Family Business: Does Marital Status and Number of Children Matter in Understanding the Gender Wage Gap?

Alessandra Oshiro

4/9/2022

The gender pay gap is still an issue in most countries around the world. Therefore, studying which other factors other than gender alone can explain this phenomenon is important. For example, there is a wage difference that can be usually observed between mothers and women without children (Cukrowska-Torzewska & Lovasz 2016). Similarly, the usually implicit, but uneven housework distribution among couples living together could also be a source of disadvantage To find out whether factoring in household and motherhood responsibilities can improve our understanding of the gender pay gap, this paper will address the following research question: *Does living with a partner and living with children influence the wage difference between women and men?*

1. Data and Variables

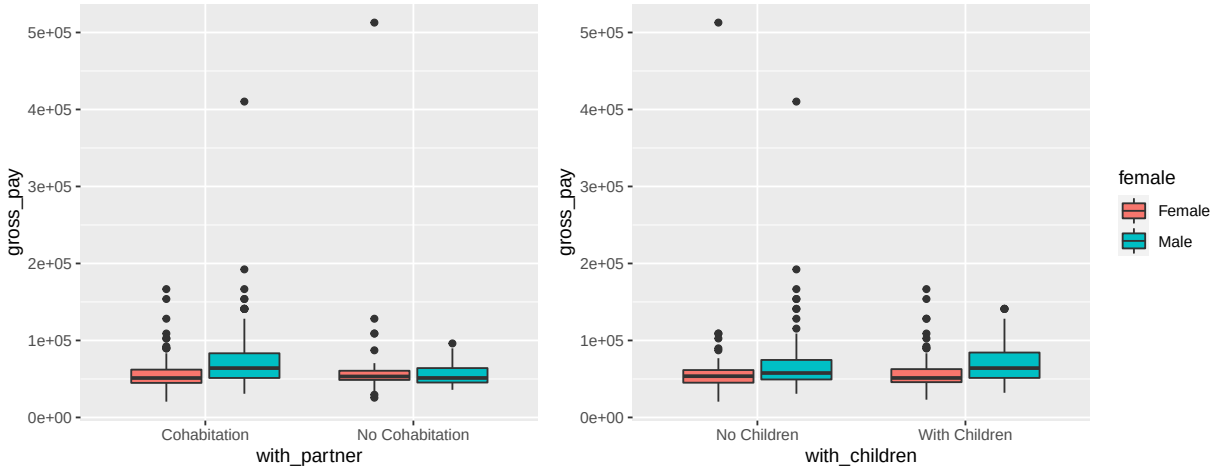
This study will use data from the fifth wave of the European Social Survey, given that it includes the "Work, Family, and Well-being" module. In particular, this paper uses data from the survey applied in Norway in 2010, and includes only respondents who report being fully employed (between 35 to 45 weekly hours of work). The reason for restricting the population to full time workers is to exclude wage differences that might be caused by variation in working arrangements. Therefore, it might not be adequate to generalize the results of the study to countries which show significant cultural and policy divergences.

The original dataset included observations that did not make sense: respondents answered that they worked 37 hours a week, and had a gross pay of 0 euro, for example. It is likely that those strange answers come from unwillingness to answer or from a misunderstanding, especially given that the ESS codebook does not specify whether the question refers to annual or monthly gross pay. These observations have been removed using the annual gross pay of the lowest paying job in Norway in 2010 (20000 euro, adjusting for inflation) as a criteria (Fløtre & Tuv 2022).

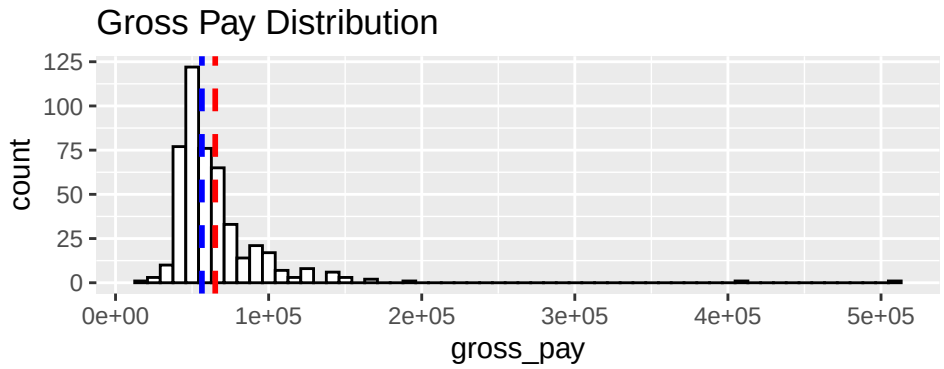
1.2 Variables

The dependent variable (`gross_pay`) is the individual's usual gross pay, converted from Norwegian kroner to euros. The independent variables (IVs) have been re-coded for clarity. The first IV is gender (`female`), which is coded 1 if the respondent is female, and 0 if they are male. The second IV (`with_partner`) reports whether the respondent lives with their partner or not, coded 1 and 0, respectively. The ESS included questions about marital status, as well as whether the respondent was living with their partner. I decided to use the latter for this study, because it is more likely to capture the cases in which women have to deal with household responsibilities, as some might live with their partners but not be formally married. The third IV indicates whether the respondent lives with children (`with_children`). It is not likely that the children living with the respondents are not their own, as all of them having at least one son, daughter, foster or adopted child living in their household in the ESS "relationship to person in household" (`rel`) variable. Finally, age (`age`) and educational level (`educ_level`) have been included as control variables, as they can be confounding. The `educ_level` variable has been re-coded as "Primary Education", "Secondary Education" (high school diploma or other non-degree studies after mandatory primary education), and "Tertiary Education" (bachelor's degree or higher) for simplicity. There were no respondents that stated having no education.

1.3 Descriptive Statistics



The sample consists of 471 observations, of which 177 are female, and 294 are male. As the boxplot on the left shows, women who live with their partners ($n = 130$) have significantly lower median gross pay than males who live with their partners ($n = 226$). In the case of women and men who do not live with their partners ($n = 47$ and $n = 68$, respectively) the gap in median gross pay is lower. The boxplot on the right shows a similar situation. The gap in median gross pay is lower between women ($n = 69$) and men ($n = 152$) that do not live with children, than what can be found when comparing women ($n = 108$) and men ($n = 142$) who do. Finally, there are several outliers in all groups, most of them being individuals with higher gross pay.



The distribution of the gross pay is also important to describe. The gross pay variable ranges from 20513 to 512821, while the median value is 56410 (blue line) and the mean is 6.510727×10^4 (red line). As it is common when working with income, this variable was not normally distributed. On the contrary, it presents a low-number inflated distribution. Because of this, several transformations were attempted: logarithmic, square-root, and cube-root. None of them achieved a perfect normalization, but the log transformation made significant improvements in contrast to the original gross pay data when assess with a Q-Qplot. Therefore, the models were run on the log-transformed data.

2. Method

The research question leads to several different hypotheses. In order to investigate whether living with a partner and the number of children are helpful in explaining the gender pay gap, it is necessary to look at the effect of each independent variable, as well as if there is any relevant interaction effect between them. Here, considering a potential interaction effect is important because it is possible that factors such as partner cohabitation or having children can have an effect on the gross pay for women, but not for men, as they imply household responsibilities that are usually taken up by females. Because of this, the analysis will test the following null hypotheses:

- H1: *Gender has no effect on the individual gross pay.*

- H2: *Living with a partner has no effect on the individual gross pay.*
- H3: *Living with children has no effect on the individual gross pay.*
- H4: *There is no interaction effect between gender and living with a partner in respect to individual gross pay.*
- H5: *There is no interaction effect between gender and living with children in respect to individual gross pay.*
- H6: *There is no interaction effect between living with a partner and living with children in respect to individual gross pay.*
- H7: *There is no interaction effect between gender, living with a partner, and living with children in respect to individual gross pay.*

To test these hypotheses, I will use three different ANOVA models, having set a significance threshold of 0.05. This method was chosen because ANOVA is the appropriate method when dealing with a regression-type question which has categorical explanatory variables, but a continuous outcome variable. Furthermore, the coefficients provided by the ANOVA summary table will allow us to test all the seven proposed null hypotheses. However, despite the data fulfilling the assumption of independent observations, gross pay is not a normal distribution and likely hetroskedastic. This needs to be taken into account when interpreting the results.

3. Analysis and Summary Tables

3.1 Model 1 ANOVA Output Table

To choose the best fitting model, three of them were tried. All of them had the log-transformed gross pay in euros (`gross_pay_log`) as dependent variable, however, they differed in the terms considered as independent variables. The first model (`model1`) included gender (`female`), cohabitation with a partner (`with_partner`), and their interaction as independent variables; as well as educational level (`educ_level`) as a control variable. According to this model, there is strong evidence to reject H1, as well as moderate evidence to reject H2 and H4. The R^2 score is 0.1928, which is quite low.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
female	1	2.69	2.69	25.63	0.0000
with_partner	1	0.82	0.82	7.80	0.0055
educ_level	2	6.76	3.38	32.15	0.0000
age	1	0.43	0.43	4.05	0.0447
female:with_partner	1	0.95	0.95	9.00	0.0028
Residuals	464	48.77	0.11		

3.2 Model 2 ANOVA Output Table

The second model (`model2`) included gender (`female`), living with children (`with_children`), and their interaction as independent variables; whereas educational level (`educ_level`) and age (`age`) were included as control variables. According to this model, there is also strong evidence to reject H1. Moreover, there is weak evidence to support the rejection of H3. Nevertheless, there is no evidence for rejecting H5, as `female:with_kids` presented a *p-value* higher than 0.05. For this model, the R^2 is even lower (0.1607).

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
female	1	2.69	2.69	24.71	0.0000
with_children	1	0.28	0.28	2.59	0.1082
educ_level	2	6.59	3.30	30.23	0.0000
female:with_children	1	0.14	0.14	1.28	0.2590
Residuals	465	50.71	0.11		

3.3 Model 3 ANOVA Output Table

The third model (`model3`) included all the relevant variables. Gender (`female`), living with children (`with_children`), and cohabitation with the partner (`with_partner`) as well as all the resulting interac-

tion terms were included as independent variables; while educational level (`educ_level`) and age (`age`) were added as control variables. Like the previous models, according to this one, there is strong evidence to reject H1. Similarly, there is moderate evidence to reject H2 and H4. Nevertheless, based on the resulting *p-values* no other null hypotheses can be rejected. This includes H3 which could be rejected by `model2`, but with weak support for it. The R^2 (0.1959) is higher than `model1` and `model2`. However, it is likely to be because it includes more terms as independent variables. The AIC comparison supports a cautious interpretation of the r-squared scores, as it points to `model1` as being the best one: the scores were 284.5804 for `model1`, 300.9206 for `model2`, and 290.7682 for `model3`. I believe this makes sense, as the first model is able to reject the same hypotheses as the third model, while being more parsimonious.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
female	1	2.69	2.69	25.51	0.0000
with_partner	1	0.82	0.82	7.76	0.0056
with_children	1	0.10	0.10	0.97	0.3261
educ_level	2	6.74	3.37	31.89	0.0000
age	1	0.54	0.54	5.11	0.0243
female:with_partner	1	0.83	0.83	7.87	0.0052
female:with_children	1	0.00	0.00	0.05	0.8307
with_partner:with_children	1	0.01	0.01	0.13	0.7162
female:with_partner:with_children	1	0.09	0.09	0.85	0.3565
Residuals	460	48.59	0.11		

4. Interpretation/Answers and Further Work

Despite the issues, the ANOVA tables provide interesting insights. First, H1 was able to be rejected in the three models, all presenting strong support for it. This might indicate that, at least among Norwegian full-time workers, there might be a gender pay gap. On the other hand, a surprising finding is that H3 could not be rejected, except for the second model, which did not include living with a partner as an independent variable. Although the literature has found that motherhood affects females' earnings, the result of this study could be explained by the existence of Norwegian policies aimed at supporting female workers who decide to have children. Therefore, it would be interesting to test these hypothesis in a wider rage of countries, which diverge on the level of support to female workers and on the number of policies aimed at reducing the gender pay gap.

Another surprising finding is that H2 and H4 could be rejected in all the models in which they were tested. This suggests that, in contrast to motherhood, living with a partner might help explain the gender pay gap in Norwegian full-time workers. The gender aspect of this is highlighted by the fact that H4 concerned the interaction effect between gender and living with a partner. Again, this might be explained by the fact that, contrary to motherhood/parenthood policies, there are no regulations for the distribution of the housework. Nevertheless, these findings need to be taken with a grain of salt, given that all of the three models which where tried showed a relatively low goodness-of-fit. The need for a careful consideration of the outcomes is further emphasized by the fact that the data deviates from the assumptions made by the ANOVA method. As mentioned before, the choice of method was mostly guided by the levels of measurement of the variables used.

5. References

Cukrowska-Torzewska, Eva and Anna Lovasz "Are children driving the gender wage gap? Comparative evidence from Poland and Hungary." *Economics of Transition* 24, no.5 (2016): 259–297. DOI: 10.1111/ecot.12090

Fløtre, Ingvild Alseth, and Nina Tuv. "Hva er vanlig lønn i Norge?" *Statistik Sentralbyrø*, March, 29th 2022. https://www.ssb.no/arbeid-og-lonn/lonn-og-arbeidskraftkostnader/artikler/hva-er-vanlig-lonn-i-norge

ESS Round 5: European Social Survey Round 5 Data (2010). Data file edition 3.4. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC. doi: 10.21338/NSD-ESS5-2010.