# The Gender Gap in Compensation among UK firms

An-Chiao Liu, Gabriel Naylor-Leyland & Jasper Ginn

The Gender Pay Gap defines the difference in hourly wage between men and women. This can be measured either by the mean wage, the median wage, or any other average. It is of interest not only to assess the size of the gender gap across a population, but also to uncover the driving factors behind its existence. For instance, are certain high-wage industries dominated by male workers? Does the population wage gap increase in positions of high wage? Do men and women earn comparable salaries for comparable jobs? Therefore the gender pay gap is 20%

```
# Number of males/females
gpg_core %>%
  # Subset for male / female columns
  select(male, female) %>%
  # Reshape the data from male, female column to variable, value columns where variable == female or ma
  gather(variable, value) %>%
  # Plot
  ggplot(., aes(x = variable, y=value, fill = variable)) +
    geom_jitter(alpha=0.2) +
    geom_boxplot() +
    theme_bw()
```
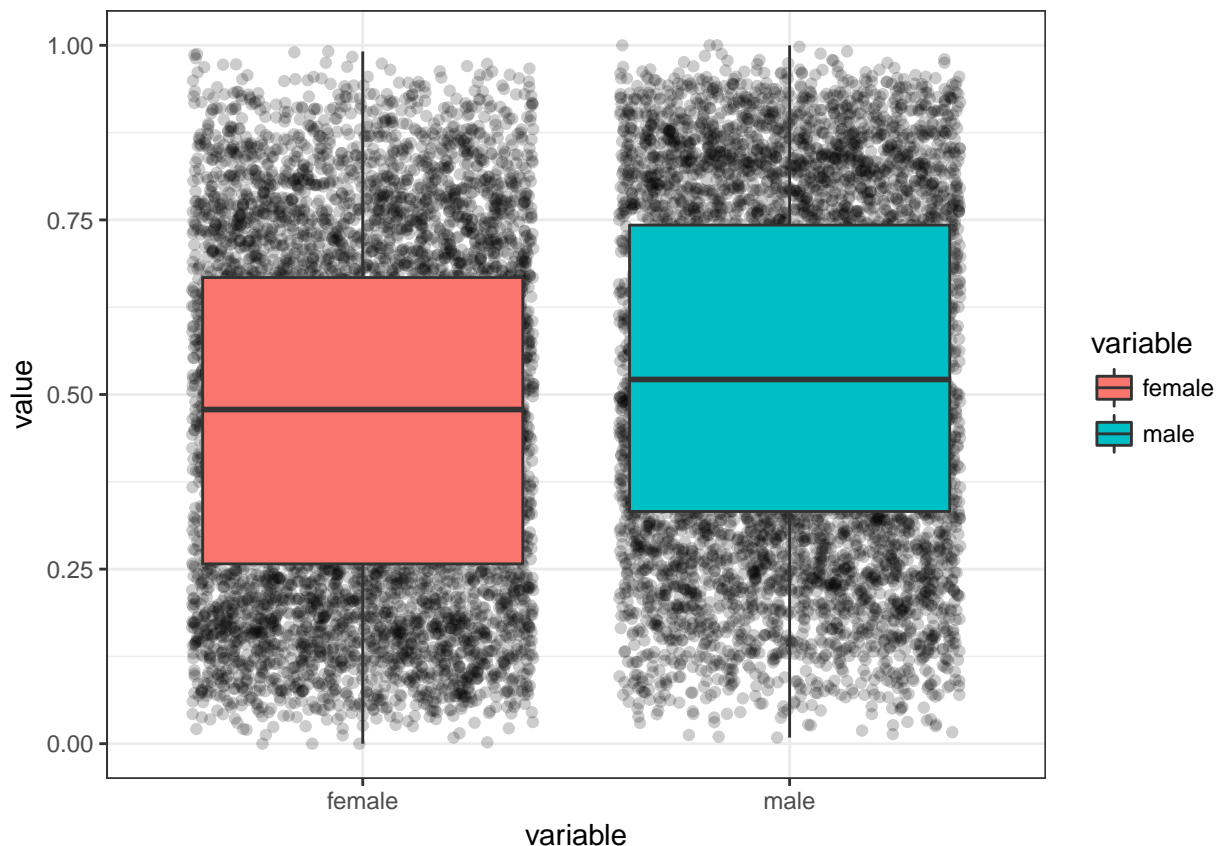

This report will focus on the wage gap in the UK in 2017. The data we analyse throughout the report comprise all companies with over 250 employees, and a small amount of smaller companies. As of 2017,

it is required by the UK government that all such companies publish information pertaining to the pay of employees for the purpose of showing whether there is indeed a difference in the pay of male and female employees. This dataset is well protected against non-response errors, given that failure to comply with the regulations can lead to enforcement from the Equality and Human Rights Commission [1]. In addition, non-sampling errors are also well protected via the use of clear and simple categories, for example, size brackets for the number of employees. Our aim is to create an efficient sampling design, from which we can make accurate calculations by using a smaller sample. We will compare the precision of a number of different such designs. Their accuracy will be measured against the results obtained from the complete data analysis.

One element of the data which we will refer to throughout this report is the Standard Industrial Classification (SIC). This is a widely used system for classifying companies into certain larger industries. Another important variable in the data is the EmployerSize, which categorises each company into seven distinct sizes: size not provided, less than 250, 250 to 499, 500 to 999, 1000 to 4999, 5000 to 19,999, and 20,000 or more. We will discuss in detail whether using information on the gender pay gap within these categories can lead to better sampling designs.

Before conducting our analysis, we first made some modifications to the dataset. Firstly, the information that was not necessary was stored in a separate dataframe, which we did not use. For example, the person responsible for the submission of the data was not relevant, and was thus not included in our main dataset.

Population data analysis

Our analysis found that, across the population (of provided data), the average monthly wage for men was 14.4% higher than that for women. It should be noted that this provides only an initial insight into the extent of the gender gap in the UK, and should be interpreted with caution. For instance, it can often be more relevant to assess the size of gender pay gaps in terms of the median wage, so that the few disproportionately high earners do not skew the results. To improve the accuracy of this 'population' gender gap we should have weighted the population mean based on the size of the companies, so that the larger companies hold more influence on the population statistics. However, with the data provided, it would be difficult to achieve a meaningful accuracy, given that we are not provided with the exact size of the companies.

A brief analysis into the prevalence of the gender gap, namely in how many companies there exists a disparity in wages based on gender, revealed the following: 749, or 11.15%, of the companies in the dataset exhibited a gender gap (in the average hourly rate) favouring women, whereas 5910 companies, or 88.03%, paid men a higher average wage. (The remaining 0.008% of companies exhibited no gender wage gap). This indicates that many of the industry-wide differences in pay are likely to favour men. These figures are accompanied by the histogram (Figure XX), showing a heavy skew towards a 'positive' pay gap. Recall that a positive pay gap represents one in which males receive a higher wage than females.

The gender pay gap in the UK is also evident in the distribution of bonus payments; defined as any payment additional to an employee's salary. The gender gap for bonus payments stands at 14.1% at the UK population level, again in favour of men. As previously, this statistic falls short of a true population average, owing to the lack of appropriate weighting. However, for the purpose of brevity, these will be referred to herein as the true population values. As previously mentioned, comparing the median (hourly) wage within a company can be a more effective way to study the gender pay gap. Therefore, by averaging the median wage across all UK companies, we found that the median male employee earns 12.2% more than his female counterpart.

A glance at the distribution of both positive and negative wage inequalities again proves interesting: 903, or 13.45%, of UK companies pay a higher median wage to women in comparison to men - a significantly smaller figure than the 5258 companies, or 78.32%, who provide a higher median wage to males. (The remaining 552 companies, or 8.22%, provide an equal median wage). Similarly to the analysis of the mean wage, we can see from the histogram (Figure XX) that the distribution is strongly skewed in favour of higher male income.

Using the data on the gender of employees working within each income quartile (herein referred to as lower, lower middle, upper middle, top), we were able to get a sense of the distribution of employees at

the company level. For example Figure XX shows that a higher concentration of women currently work in the lower quartile, while male employees dominate the top and upper middle quartiles. This is likely to be one of the driving factors behind the 14.4% nation-wide pay inequality. However, in order to further this analysis, we would require data pertaining to the size of the gender pay gap within each income quartile. For example, we know that males dominate the top quartile - and females the lower quartile - but we do not know in which quartile is the pay gap more pronounced.

Sampling methods

The focus of this report largely falls on sampling methods, with the aim of making inferences on the target population using a (smaller) designed sample. The target population for our sampling is the UK working population; the goal of assessing the gender wage gap is to do so in the context of its widespread presence in the entire population. In the current situation, whereby companies with over 250 employees are required to provide relevant data, we can work with a dataset containing 6713 companies. The sample size in mind when conducting the following sample analyses will be 1000. Thus, we endeavour to make calculations related to the gender pay gap on such a sample which accurately reflect the UK population.

Firstly, we opted to generate a sample of 1000 random companies out of the 6713 available, with each company given an equal probability of being sampled (1000 / 6713 = 14.9%). This method is known as a simple random sample (SRS), and requires the assumption that a random sample is representative of the target population. The sample is taken without replacement, i.e. no company can be included more than once in the sample.

Working with this SRS of 1000 companies, we found that the gender pay gap stands at 12.56% and 14.88% when assessing the median and mean wage, respectively. Both of these pay gaps are in favour of men. These values do not differ drastically from the population values of 12.2% and 14.1%. 95% confidence intervals for these percentages (.. , ..) show little variation from our original population values. The standard error for the estimate for the difference in mean hourly wage using the SRS is 0.0041. Using this sampling design, we achieve a margin of error of 0.0041*1.96 = 0.8% at the 5% level. This relates to the sampling error - the difference between the sampled statistic and the population statistic is just 0.8%. We will use this SRS as a reference to compare all future sampling designs against. This will be done via the design effect (Variance_design / Variance_SRS).

When the statistic of interest (here, gender wage inequality) is known to correlate in some way with information provided in the sampling frame, it can be useful to stratify. Arranging the data into strata prior to sampling can optimise the sampling design in the sense that the standard error of the 'predictions' can be reduced, thus improving the precision. This will only be the case if we see small within-strata variance, and larger between-strata variance. This will make for a more efficient design for the reason that, if a large portion of the total variance is between the strata, only a small sample from each stratum will be necessary to make accurate predictions.

Based on rudimentary analysis of the two potential stratification variables SIC and EmployerSize, and with the sample of 1000 still the goal, we decided to stratify based on SIC division. Evidence supporting this decision is provided in Figure XX: violin plots of the strata. From these plots we can clearly rule out EmployerSize as a useful stratification variable given the great variation in wage difference within each employer size category. Additionally, post-hoc tests following an ANOVA on the different categories of EmployerSize revealed very little evidence of between-group differences (only one of the 21 pairwise comparisons showed a significant difference). On the contrary, running an ANOVA on the SIC divisions yielded many significant pairwise differences in the median pay gap.

SIC division was identified as a more useful stratification variable. However, there was yet an obvious problem in that two of the divisions comprise just three companies each: activities of extraterritorial organisations and bodies, and activities of households as employers. We excluded these two divisions for the following reasons: 1. They comprise just 6 companies, making it difficult to sample in a representative way. 2. They

contain less than 0.09% of population. 3. The divisions themselves have unique, specific characteristics. Therefore, it is not possible to combine them with other divisions in a meaningful way.

We then used a proportional-to-size (PPS) procedure to draw a stratified sample of (at most) 1000 companies. This technique ensures that the drawn sample well represents the population, given that the proportions of the divisions within the sample match those within the population. For example, the division 'education' contains 986 companies, or 14.69% of the population. Within the PPS sample, education will comprise as close to this figure as possible. This method of sampling should ensure lower sampling errors, when comparing with an SRS. This is because a PPS stratified sample is expected to better represent the target population (in this case all UK companies with over 250 employees).

The PPS sample tells us that there is a 14.24% difference in the average wage - closer to the population value (14.4%) than the SRS (14.88%). The margin of error here is 0.74% (s.e. = 0.0037645), which gives a sense of how far our predictions of the wage difference differ from the population values. The margin of error is, as expected, slightly lower than that of the SRS (0.8%). The design effect of 0.8771 means a sample of only 877 would be required to achieve the same precision as our earlier SRS. These figures indicate that we can more efficiently (design effect) and accurately (margin of error) make inferences on the population with this sampling design.

However, it is still possible to optimise the sample, via the choice of stratum sample size. This can be done by using a Neyman Allocation. This method allocates an optimal size for each stratum via the formula $(N\_h * V\_h) / sum(N\_h * V\_h)$, where N_h and V_h are the stratum sizes and variances respectively. In this way, it is possible to oversample from certain strata - namely those with higher variances - in order to achieve a smaller standard error. According to this sampling design, the gender pay gap stands at 14.05%. While this underestimates the pay gap slightly more than the PPS sample, the standard error of 0.0037412 is marginally smaller in comparison. This means we now have a smaller, better design effect of 0.8685. Consequently, we consider this design more efficient than the PPS, requiring fewer companies to achieve the same precision. We therefore conclude that, when seeking to maximise precision within a sample of 1000 companies, stratifying based on SIC division and using Neyman's optimal allocation is better than taking an SRS or a PPS stratified sample. Other sampling designs are beyond the scope of this report.

Sample size determination based on the coefficient of variation

Another common way in which to sample is to base the sample size around a prerequired level of precision. In this situation, the precision of the sample estimates dictates the sample size, as opposed to a fixed budget or sample size. With this technique in mind, we aim to find the sample size required in order to achieve a coefficient of variation of 0.01. The coefficient of variation (CV) is defined as the standard error of the wage gap, relative to the size of the wage gap - a standardised error term.

Firstly, we calculated the existing CV, taken from the optimised stratified sample, to be 0.026. This is significantly higher than the target of 0.01. In order to calculate the optimal sample size in this situation, we can systematically increase the sample size, n, until the CV falls at or below the target of 0.01. This algorithm is implemented in R by the use of for loops and if statements. By doing so, we find that a sample of 3919 is sufficient. Therefore, with a level of precision as determined by a CV of 0.01, it would be necessary only to sample 58% of UK companies with over 250 employees, while still obtaining the same information.

It is also of great interest to sample at the level of the individual employee, rather than at the company level. This can be done with a cluster sampling design. This method requires the weighting of the data according to the size of the company. In absence of the exact company size data, we opted to use the midpoints of the size categories as surrogate company sizes. The largest size category was an exception; we used the value of 20,000 to weight these companies. Companies not providing the number of employees were weighted as 0, and thus excluded from this analysis. Using this sampling design, we found that a weighted average for the gender pay gap stands at 16%, in favour of men.

Discuss potential non-sampling, sampling, and non-response errors.

8. Another way to think about sampling is not at the company level, but at the level of the individual. Unequal pay at very large companies of course has a larger impact than at small companies. We lack data for companies with <250 employees, but could still weight the data according to the size of the company to inspect the gender pay gap for individuals in the UK who work at companies with > 250 employees.

a. Describe what you would do if the goal was not to study the gender pay gap at companies, but for individuals at companies.

b. Use the strategy you describe under a. to finally say something about the size of the gender pay gap.

References: [1] BRIEFING PAPER Number 7068, 28 September 2018 The Gender Pay Gap www.parliament.uk/commons-library | intranet.parliament.uk/commons-library | papers@parliament.uk | @commonslibrary By Feargal McGuinness Doug Pyper ## TODO:

- Stratification:
- Compare the two variables
- Recode the SIC variables
- Horvitz-thompson (Neyman) optimal allocation
- How to oversample from specific strata to reduce overall variance (See code Peter week 5 on stratification)

throwaway work: #show the violin plot here to show that the gap differs among size groups (the reason to stratify here) Another variable on which to stratify is EmployerSize. The data are already provided in seven categories based on the number of employees. Therefore, by using these as the strata, we can improve both the precision of the estimates within each strata, and thus the overall precision. We could not sample an equal number (1000 / 7) of companies from each strata, owing to there not being a sufficient number of companies in each strata. Therefore, we sampled from each size division with equal probability. (CHECK! THIS WAS CONFUSING). (We have roughly taken a proportion of $1000/6713 = 14.89\%$ from each size division). In order to do this, it was first necessary to calculate the proportion of all companies falling in each size division. This proportion was then multiplied by 1000 for each strata to achieve the desired sample size. Given that we preferred our sample to be at most 1000 companies, we rounded the final stratum sizes down where necessary. #report SE and confidence intervals, talk about the features (errors) of this particular design

Given that the premise of stratified samples is to increase the precision/ lower the SE, this sample has not performed its desired function. Therefore, it was necessary for us to implement an optimal allocation strategy, so that each size division is more apporopriately represented in the sample. It is of interest to take a larger sample from strata in which there is more variation. For this, we (attempted to use) used the Neyman Allocation. . . .