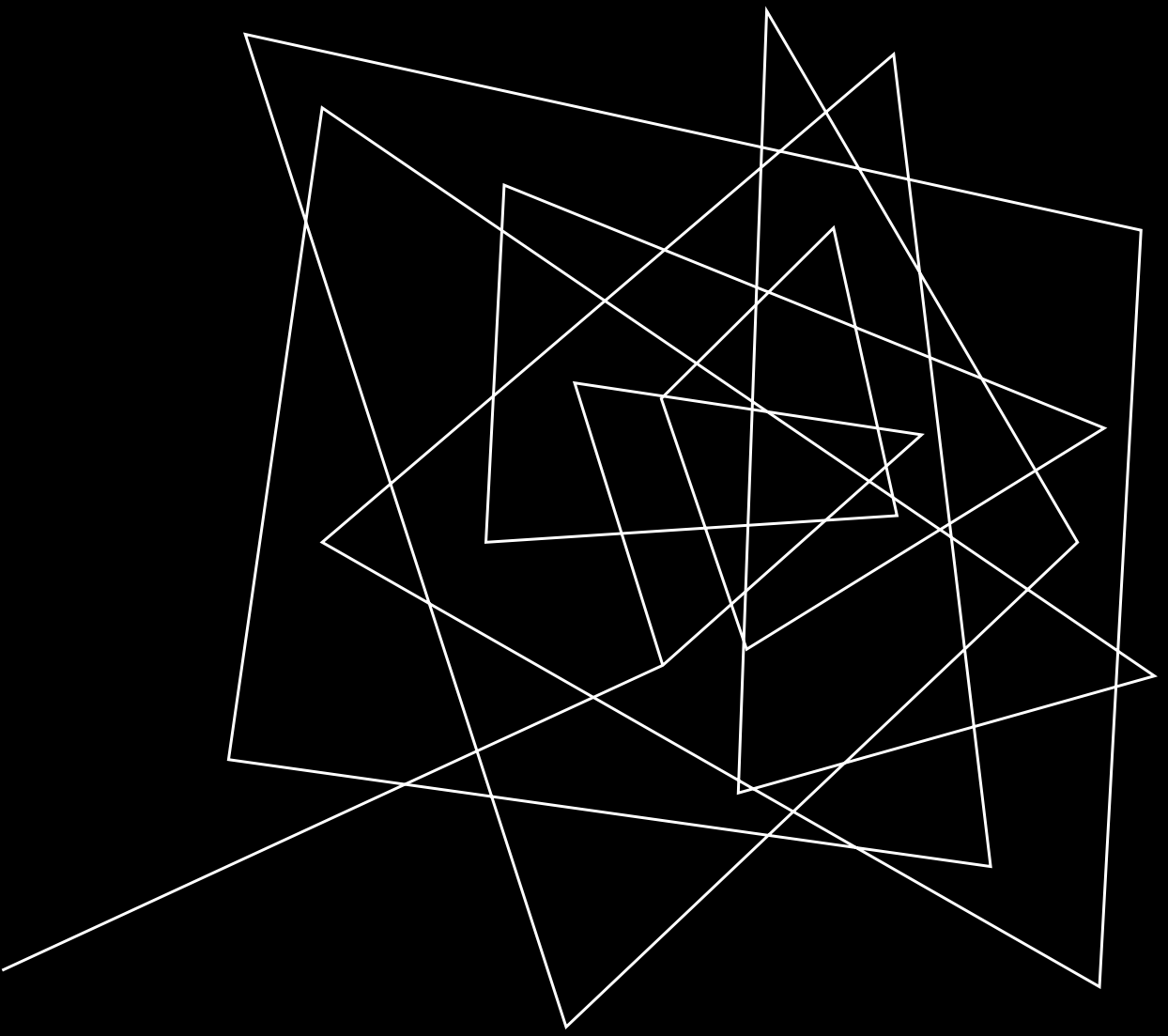
IS THERE REALLY A POT OF GOLD AT THE END OF THE RAINBOW: AN INVESTIGATION OF SEXUAL ORIENTATION PAY GAPS IN THE UK

Scott Oatley



PAY GAPS

PAY GAPS

- Pay gap research has a massive history in econometrics and sociological literature
- Mainly started to assess gender pay gaps and ‘glass ceilings’ (Olsen et al 2018)
- Has been extended to assess ethnic pay gaps (Brynin and Güveli 2012) and relatively recently to investigate social class pay gaps (Laurison and Friedman 2016)
- Very little work has been done to extend analysis to other areas of social stratification – sexuality
- Some pioneering work has been done
  - Vast majority of work on this area deals with flawed operationalisation of sexuality data
    - Couple linkage, large assumptions (just because someone is in a relationship with X doesn’t mean they are ‘out’)
    - Bierasure
  - Is American
  - The two are not necessarily related but who knows...

EXISTING LITERATURE

- American literature on the topic dominants

- Started with a tradition started by Badgett (1995)
  - Gay and bisexual male workers earned from 11% to 27% less than heterosexual counterparts
  - Also, evidence to suggest same for gay and bisexual women

- Followed up by a variety of others
  - Mostly use the General Social Survey in USA
  - Most modern data in the US finds consistent evidence for male bisexual pay penalty, mixed findings for gay men and typically finds pay premiums for gay and bisexual women (Drydakis 2022)

- A very small, growing literature in the UK but mainly uses the Labour Force Survey or the UK Integrated Household Surveys (Aksoy et al 2018)

- UKHLS has access to sexuality at multiple wave points but is yet an untapped datasource

THEORY

- What theory explains these gaps?

- Becker’s (1957) taste discrimination
  - Blinds the employer to the true monetary costs of hiring a minority worker. An employer who discriminates will, following this model, act as if the costs of hiring a minority worker exceed the actual costs.
  - Prejudiced workers (customers) may act as if their wage (the price of the good they want to sell) is a fraction equal to their discrimination coefficient lower (higher) if they have to interact with a minority worker.
  - discrimination will decrease hiring chances for the minority worker, at least in sectors dominated by majority workers (Borjas, 2009)

THEORY

- Arrow’s (1973) statistical discrimination

    - Statistical discrimination occurs when, as a time-efficient and profit-maximising response to low information and uncertainty about the actual productivity of individual job candidates, employers take into account their views about the relative productivity-related characteristics of different groups (based on information that might be imperfect) to predict a particular applicant’s productivity.

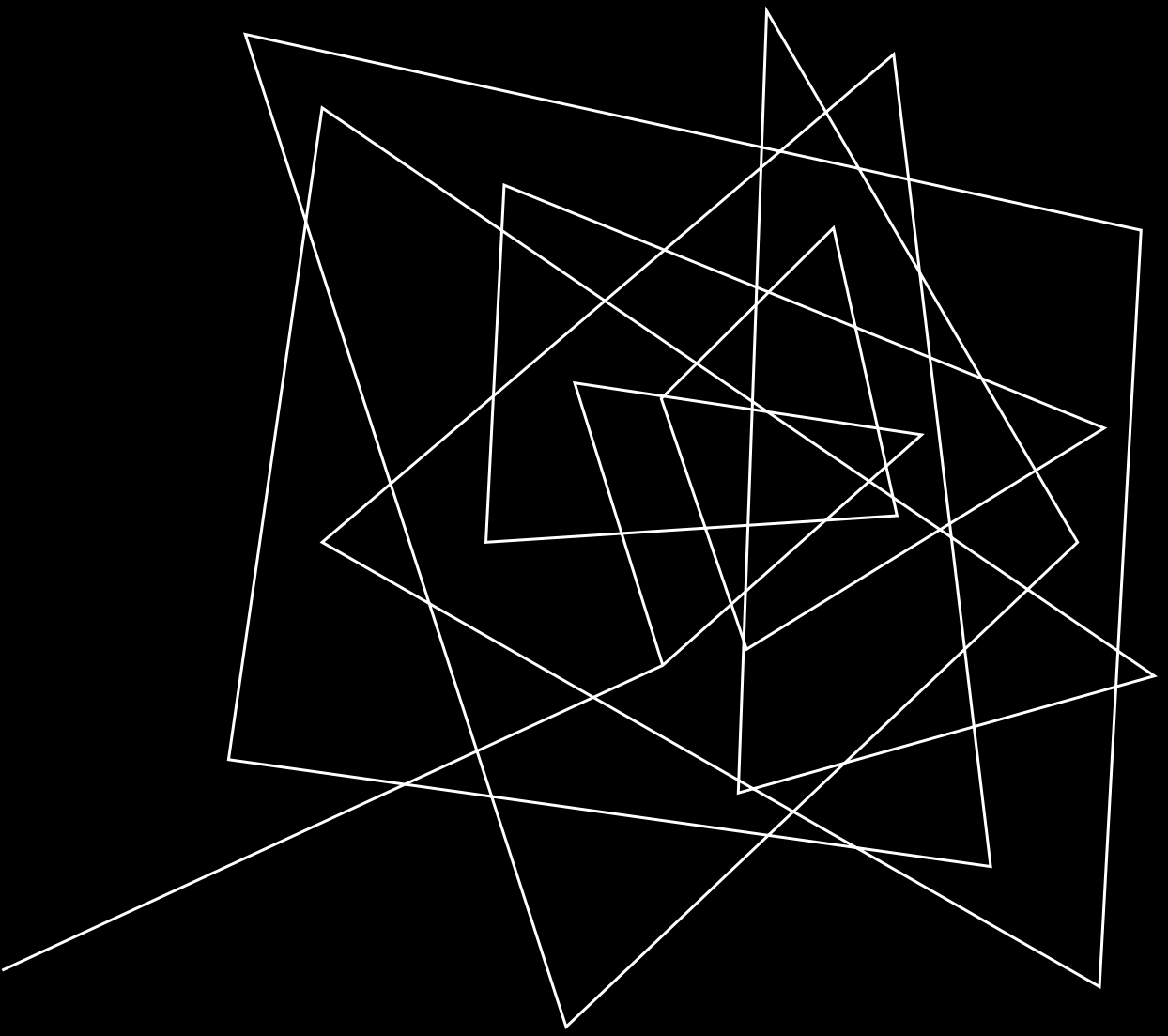
    - Several factors may lead employers to expect higher average productivity from lesbians relative to straight women. First, lesbians are documented as being, on average, more ‘masculine’, that is, more dominant, autonomous and assertive. This characteristic may match well with some specific jobs and may adhere to the ideal of masculinity that is associated with labour market success (Berg and Lien, 2002; Blandford, 2003; Clain and Leppel, 2001)

THEORY

- Second, lesbians are documented as having on average, a more committed and continuous labour market participation. This characteristic results from two different factors. On the one hand, on average, lesbians have children less frequently than heterosexual women and, due to a less traditional division of labour within the household, lesbians, on average, engage in less rearing tasks, which are conditional on having children in the first place, than heterosexual women do. On the other hand, this less traditional division of labour also results in the reduction of other household responsibilities.

- The fact that gay men are (perceived as), on average, less masculine or, more concretely, less dominant, autonomous, and assertive (Blashill & Powlishta, 2009; Chung & Harmon, 1994; Ellis & Ratnasingam, 2012; Haddock, Zanna, & Esses, 1993; Jackson & Sullivan, 1989; Kite, 2011; Kite & Whitley, 1996; Kurdek, 2006).

- Individual homosexual men may be statistically discriminated based on the fact that the group of gay men is documented and perceived as being less masculine, an ideal to which employers adhere, at least for particular occupations (Berg & Lien, 2002; Blandford, 2003; Kite & Deaux, 1987; Madon, 1997).



SOME SEXUALITY DETAILS IN UKHLS

MEASURES

- Prior to investigation of pay gap related data

- Sexuality in the UKHLS presents a vital opportunity to explore key aspect of stratification often overlooked

- Some data that may be of interest:
  - Couple pairings
  - Occupational Sorting

COUPLE PARINGS AND SEXUAL ORIENTATION

samesexrel	sexuality_new			Total
	Heterosex	Homosexua	Bisexual	
same-sex	51	1,325	115	1,491
	0.05	96.64	13.23	1.49
hetero	97,745	46	754	98,545
	99.95	3.36	86.77	98.51
Total	97,796	1,371	869	100,036
	100.00	100.00	100.00	100.00

OCCUPATIONAL SORTING

Production, works and maintenance managers	2.94	Straight Male
Marketing and sales managers	2.99	Straight Male
Information and communications technology	2.01	Straight Male
Software Professionals	2.66	Straight Male
Elementary occupations in goods handling	2.19	Straight Male
Secondary education teaching professionals	2.81	Straight Female
Primary and nursery education teaching	3.48	Straight Female
Nurses	5.29	Straight Female
Accounts and wages clerks, book-keepers	3.35	Straight Female
General office assistances/clerks	4.43	Straight Female
Care assistances and home carers	5.25	Straight Female
Educational assistants	4.42	Straight Female
Sales assistances	4.27	Straight Female
Cleaners, domestics	2.26	Straight Female

OCCUPATIONAL SORTING

Personnel, training and industrial relations	3.82	Gay Male
Information and communications technology	2.18	Gay Male
Customer care managers	2.05	Gay Male
Software professionals	2.05	Gay Male
Medical practitioners	2.86	Gay Male
Secondary education teaching professionals	3.14	Gay Male
Primary and nursey education	2.46	Gay Male
Public service administrative professionals	3.27	Gay Male
Nurses	2.18	Gay Male
Youth and community workers	2.59	Gay Male
Housing and welfare officers	3.41	Gay Male
Vocational and industrial trainers	2.73	Gay Male
Accounts and wage clerks, book-keepers	2.05	Gay Male
Care assistants and home carers	3.41	Gay Male
Office Managers	2.24	Gay Female
Hospital and health service managers	2.84	Gay Female
Social services managers	2.24	Gay Female
Mangers and proprietors in other services	2.09	Gay Female
Higher education teaching professionals	2.69	Gay Female
Secondary education teaching professionals	3.29	Gay Female
Primary and nursey education teaching	4.78	Gay Female
Archivists and curators	2.39	Gay Female
Nurses	4.04	Gay Female
Housing and welfare officers	2.54	Gay Female
Sales representatives	2.09	Gay Female
Nursing auxiliaries and assistants	2.84	Gay Female
Customer service occupations	2.09	Gay Female
Kitchen and catering assistants	2.84	Gay Female

OCCUPATIONAL SORTING

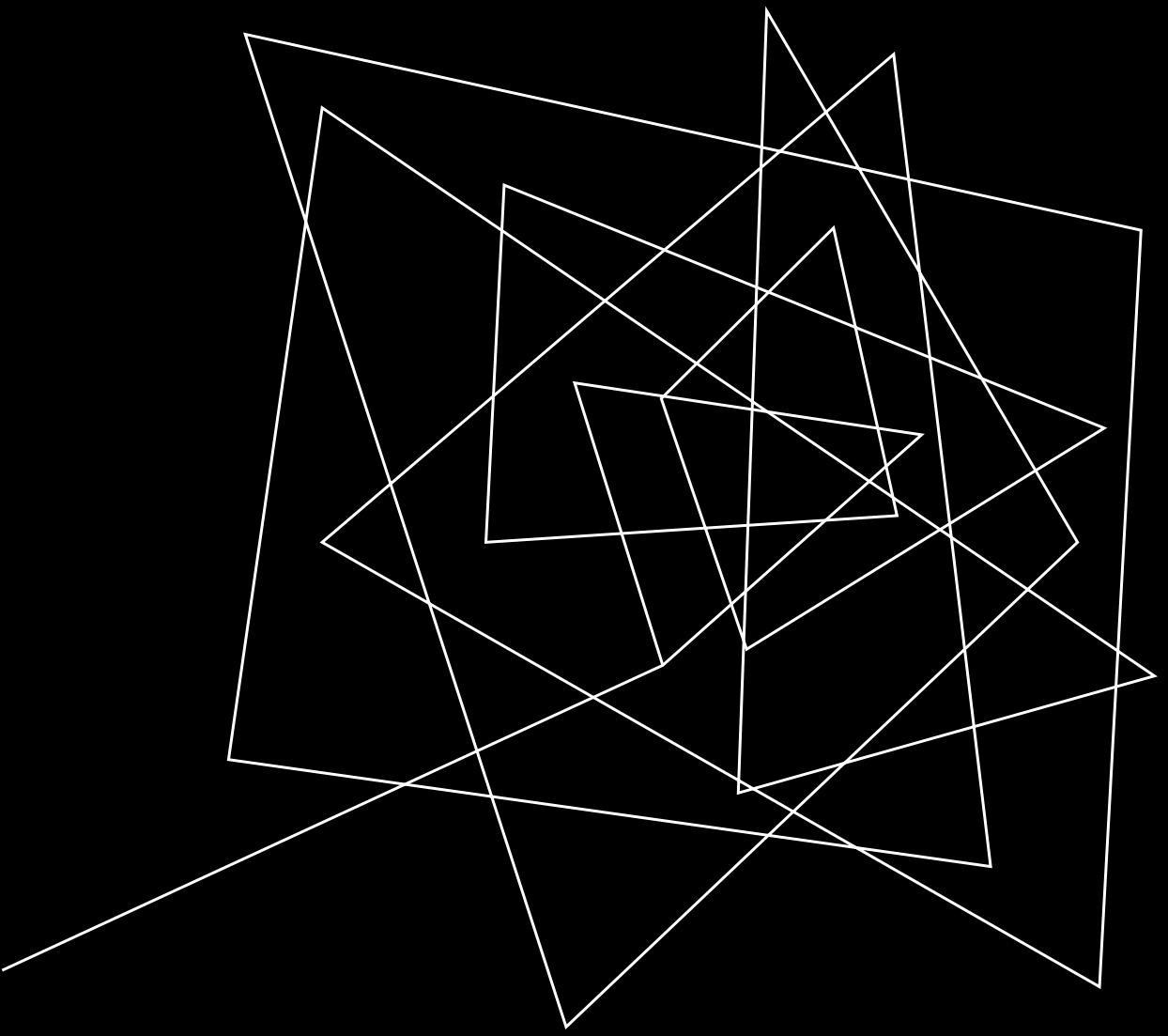
Personnel, training and industrial relations	3.82	Gay Male
Information and communications technology	2.18	Gay Male
Customer care managers	2.05	Gay Male
Software professionals	2.05	Gay Male
Medical practitioners	2.86	Gay Male
Secondary education teaching professionals	3.14	Gay Male
Primary and nursey education	2.46	Gay Male
Public service administrative professionals	3.27	Gay Male
Nurses	2.18	Gay Male
Youth and community workers	2.59	Gay Male
Housing and welfare officers	3.41	Gay Male
Vocational and industrial trainers	2.73	Gay Male
Accounts and wage clerks, book-keepers	2.05	Gay Male
Care assistants and home carers	3.41	Gay Male
Office Managers	2.24	Gay Female
Hospital and health service managers	2.84	Gay Female
Social services managers	2.24	Gay Female
Mangers and proprietors in other services	2.09	Gay Female
Higher education teaching professionals	2.69	Gay Female
Secondary education teaching professionals	3.29	Gay Female
Primary and nursey education teaching	4.78	Gay Female
Archivists and curators	2.39	Gay Female
Nurses	4.04	Gay Female
Housing and welfare officers	2.54	Gay Female
Sales representatives	2.09	Gay Female
Nursing auxiliaries and assistants	2.84	Gay Female
Customer service occupations	2.09	Gay Female
Kitchen and catering assistants	2.84	Gay Female

OCCUPATIONAL SORTING

Marketing and sales managers	3.93	Bi	Male
Mangers and proprietors in other services	2.11	Bi	Male
Software professionals	7.55	Bi	Male
Scientific researchers	3.32	Bi	Male
IT user support technicians	2.72	Bi	Male
Police officers (sergeant and below)	2.42	Bi	Male
Stock control clerks	2.42	Bi	Male
Library assistants/clerks	2.11	Bi	Male
Nursing auxiliaries and assistants	3.32	Bi	Male
Care assistants and home carers	4.53	Bi	Male
Sales and retail assistants	3.32	Bi	Male
Taxi, cab drivers and chauffer's	3.02	Bi	Male
Kitchen and catering assistants	3.02	Bi	Male
Security guards and related occupations	6.34	Bi	Male
Senior officials	2.84	Bi	Female
Residential and day care managers	2.62	Bi	Female
Teaching professionals n.e.c	3.93	Bi	Female
Nurses	3.71	Bi	Female
Business and related associate professionals	4.8	Bi	Female
Pensions and insurance clerks	2.4	Bi	Female
Care assistants and home carers	3.49	Bi	Female
Air travel assistants	2.18	Bi	Female
Sales and retail assistants	2.4	Bi	Female
Customer service occupations	4.15	Bi	Female
Cleaners, domestics	2.84	Bi	Female

MEASURES

- Lesbian and bi women sort into managerial occupations
- Gay and bi men appear to sort into more ‘traditionally feminine’ occupations
- Straight men are rather boring



MEASURES

MEASURES

- Dependent Variable = Log Annual Net Labour Income

- Derived from three sources:
  - Net usual pay
  - Net self-employment income
  - Net pay in second job

  - Negative values solely come from self-employed net income

- Derived net income measure is monthly

- Income is equivalised using the CPI pegged to 2016 (comparison to Laurison and Friedman 2016)

- Gamma distributed resulted in log income transformation

MEASURES

- A lot of them...

- Primary covariates:
  - Gender
  - Parental Status
  - Marital Status
  - Sexuality

- Demographic covariates:
  - Age + Age2 (centred + /100)
  - Housing Tenure
  - Parental NS-SEC
  - Ethnic Identity
  - Highest Educational Qualifications
  - Long Term Illness

MEASURES

- A lot of them...

- Regional:
  - Urban
  - Government Regions – 12 regions (Scotland is apparently a region...)

- Workplace:
  - Industry
  - Size
  - Workhours (/100)
  - SOC 2000

- Time:
  - Wave

MEASURES

- A little discussion of the important measures

- Parental status
  - For men: evidence of stability?
  - For women: Maternity leave, more kids, leaving soon?

- Marital Status
  - For men: stability
  - For women: children soon

- Sexuality:
  - For gay men: effeminate, not a man’s man
  - For bi men: ???
  - For gay women: masculine, no children
  - For bi women: ???

MEASURES

- Strong case for interactions!

- A lot of them...
  - OG model had significant 4-way interactions in fixed effects
    - Actually semi-decent n's except for one interaction combination
    - I also just didn't want to attempt to interpret a 4-way interaction on the fly...

Table 1: Descriptive Statistics for Model

	n	%
Gender		
male	60,898	43.68%
female	78,533	56.32%
Current housing tenure		
owned/privately rented	124877	89.56%
social housing	14,554	10.44%
sexuality_new		
Heterosexual	135414	97.12%
Homosexual	2,358	1.69%
Bisexual	1,659	1.19%
child		
No Child	85,972	61.66%
Child	53,459	38.34%
nssecdom_new		
Large employers & higher management	5,752	4.13%
Higher professional	7,553	5.42%
Lower management & professional	13,942	10.00%
Intermediate	9,299	6.67%
Small employers & own account	12,088	8.67%
Lower supervisory & technical	11,445	8.21%
Semi-routine	10,595	7.60%
Routine	13,763	9.87%
Not in employment	54,994	39.44%

MEASURES

ethnic		
White	122263	87.69%
Mixed/Multiple Ethnic Groups	2,506	1.80%
Indian	4,190	3.01%
Pakistani and Bangladeshi	3,565	2.56%
Chinese	694	0.50%
Any other Asian Background	1,150	0.82%
	4,801	3.44%
Black/African/Carribean/Black British		
Other	262	0.19%
hed		
Degree/Other Higher	70,869	50.83%
A'level	30,844	22.12%
GCSE/Other	34,177	24.51%
No Qualification	3,541	2.54%

ill		
No	105027	75.33%
Yes	34,404	24.67%
region		
North East	5,366	3.85%
North West	14,237	10.21%
Yorkshire and the Humber	11,120	7.98%
East Midlands	10,472	7.51%
West Midlands	11,002	7.89%
East of England	12,450	8.93%
London	14,439	10.36%
South East	17,143	12.29%
South West	11,847	8.50%
Wales	9,522	6.83%
Scotland	13,284	9.53%
Northern Ireland	8,549	6.13%
urban		
Rural	33,362	23.93%
Urban	106069	76.07%
sector		
Public	54,305	38.95%
Private	85,126	61.05%

industry		
Public Admin, education, and health	49,163	35.26%
Agriculture, forestry, and fishing	511	0.37%
Energy and water	1,370	0.98%
Manufacturing	15,650	11.22%
Construction	5,035	3.61%
Distribution, hotels, and restaurants	21,678	15.55%
Transport and Communication	9,050	6.49%
Banking and finance	6,687	4.80%
Other services	30,287	21.72%
size		
500+	30,811	22.10%
50-499	47,391	33.99%
25-49	19,105	13.70%
Less than 25	42,124	30.21%

UKHLS wave		
3	17,859	12.81%
4	15,220	10.92%
5	14,685	10.53%
6	12,273	8.80%
7	12,312	8.83%
8	11,172	8.01%
9	13,462	9.65%
10	11,978	8.59%
11	10,710	7.68%
12	8,945	6.42%
13	8,521	6.11%
14	2,294	1.65%

	Mean	SD
logincome	9.72	0.65
Age Centred on 20 Divided by 100	-0.00	0.12
Age Squared Centred on Grand Mean Divided by 100	1.35	1.42
Workhours Divided by 100	-0.01	0.10

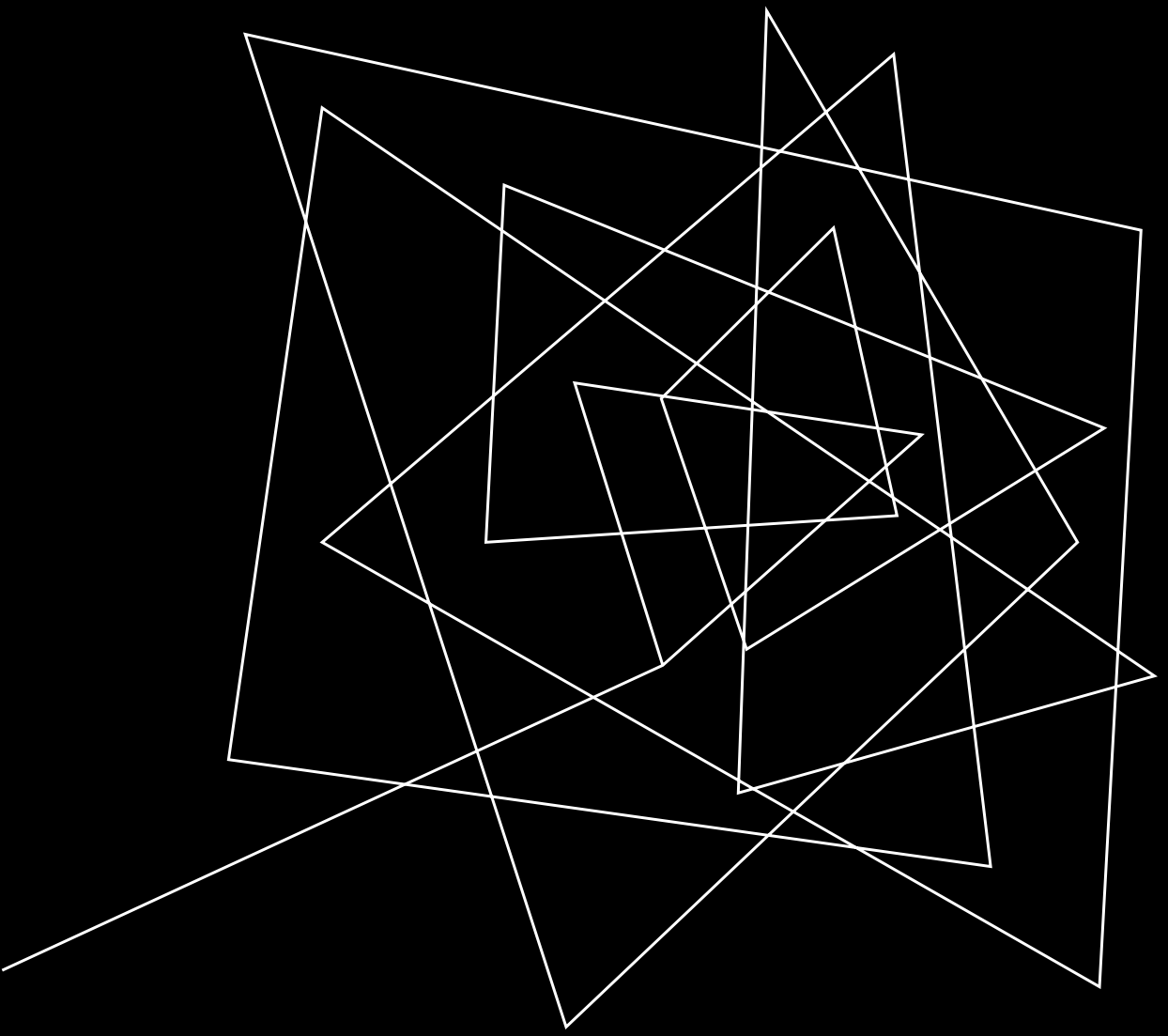
n	139431

WAVES

- Quick note on repeated measures

- For sexuality based measures
  - Each wave has between 2-300 Homosexuals
  - Each wave has between 1-200 bisexuals

- n is in my view robust enough over waves...



MODELLING STRATEGY

DATA STRUCTURE

- Any Model has to consider the longitudinal design of dataset
  - For income, a growth curve model is most appropriate
    - How does income change (grow) overtime

- The model also has to consider the dependent variable
  - Net Labour Income
    - Continuous
    - Follows a gamma distribution

  - Possible to run a growth curve model with a link logit, family gamma form
  - Also viable to log transform income – typically the norm in econometrics literature on income gaps

DATA STRUCTURE

- Literature is very clear on the fact that income is differentiated by occupation

- Is it better to stick current NS-SEC measure in fixed effects section, or would be better to use current occupation as a higher-level random effect?
  - Former: Statistically simpler and less computer-intensive
  - Latter: More grounded in material reality

- The model is already a growth curve model. Adding an additional higher-level component of occupation makes this a three-level MLM
  - Occupation is not nested neatly within individual-waves
  - Cross-classified growth curve model is appropriate

DATA STRUCTURE

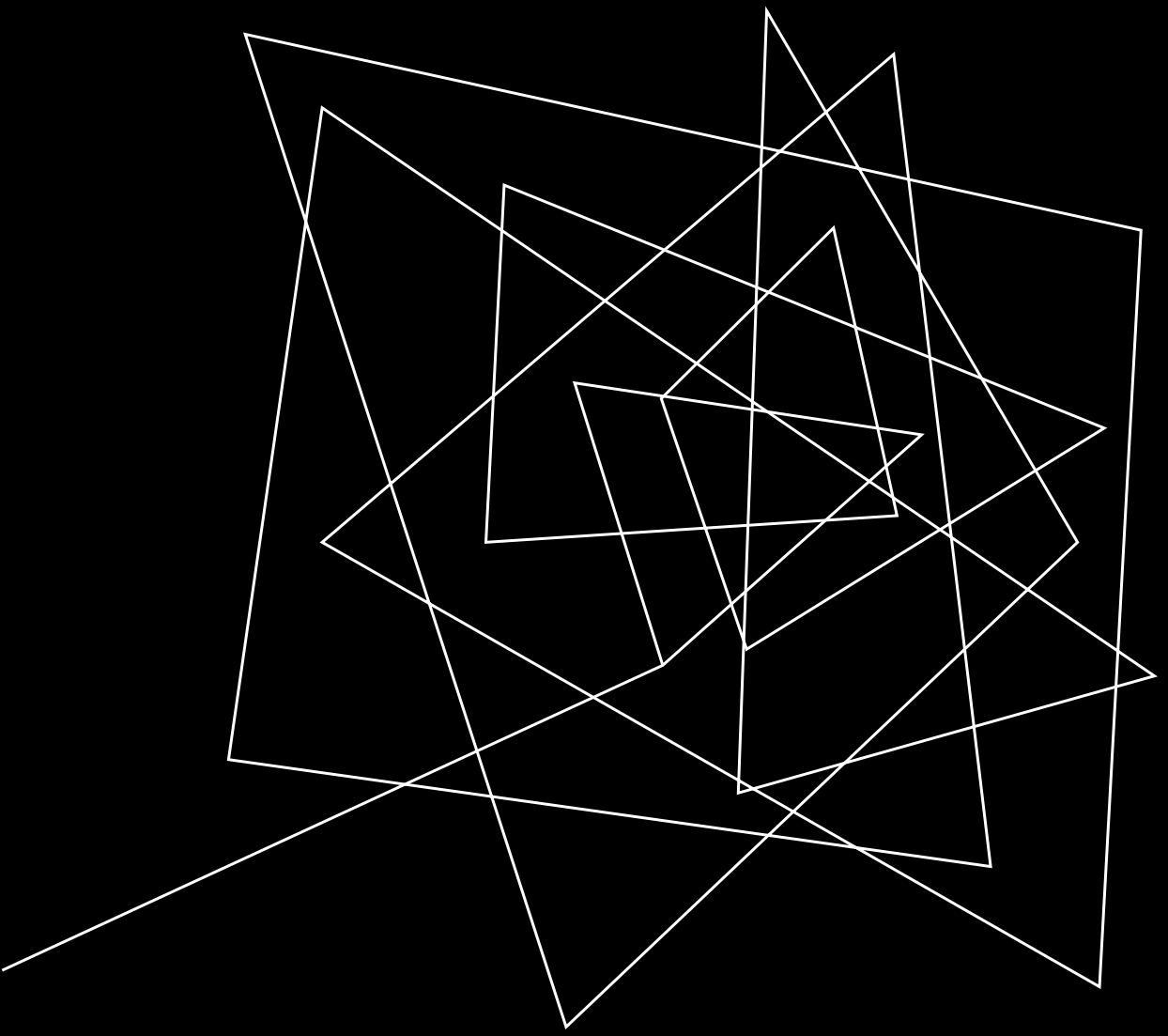
- Literature is also clear on the fact that different sexualities have varying degrees of authority, occupational sorting, working hours etc
- Some of this can be assessed in the form of a random slope
- Having working hours as a random slope would allow each group line to vary, allowing each set of explanatory variables to vary for that grouping
- Final model: Cross-Classified Growth Curve Model with a Random Slope Component
- This model has been tested against simpler models: the Intercept model, working up to the cross-classified growth curve model. AIC and BIC confirm better model fit.

DATA STRUCTURE

- Two issues with this model choice:

- SVY adjustments are impossible under a cross-classified modelling structure
  - The fact that sexuality has zero weighting in UKHLS makes this a little less awful but the design and SVY adjustments are needed!
    - Common issue with pretty much any advanced quant models

- Disaggregation also becomes impossible using standard pay gap related decomposition techniques
  - Blinder-Oaxaca decompositions allow researchers to disaggregate pay gaps from matters of merit from matters of discrimination
  - This decomposition is also impossible in the growth curve model as well
    - There is an experimental (extremely) xtoaxaca command that performs sometimes...



MODEL

MODEL

- Total n=139,431

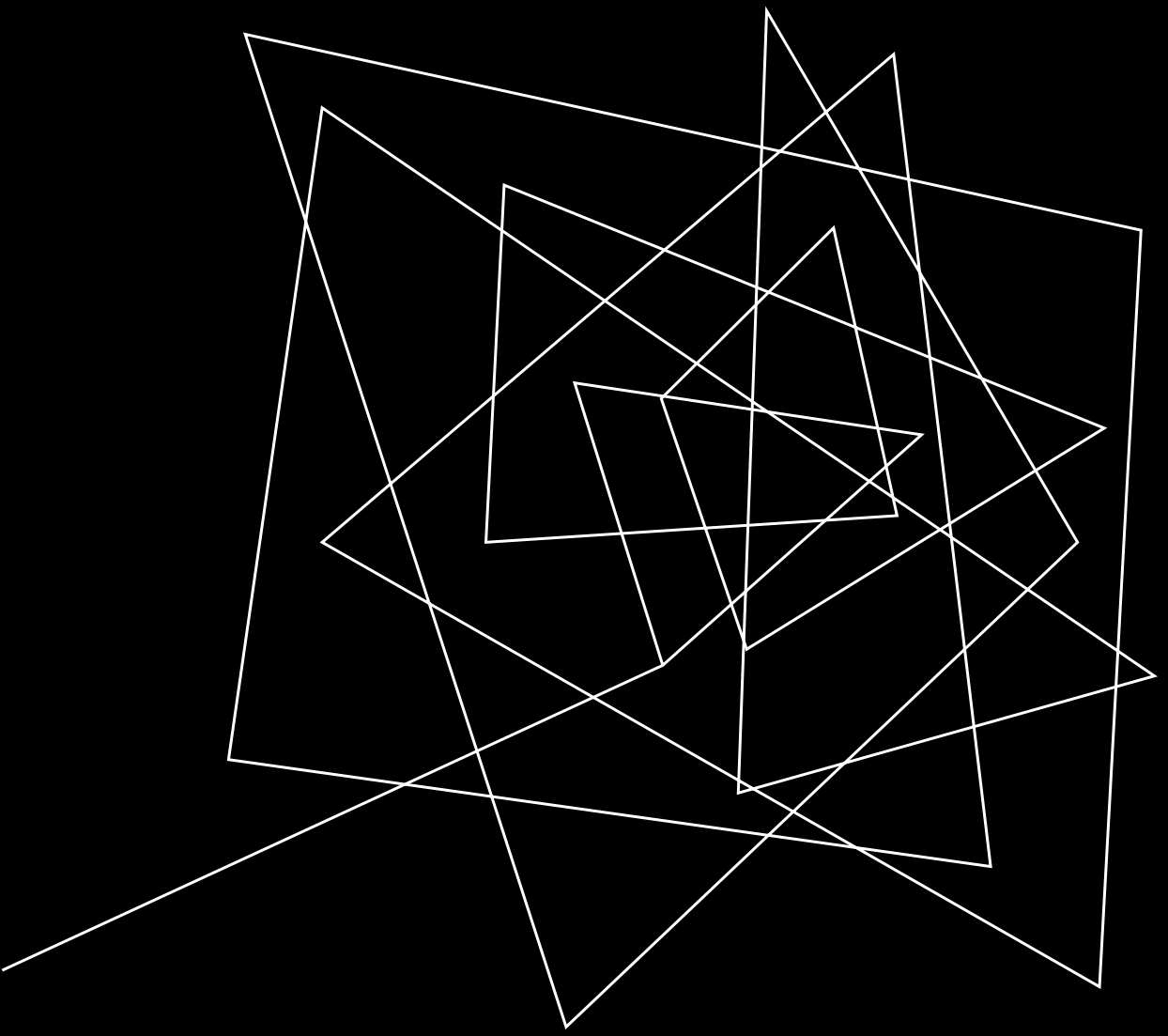
- This n spans over 11 waves of UKHLS data
  - From wave 3-14

- (Work of Matthews et al (2024) has really helped in my view push for use of sexuality data in UKHLS)

- Substantial missingness
  - Primarily coming from two main sources: working population and sexuality
    - Bulk of missing data comes from the fact that a large proportion of the UKHLS working population are in fact not workers
    - Additionally, a large amount of missingness comes from people either refusing to answer the sexuality data, or answering in a way that is difficult to model i.e ‘other’
    - Multiple Imputation Strategies are planned

WAVE DATA

- Notice I said waves 3-14
- Sexuality first collected from wave 3, periodically thereafter
- Use Last Observation Carried Forward but NOT backward to fill in ‘fixed’ effects variables
- Plausible assumption that someone that has ‘come out’ at wave three but did not answer at wave nine that they are still ‘out’
  - Very strong (too strong) assumption to assume someone that didn’t answer at wave three but did at wave nine was ‘out’ at wave three



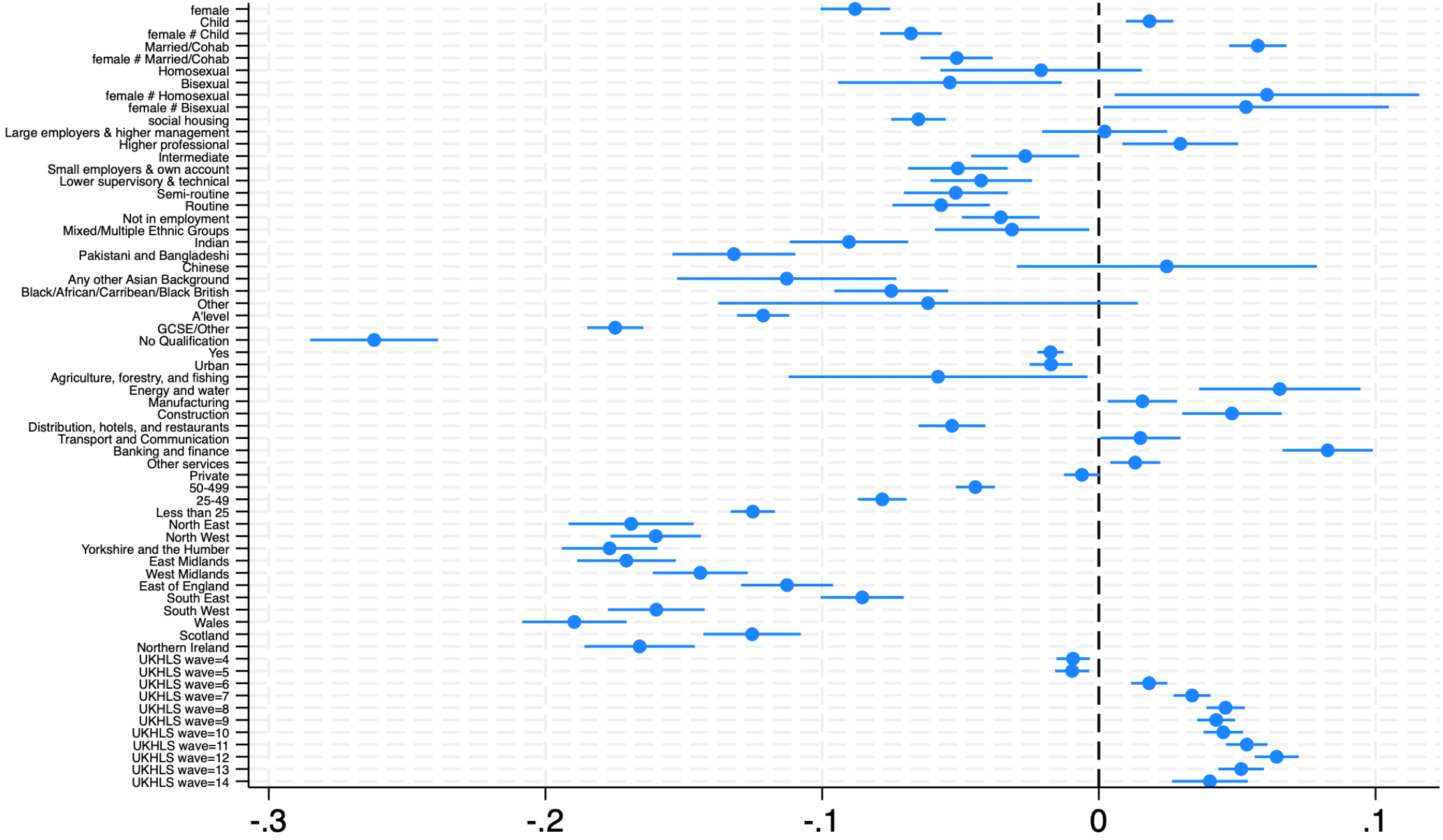
RESULTS

PRESENTATION OF RESULTS

- Very work in progress
- Had hoped to have some nice graphics for you
- Coefplot
- Mainly reporting textually the key fixed effects components
- Random effects reported and caterpillar plot
  - VPC somewhat pointless with a model like this
    - (Open to differences of opinion on this)

FIXED EFFECTS

- Notice I said waves 3-14
- Sexuality first collected from wave 3, periodically thereafter
- Use Last Observation Carried Forward but NOT backward to fill in ‘fixed’ effects variables
- Plausible assumption that someone that has ‘come out’ at wave three but did not answer at wave nine that they are still ‘out’
  - Very strong (too strong) assumption to assume someone that didn’t answer at wave three but did at wave nine was ‘out’ at wave three


female
Child
female # Child
Married/Cohab
female # Married/Cohab
Homosexual
Bisexual
female # Homosexual
female # Bisexual
social housing
Large employers & higher management
Higher professional
Intermediate
Small employers & own account
Lower supervisory & technical
Semi-routine
Routine
Not in employment
Mixed/Multiple Ethnic Groups
Indian
Pakistani and Bangladeshi
Chinese
Any other Asian Background
Black/African/Carribean/Black British
Other
A'level
GCSE/Other
No Qualification
Yes
Urban
Agriculture, forestry, and fishing
Energy and water
Manufacturing
Construction
Distribution, hotels, and restaurants
Transport and Communication
Banking and finance
Other services
Private
50-499
25-49
Less than 25
North East
North West
Yorkshire and the Humber
East Midlands
West Midlands
East of England
South East
South West
Wales
Scotland
Northern Ireland
UKHLS wave=4
UKHLS wave=5
UKHLS wave=6
UKHLS wave=7
UKHLS wave=8
UKHLS wave=9
UKHLS wave=10
UKHLS wave=11
UKHLS wave=12
UKHLS wave=13
UKHLS wave=14
-.3
-.2
-.1
0
.1

FIXED EFFECTS

- Notice I said waves 3-14
- Sexuality first collected from wave 3, periodically thereafter
- Use Last Observation Carried Forward but NOT backward to fill in ‘fixed’ effects variables
- Plausible assumption that someone that has ‘come out’ at wave three but did not answer at wave nine that they are still ‘out’
  - Very strong (too strong) assumption to assume someone that didn’t answer at wave three but did at wave nine was ‘out’ at wave three

FIXED EFFECTS

- Intercept = 10.18 = £26,370
- Net of all effects*

- Women = -8%

- Male Parental Status = 2%
- Female Parental Status = -7%

- Male Married = 6%
- Female Married = -5%

FIXED EFFECTS

- Male Homosexuals = N/A (Male Homosexual gap (and others) disappears once we control for working hours)
- Male Bisexuals = -5%
- Female Homosexuals = 6%
- Female Bisexuals = 5%

- Putting this in perspective:
  - Sexuality based income gaps are around the same size as social class-based income gaps (around +-5%)

RANDOM EFFECTS

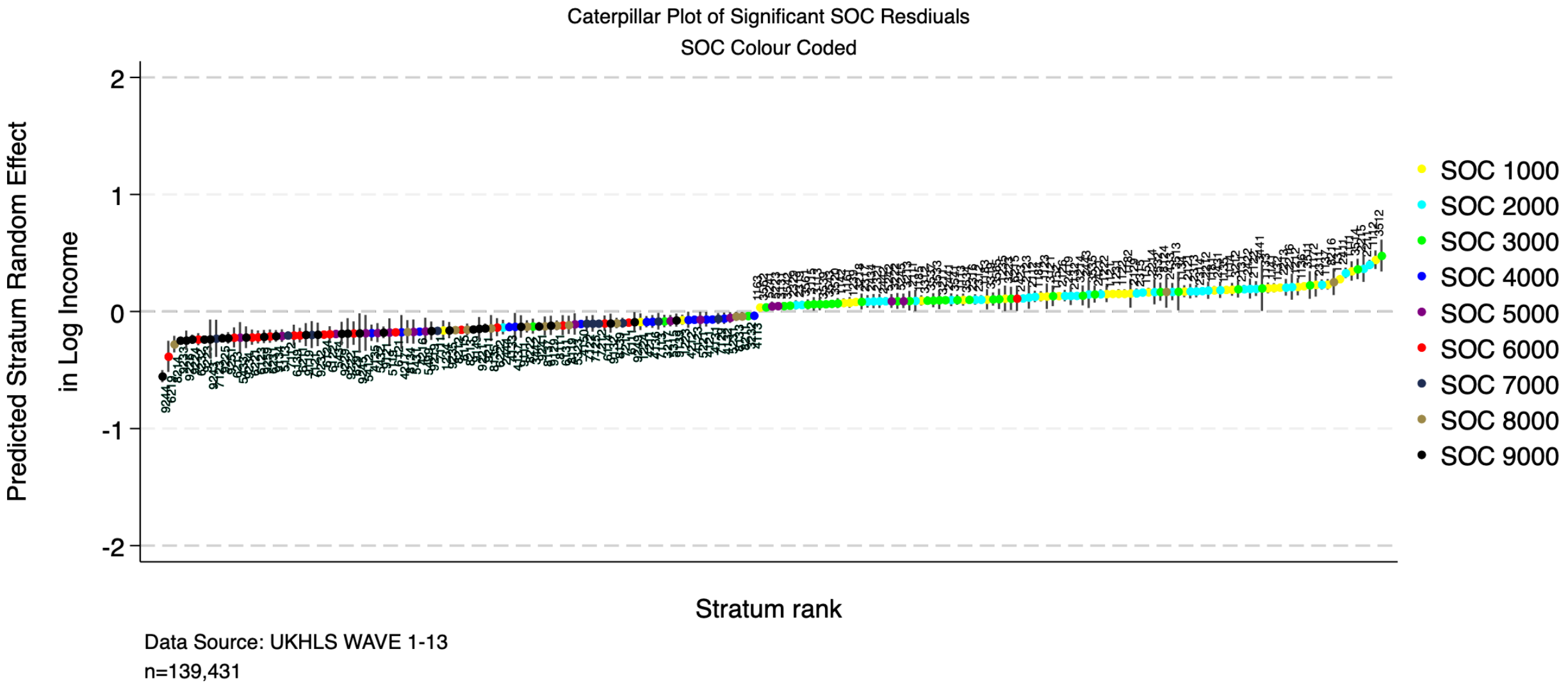
var(R_jbsoc00)	0.02	
	(0.00)	
var(workhours_100)	45007.06	
	(804.68)	
var(_cons)	0.07	
	(0.00)	
cov(workhours_100,_cons)	-7.00	***
	(0.68)	
var(e)	0.07	
	(0.00)	
Number of observations	139431	
AIC	86774.26	
BIC	87492.97	

RANDOM EFFECTS

- Variance seems very small but remember log income
- The common variance parameter estimated for SOC codes reports a between-SOC variance of 0.02
- The estimated individual level residual error variance is reported as 0.07
- The constant is 0.07
- The variance for working hours suggests that there is significant variation in individuals' income based on their working hours
- The positive covariance reported as -7/100=-0.07 for working hours across individual life course suggests a pattern of fanning in. This suggests that individuals at the mean working hours with higher initial log incomes tend to have slower increases in income with working hours increases, suggesting a grouping together effect (fanning in)

RANDOM EFFECTS

- Occupations make up a small but important amount of income variation
- Individual life course makes up a much larger proportion of income variation
- Equal amounts of variance related to error variance
- Variance in working hours sees a large amount explained by differences in working hours
- Covariance suggests a fanning in approach


Caterpillar Plot of Significant SOC Resdiuals
SOC Colour Coded
Predicted Stratum Random Effect
in Log Income
2
1
0
-1
-2
Stratum rank
SOC 1000
SOC 2000
SOC 3000
SOC 4000
SOC 5000
SOC 6000
SOC 7000
SOC 8000
SOC 9000
Data Source: UKHLS WAVE 1-13
n=139,431

OCCUPATIONAL VARIATION

- Displays residual variation of log annual net income across occupations
- Of the 347 occupations, 238 have a residual significant from zero
  - 63 occupations have a significant negative residual
  - 175 occupations have a significant positive residual
- SOC groups 1000-3000 and SOC 5000 have most occupations with positive residuals
- Comparatively SOC groups 6000-7000 and SOC 9000 all have most occupations with negative residuals
- Only SOC group 4000 have a majority of occupations that are neither have positive or negative residuals