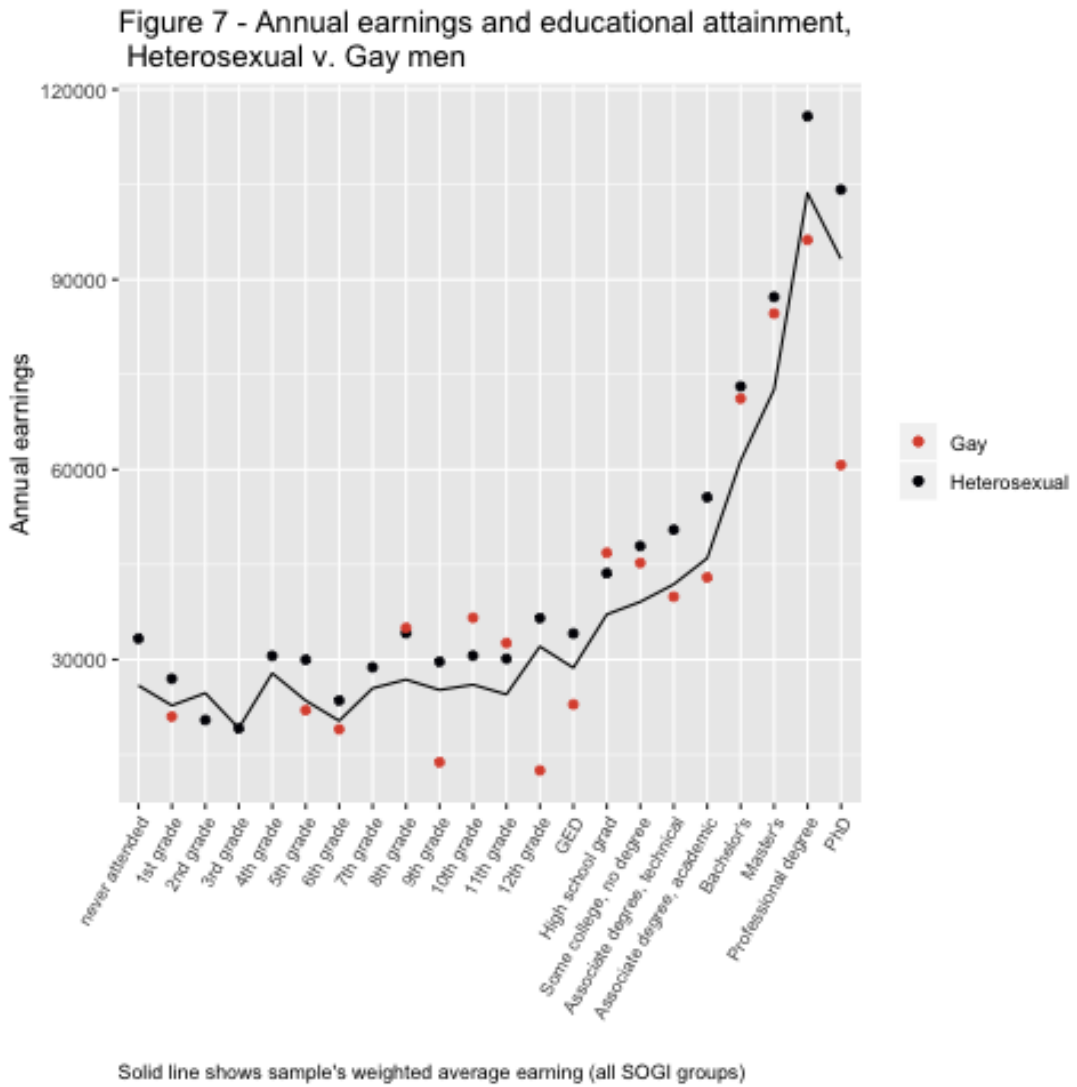


(c) Male only

SOGI wage gap: Education



SOGI wage gap: Education



Regression model

- Linear probability employment model

$$\text{EMP}_i = \alpha + \beta_1 SOGI_i + \beta_2 X_i + \epsilon_i \quad (1)$$

- Log-earning regression model

$$\text{Log Earnings}_i = \alpha + \beta_1 SOGI_i + \beta_2 X_i + \epsilon_i \quad (2)$$

- Separate for female and male workers, FT employed only

Log-earnings regression

$$\text{Log Earnings}_i = \alpha + \beta_1 SOGI_i + \beta_2 X_i + \epsilon_i$$

- $LogEarnings_i$: the natural log of annual earnings (FT workers)
- $SOGI_i$: vector of sexual orientation dummies
- X_i: vector of controls that includes (i) month and year of survey dummies and (ii) demographic characteristics: age and age squared, races, ethnicity, education, relationship/marital status, any children 0–5 in the household, any children 6–17 in the household, and region
- X_i also includes (iii) job characteristics: years of job tenure and its square; dummy variables for firm size, sector of employment, 24 industry dummies, 26 occupation dummies, and dummy variables for whether workers' personal earnings or job tenure responses are topcoded

Replicated result

Table: Earnings regression - replicating Carpenter and Eppink

	Dependent variable: Log annual earnings			
	Female, baseline	Male, baseline	Female, full model	Male, full model
	(1)	(2)	(3)	(4)
Homosexual	0.159***	0.078	0.087**	0.095**
	(0.052)	(0.055)	(0.043)	(0.037)
Bisexual	−0.083	−0.099	−0.028	−0.009
	(0.112)	(0.105)	(0.091)	(0.068)
Controls included:				
Month & year interview dummies	Yes	Yes	Yes	Yes
Demographics			Yes	Yes
Job characteristics			Yes	Yes
Observations	17,255	19,387	17,224	19,336
R^2	0.003	0.016	0.354	0.349

Note: *p<0.1; **p<0.05; ***p<0.01. Robust standard errors in bracket.

Robustness check

Table: Gay earnings gap - robustness check

Specifications		Coef. on Gay dummy
	(A)	
Baseline (interviews dummy only)	$N_{gay} = 430$	0.070
		(0.049)
Baseline + demographic	$N_{gay} = 430$	0.063
		(0.046)
Full model	$N_{gay} = 430$	0.088**
		(0.038)
	(B)	
Young - Aged 25-44	$N_{gay} = 248$	0.032
		(0.054)
Old - Aged 45-64	$N_{gay} = 182$	0.170***
		(0.053)
Partnered workers	$N_{gay} = 145$	0.040
		(0.050)
Single workers	$N_{gay} = 256$	0.159**
		(0.065)
Workers with less than BA degrees	$N_{gay} = 193$	0.087
		(0.062)
Workers with BA degrees or more	$N_{gay} = 237$	0.105**
		(0.053)
Public sector	$N_{gay} = 65$	0.090
		(0.058)
Private sector	$N_{gay} = 365$	0.096**
		(0.043)
At least 500 workers at firms	$N_{gay} = 106$	0.150***
		(0.052)
Less than 500 workers at firms	$N_{gay} = 324$	0.072
		(0.048)
	(C)	
No occupation dummies	$N_{gay} = 430$	0.093**
		(0.039)
No industry dummies	$N_{gay} = 430$	0.081*
		(0.044)
No occupation and industry dummies	$N_{gay} = 430$	0.056
		(0.041)

Note: *p<0.1; **p<0.05; ***p<0.01. Robust standard errors in bracket.
Unless specified, model used is the full saturated earning regression

Taking stock

- Succesfully replicated summary statistics and main results
- The gay male premium in annual earnings seems to be consistent across sub-sample
- Especially among single, older, highly educated workers in large, private-sector firms
- Statistical significance seems to be driven by occupation and industry controls

Urban/rural difference

- NHIS public data does not include indicator urban v. rural locations (confidentiality), but lots of health-related question
- Public health literature: urban-rural health gradient
- Health conditions and behaviors as (imperfect) proxy for urban/rural
- Coronary heart disease, diabetes, cervical or colonrectal cancer; smoking; BMI; having difficulty finding physicians; delay getting healthcare due to lack of tranposrtation; last time visiting a dentist, mental health practitioner, or specialist; having private health insurance coverage

Health-augmented model

Table: Augmented earnings regression

	Dependent variable: Log annual earnings	
	Women	Men
	(1)	(2)
Homosexual	0.210	0.028
	(0.177)	(0.438)
Bisexual	−0.974**	1.974
	(0.432)	(2.284)
Controls included:		
Month & year interview dummies	Yes	Yes
Demographics	Yes	Yes
Job characteristics	Yes	Yes
Health vectors	Yes	Yes
Observations	386	186
R^2	0.693	0.714

Note: *p<0.1; **p<0.05; ***p<0.01. Robust standard errors in bracket. Full model description in the text. Full resutls are available upon request.

Oaxaca - Binder decomposition

- Evaluate relative impact of various demographics and labor market characteristics in contributing to the total earnings gap

$$
\begin{aligned}
&\overline{\text{Log Earnings}_{gay}} - \overline{\text{Log Earnings}_{straight}} \\
&= \beta_{gay}\overline{X_{gay}} - \beta_{straight}\overline{X_{straight}} \\
&= \beta_{gay}(\overline{X_{gay}} - \overline{X_{straight}}) + (\beta_{gay} - \beta_{straight})\overline{X_{straight}} \\
&= \text{Difference in characteristics} - \text{Difference in returns}
\end{aligned}
$$

OB decomposition result

Table: Oaxaca-Blinder decomposition of gay men's wage premium

Decomposition	Attributable to difference in:		% Relative to total gap	
	Characteristics	Returns	Characteristics	Returns
	(1)	(2)	(3)	(4)
Age	-0.011	1.078	16	1,618
Race and ethnicity	-0.019	0.074	29	112
Education	0.058	-0.034	87	52
Relationship	-0.018	-0.098	27	147
Children	-0.163	0.181	245	272
Region	0.001	-0.033	2	49
Tenure	-0.013	-0.110	20	166
Sector	0.000	0.010	0	15
Firm size	0.009	-0.198	13	297
Industry	-0.089	0.240	134	360
Occupation	0.058	-0.494	88	742
Survey wave	0.005	-0.113	7	170
Intercept	0.000	-0.252	0	379
Total	-0.183	0.249	-275	375

OB decomposition interpretation

- 0.067 log point in total (unconditional) earnings gap = 0.183 log point penalty due to characteristics + 0.249 log point premium due to returns
- Should straight men have the returns to characteristics as gay men do, we would expect their earnings to be 0.183 log point (20 pp) higher given their characteristic
- If both groups share the same average characteristics as straight men currently do, then the differential returns would give an even bigger gay earnings premium (0.249 log points or 28 pp).
- A lot of inconsistency, e.g. opposite effect of age v. tenure, relationship v. children, industry v. occupation

Occupation and industry sorting

- Endogenous selection bias ? Should we “control” for occupations and industries?
- Small sample - comparison within each occupation-industry cell

Industry distribution

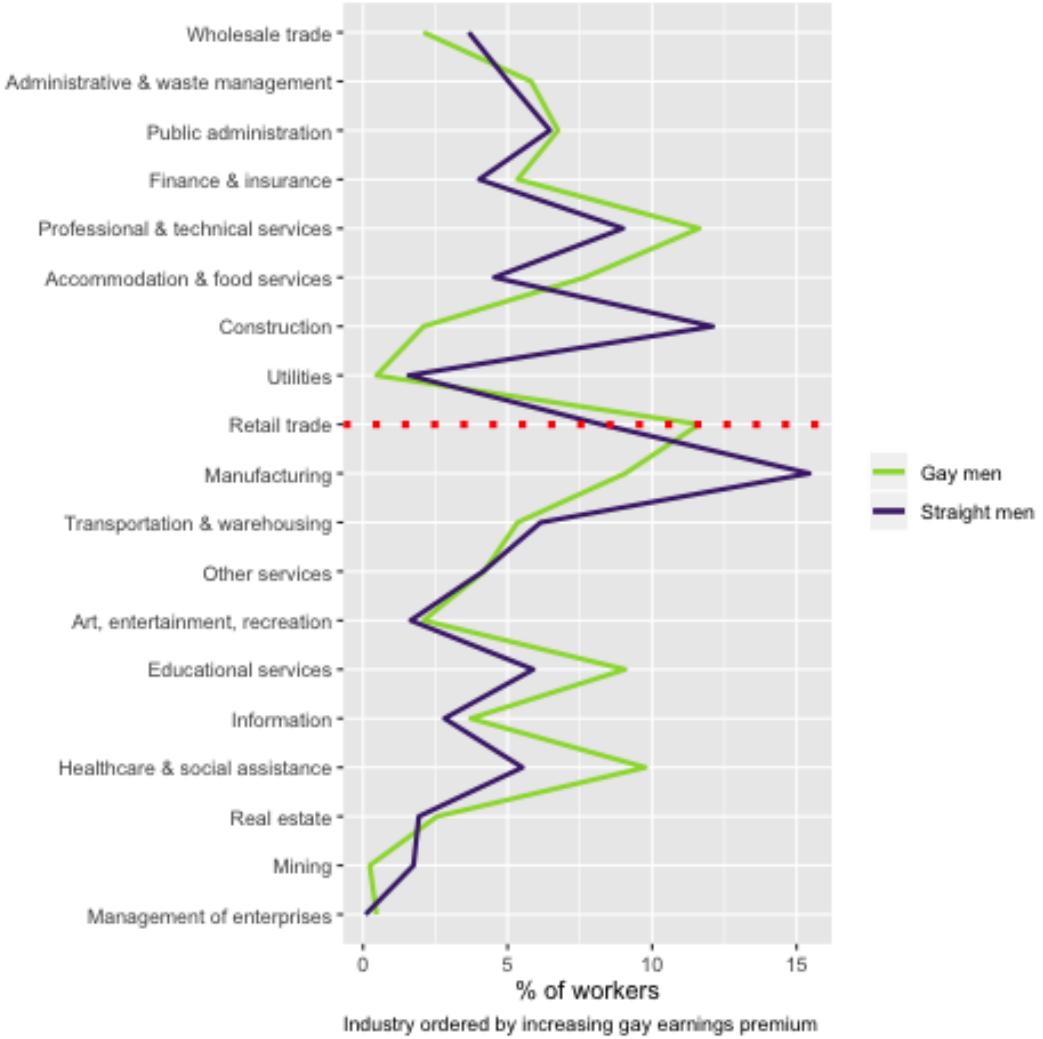
Table: Distribution of workers and earnings by industry

Industry	Gay men		Straight men		Average earnings		
	Count	% of total	Count	% of total	Gay	Straight	Difference
Wholesale trade	9	2.1	635	3.7	105,258	65,048	40,211
Administrative & waste management	25	5.8	873	5.0	63,372	43,513	19,859
Public administration	29	6.7	1122	6.5	89,930	73,883	16,046
Finance & insurance	23	5.3	698	4.0	107,692	93,296	14,396
Professional & technical services	50	11.6	1,561	9.0	100,469	88,056	12,413
Accommodation & food services	33	7.7	788	4.5	44,089	34,375	9,714
Construction	9	2.1	2,100	12.1	57,838	48,445	9,393
Utilities	2	0.5	274	1.6	77,684	76,201	1,483
Retail trade	50	11.6	1,431	8.2	52,826	51,455	1,371
Manufacturing	39	9.1	2,680	15.4	60,887	61,376	-489
Transportation & warehousing	23	5.3	1,067	6.1	51,898	55,404	-3,507
Other services	18	4.2	721	4.2	43,263	47,041	-3,778
Art, entertainment, recreation	9	2.1	286	1.6	41,760	45,995	-4,234
Educational services	39	9.1	1,022	5.9	52,101	58,267	-6,166
Information	16	3.7	489	2.8	67,217	76,226	-9,008
Healthcare & social assistance	42	9.8	957	5.5	55,921	69,002	-13,081
Real estate	11	2.6	335	1.9	46,174	62,382	-16,208
Mining	1	0.2	303	1.7	65,000	87,107	-22,107
Management of enterprises	2	0.5	15	0.1	50,311	92,962	-42,651

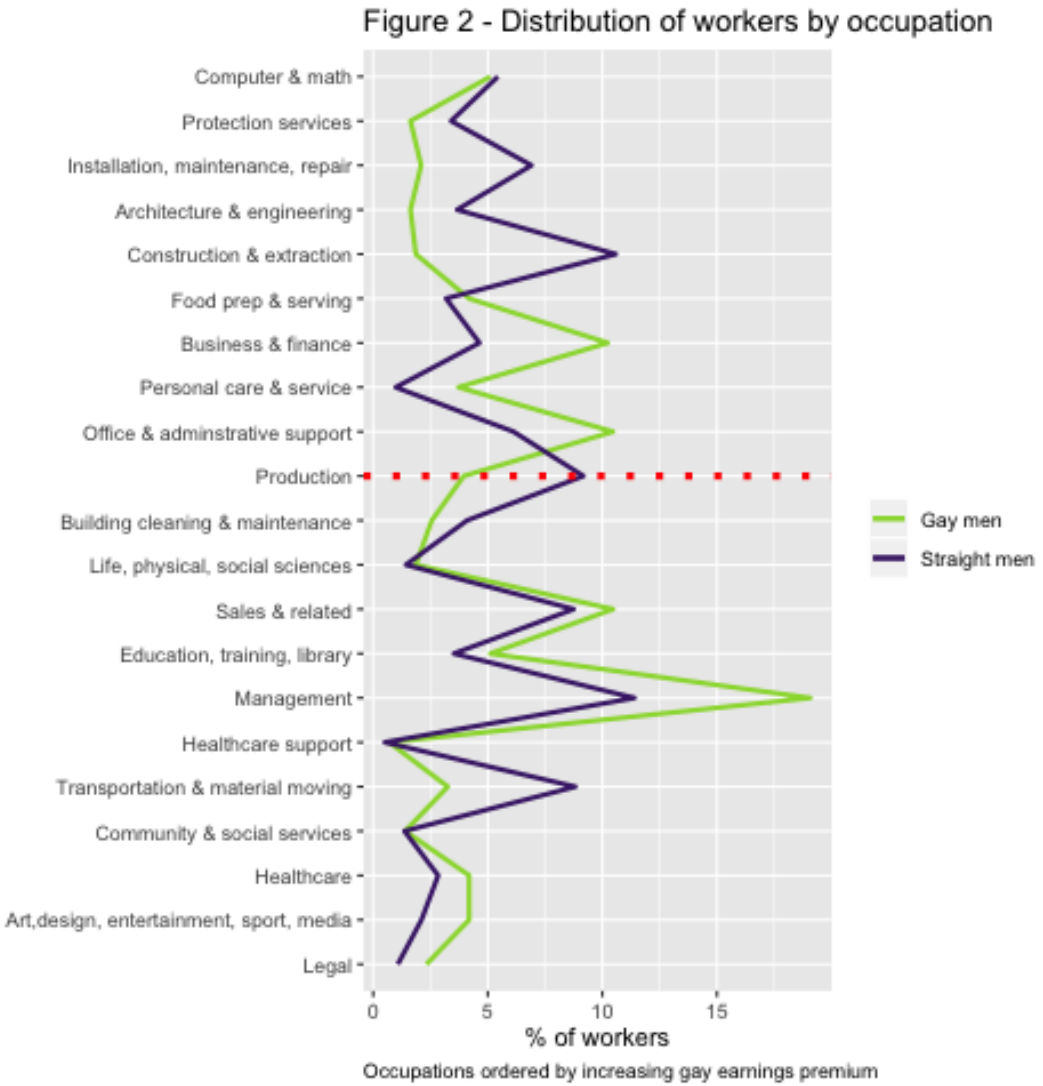
*Note:*Industries are in descending order of the difference in weighted average earnings between gay and straight workers.

Industry distribution

Figure 1 - Distribution of workers by industry



Occupation distribution



Occupation & industry

- Gay men are concentrated in occupations/industries both where they have earnings advantage and where they are paid much less than straight men.
- No discernible pattern of systematic sorting. 46.5% of gay men vs. 45.4% of straight men are working in industries in which gay men earn a premium. Similarly, 59.1% of gay men and 55.1% of straight men are working in occupations in which gay men earn a premium.
- But very small sample: many occupations and industries have less than 20 gay men workings (a few have none).

Occupation & industry restriction

Table: Earnings regression - industries and occupations with $N_{gay} \geq 30$

Specifications	Coef. on Gay dummy
Baseline (interviews dummy only)	0.027
	(0.057)
Baseline + demographic	0.068
	(0.052)
Full model without occupation and industry control	0.065
	(0.047)
Full model	0.085*
	(0.044)

Note: *p<0.1; **p<0.05; ***p<0.01. Robust standard errors in bracket.
N = 11,894; $N_{gay} = 341$

Tips for replication

- Start early, and work on it weeks by weeks !
- Pay attention to descriptive statistics: they can point you to the way
- Explore different extension paths, you don't know *a priori* where they might lead to
- Talk to other folks for ideas and suggestions