Gender Pay Gap

The data is from Understanding Society: Waves 1-12, 2009-2021 and Harmonised BHPS: Waves 1-18, 1991-2009, found from the uk data service. We use from 2009 onwards for simplicity, and the unit of obs is person-year (superopeid and istrtdaty)

Pay transparency law in 2017 that requires all companies above 200 to report median gender pay gap (+ other summary stats). But cant really find linked data as literature has access to that use an event study, looking at people at firms bigger than 200 after 2017. Using a DiD/ES. But they have data on the women at these firms whereas we only have the data for women in general, we can def argue that understanding society is a representative survey, but even then we would only be finding the ATE (not sure if L – ATE or)

We can use understanding society data that is essentially big data and look at how the pay gap has evolved in response to this, but this might be too diff-in-diff like, as it should more be an application of stuff we have learnt in class. It is a longitudinal study of about a thousand survey answers or so ranging from background to pay to pension to house to food (I think)

So far in the code I have computed the average wage differential (this could be changed but just median of all men in a year – median of all women in a year). So, this is the average, we have every body’s actual wage, maybe we could compute their distance from the average wage differential (IE their wage gap is 75% larger than most women), then can we use model selection on the big data to see if this uncovers any inference (include immutable characteristics but also time varying?)

Or we can match men to women on a load of characteristics (possibly found using model selection), like this paper: <https://link.springer.com/article/10.1007/s00148-019-00743-8> which estimates the effect of being female on wages.

The pay variable we have is the recommended one, basic hourly rate, so accounts for pt/ft male/female etc etc