Hagtíðindi
Statistical Series

Greinargerðir
Working papers

7 September 2021

Icelandic gender pay gap analysis 2008-2020

Rannsókn á launamun karla og kvenna 2008-2020[1]

Introduction

Goal

The purpose of the present paper[2] is to estimate the difference in hourly wages between female and male employees and its time evolution in Iceland. The paper is part of a study conducted in cooperation with the Prime Minister's Office. The wage difference estimates account for a set of measured individual and employment characteristics which include experience in a given company, age and other demographic attributes of employees, education, occupation, economic activity of employer, female/male proportion of employees in the occupation category, economic sector and activity, size and location of company.

On the one hand, the gender wage gap is due to differences between the characteristics of female and male employees. On the other hand, the gap is driven not only by this isolated (*additive*) gender effect but also by the way its impact *varies* according to the values of the employees' characteristics (*interaction* effect). One cannot thus isolate the effect (on wages) of being a woman/man employee from the effect of having a given level of education or being a parent or having a given occupation.

One may therefore answer questions like: is the effect (on wages) of working fulltime or being married or having a non-Icelandic background different for men and women? Does this change with time? Does the wage change with the level of education or with the economic activity in a different way for men and women? This shows, as is pointed out in [1], that „the concept of a single gender pay gap is a too simplistic representation of reality“.

For the Icelandic data, we conclude that occupation, employment characteristics and economic activity explain a large part of the differences in wages, but also that several characteristics of employees/employment have significantly different effects on wages, depending on gender. We investigate in addition how the gap depends on characteristics *away* from the reference or mean characteristics.

---

[1] See Icelandic Working paper at http://hagstofan.s3.amazonaws.com/media/public/2021/93d565e8-b337-4013-9fee-855b080681f4.pdf

[2] **Acknowledgement**: *Authors are grateful to Dr. Kimberly Zieschang, Anton Örn Karlsson and Eyjólfur Sigurðsson for most useful and extensive discussions.*

Wage differences are even more complex than the ones described by using only their means, since for each gender and any combination of characteristics' values one may observe an entire distribution of wage differences. The difference in means is however the descriptive measure most often used in literature and the main object of this study, but other distributional aspects (e.g. spread, skewness) or equivalent distributional descriptors are desirable and will be briefly addressed.

In addition to point estimates, we report uncertainty measures of the modelled wage differences which depend on both model choice and model parameters and we employ for this purpose both frequentist and Bayesian statistics.

Focus and main findings

We focus on building the statistical (multilevel) **models** needed for testing the main research hypotheses and on explaining the models' results.

The present analysis has two main types of **findings**, concerning:

(i) the *isolated effect of gender* on wages, summarised by the so-called *adjusted wage gap*[3], particularly useful for comparisons with similar studies for other time periods and/or data sets. By using additive wage models, the average wages are compared between employees of different genders but identical with respect to other attributes, at these attributes' fixed (reference or mean) values.

Our estimate of the *adjusted values of the wage gap*, i.e. the difference in means obtained as a result of (additive) *modelling while accounting for the influence of other covariates than gender and for data complex correlation structures*, is negative and has decreased in absolute value from about 6.4% in 2008 down to 4.9% in 2016 and to about 4.1% in 2020. These values are in rather good agreement with previous publications of Statistics Iceland [H3], regarding the years 2000-2016 and the comparison is one of the main reasons for fitting these particular models.

This measure has been estimated for *separate economic sectors* as well. The different values of the gap for such subsets are already a manifestation of the fact that there is a significant *interaction* effect to be taken into account, in this case between gender and economic sector variables. The absolute value of the gap on the private sector (A) has decreased from 6.7% in 2008, down to 6.1% in 2016 and 5.6% in 2020. In the same period, the gap on the government sector (R) has decreased from 4.9% down to 3.3% , while for the municipalities sector (M) from 6.1% in 2008, down to 2.7% in 2020.[4]

The **declining trend of the adjusted wage gap** and the accompanying changes in the main effects of the other covariates related to it are illustrated in Table 4 (4a - for the fuller model M and 4b for the simpler model M1). For example, the (additive, main) effect on wages of age, length of employment with same company, marital status, not having an Icelandic background, being a supervisor, having higher education levels, has not changed much during the past few years and remained positive, while the negative effect of having children younger than five years old has increased (in absolute value).

[3] This gap value is thus defined as the difference between the mean hourly wages of women and men, divided by the mean value of men's hourly wages, **while** controlling for any other variables. When men have higher wages than women, the gap is negative.

[4] A *cautionary remark* is in order: most of the time, values like the ones listed above make most sense when specifying the models, the covariates and their reference values.

The positive effect of having an occupation with a balanced mixture of women and men employees has increased as well[5].

Answering questions regarding how and whether those effects depend of gender is the object of our second category of findings, listed below in (ii).

Table 6 shows the evolution in the explanatory impact of the main (groups of) covariates, comparing the years 2010 and 2019. For instance, the occupational and work related characteristics (fulltime, length of employment, hour worked) have the biggest explanatory impact which has changed very little over ten years. The economic activity has also a large contribution to explaining the gaps, which has declined slowly with time (relative to total gap). The variability in wages explained by grouping according to occupation, economic activity and company has decreased during these years but remains high.

This however is a very incomplete picture as we show in what follows!

(ii) the *effect of various characteristics (of employees and work conditions)* on *the gender wage gap*. This is related to the fact that a proper wage model contains significant (gender-characteristics) interaction terms or, equivalently, to the fact that two significantly different wage models should be fit for the two genders. It is most frequently summarised by the *decomposition* of the wage gap into components (un-)explained by differences in characteristics.

Figures 9, 10 and Table 5b show how much of the gap is explained by the differences in (average) **characteristics** of men and women, versus what is due to the differences in **the *effects* of these characteristics** on wages for the two genders (i.e. by significantly different wage models for women and men). Such decomposition is not unique, see Table 5b, and it depends (for a given model) on the choice of the reference model used when doing the comparisons. One may also criticise it from the point of view of the hidden mechanisms behind the differences in characteristics. For example, even though people in the same occupation are paid equally, the career choices leading to a given proportion of men or women in that occupation may be quite biased in many cases [45]. Same argument can be made about other attributes, so that the "explained" part is in fact hiding extra, unaccounted complexity.

Table 3 illustrates the fact that certain covariates have a statistically significant advantageous effect on the wages of female/male employees. It also exemplifies how this significant effect changes with time. In Table 2 we show in detail that the interactive model and the separate different models for the two genders are an equivalent description of the interdependence between the gender variable and certain covariates.

For example, women gain **less** than men with increasing age and with increasing length of employment in same company, by being a supervisor or being married. On the other hand, women gain **more** by being in a labour union then men do, by working in a medium sized company, being highly educated and working in the government sector or for municipalities.

Both genders gain **comparable wage advantages** by working in occupations with a balanced mixture of men and women versus occupations dominated by women, and/or for a company with equal pay

[5] The corresponding variable measuring this attribute of occupational groups is *categ_propF* and defined in the section entitled „Data and model variables".

certification. Differences of occupational composition effects are non-significant in Iceland, unlike differences observed in other countries where significant advantages are reported for men, see for instance [44].

However, when looking at more than one attribute, the differences between genders become even more interesting. For instance, the advantage of being a woman in a medium size company is by far offset for a woman supervisor when compared with a man supervisor in the same type of company.

One may say that indeed, the group on the labour market with lowest wages consists of, according to Table 3, young women, recently hired by a (big) company, not married and/or with small children, who are not supervisors, not of Icelandic background, not working fulltime, in the state sector or for municipalities, in an occupation with predominantly young[6] employees and not in the capital area. This result has been confirmed in Icelandic income and living conditions surveys.[7]

The differential effect of occupation, economic activity and company has been tested by allowing the gender slope in the multilevel models (as described in what follows) vary according to these clustering factors. This proved that indeed, there exists a significant separation of gender effects by occupation, economic activity and company although the variability in wages explained by these factors has declined with time. Tables 3 and 4 give a simpler example of this latter trend (not including random slopes), where intraclass correlation, measuring how similar cross-classified observations are, decreased with time.

Methods, models and motivation

The data structure determines the type of models to be used, in our case *multilevel (linear) models* (MLM) [3]. These can be regarded as **more general** than ordinary least square (OLS), fixed effects (FE) or random effects (RE) models, being able to **accommodate complex error structures** which exist in data sets like ours, with repeated observations for same individuals and/or multiple correlations between individuals due to shared (cross classifying) attributes of employment, e.g. working in same company, having same occupation or working in same economic activity. The same feature allows for **better use of data** when very few records are available for certain cross-classifications, due to modelling.

Ignoring clustering effects in data and using simple regression models would produce biased significance tests since standard errors are underestimated, i.e. inflating the *significance* of associations. This is mainly due to the fact that the main OLS assumptions are not valid for these data structures. Moreover, when the relation between outcome and predictor is different between groups versus within groups, MLM will identify it correctly while OLS will only find the overall relation (see Simpson‘s paradox, i.e. trend in combined data is the reverse of the trend in each group [3]).

One has two choices when fitting these models, i.e. using a frequentist or a Bayesian approach. The frequentist estimates (obtained by maximum likelihood - ML or restricted maximum likelihood - REML) underestimate the parameter *uncertainty* to some extent but are obtained faster than their Bayesian companions. A middle ground is however achievable by creating simulations-based posterior distributions for the parameters of interest which also have the advantage of being easily interpretable.

---

[6] Less than 35 years old, see the variable *categ_propY* in the section „Data and model variables“ for detailed definitions

[7] See, for example: https://hagstofa.is/utgafur/frettasafn/lifskjor/felagsvisar-serhefti-um-fjarhag-heimila/

Bayesian methods are also useful when trying to decide which variables should enter the model. For instance, multi-model Bayesian inference would sum and average across the space of all possible models, weigh them according to their posterior probabilities and give the posterior probability that each effect is statistically significant, for all (ranked) models. We used this method as a preliminary validation of the model components.

In the present analysis, we build multilevel models with random effects (variability) of (i) individuals' *intercepts*, which are due to the temporal correlations between repeated observations and (ii) for company, occupation group and economic activity variables which are due to the correlations between observations belonging to the groups these variables define. We compare models with and without random effects on the *slopes* of several predictors and various error correlation structures in order to select the most parsimonious and well performing (in terms of computing time) one.

For time growth models of the regular hourly wages[8], independent variables (referring to individual attributes or to company, occupation and economic activity attributes) are also included with fixed/random intercepts and slopes and their interaction with the gender variable is tested (cross level interactions).

Individual characteristics belong to two categories: time-varying and constant in time. The time invariant characteristics only have an effect on wage levels when comparing different individuals while the time-varying ones (like age, experience, length of employment in company, total working hours) are changing both *within* each individual, mostly increasing with time, and *between* individuals. These effects are not always identical and are estimated accordingly.

We test the hypothesis that the main individual and work related characteristics have different effects on hourly wages of men and women. This is most efficiently done by assessing **the joint statistical significance** of the interactions between the gender variable and the characteristics of interest, in a unique model. If the interactions are significant, then *one cannot isolate the effect of gender* from the effect of the characteristics and one should not report a unique measure of the gender effect.

In this case, an equivalent approach for comparing the conditional means (at fixed values of the other predictors: either means of continuous ones if centred, or reference levels of categorical variables) of the hourly wage distributions for the two genders is to fit two different models, one for each gender, and compare the conditional means (calculated for common reference values of the covariates).

Two main types of pay gaps are of interest:

(i) The ***un-conditional*** wage gap:
If the wage models contain only an intercept and the gender dummy variable (with or without random error terms, depending on testing purposes), then the results estimate the mean wage for each gender, i.e. irrespective of other possible predictors. The models can be fit for each time point or as a joint growth curve with reference point defined by the start or end of the analysed period and with estimated rates of change for time and gender effects. If random effects are included in the model, the gap is corrected for the bias due to clustering of similar items.

(ii) The ***conditional*** wage gap:
If the wage models contain multiple additional independent variables, then the results are estimates of the conditional means of wages for each gender, which depend on the

[8] All basic wages paid for both day-time and shift-work hours as well as fixed wage contract hours, including additional payments and bonuses settled regularly in each wage period by the employers. Overtime and other irregular payments are excluded as well as employers' social contributions, including employers' payments in pension funds, and taxes

reference values chosen for the additional „explanatory“ variables, on the model itself and the available data. In this case, one evaluates the size and uncertainty measure of the MLM pay gap.

Decomposing the pay gap „explained“ by the (differences in) characteristics included in the model into independent contributions due to various factors has been traditionally done for OLS models by using the Oaxaca-Blinder techniques. In the present paper, we build the corresponding decomposition for the multilevel models at fixed time. We also describe the contributions of each characteristic (group) to the wage differences between women and men and its time evolution.

Data and model variables

The data-set used for this analysis consists of about one million records from Statistics Iceland's data on wages combined with demographic and employment data and covering the period 2008-2020[9].

The following set of variables has been used for modelling:

- **outcome** of interest (*wageHourly*): (logarithm of) regular hourly wages, observed yearly, for individual employees
- variables which **group** observations into clusters:
  individual identifiers (*id*, needed due to time correlated observations for each individual),
  company identifiers (*company*) ,
  Nace2 classification codes of economic activities (*nace2*),
  occupation codes (*occupation4*, 4 digits)
- **individual** attributes :

  - education (*educ1*, encoded as e2=10:29, e3=30:49, e4=50:69, e5=70-89 *ISCED* levels), length of employment in company (*lenEmployComp*) and its squared value, (scaled-) total hours worked (*totalHoursScaled*, i.e. divided by 365), age (*age*) and age squared (only mean centred when models were fitted for fixed time values but decomposed into age-*within* and age-*between* individual variations and centered accordingly when time growth curves were modelled).
  - fulltime working (*fulltime*), labour union membership (*inlabunion*), registered apprentice (*regapprentice*), registered student (*regstudent*), background (*backgr,* as Icelandic or not), supervisor (*supervisor*), craft worker (*ctworker*), monthly earnings (*monthlyearn*), shift premium (*shiftPremium*), all these variables having only 0,1 values.
  - marital status (*marital*), having children of ages less than 2 (*childage0to2*), between 2 and 5 (*childage2to5*), or between 6 and 16 years old (*childage6to16*), all binary variables as well.

- **company**[10] attributes:

  economical sector (*econSect*, A – private sector, R- state sector, M - municipalities), size of company (*sizecompany*, small: less than 49 employees, medium: between 50 and 249, large:

[9] Wage data are based on survey on a sample of private companies and municipalities (local government) with 10 or more employees. For central government wage data cover all its employees.
[10] This is a generic term, which does include private sector companies but also employers from the state economic sector and municipalities.

over 250 employees), capital area location of company (*capitalareaComp*: 0 or 1), equal-pay certificate (*equalpay*: 0 or1).

- **occupational** attributes:

  proportion of women employees (*categ_propF*, low: < 33.3%, medium: between 33.3 and 66.6%, high: > 66.6%),

  proportion of employees older than 35 (*categ_propY*, low: < 33.3%, medium: between 33.3 and 66.6%, high: > 66.6% employees).

Previous studies

Statistics Iceland has previously analysed the gender pay gap and published several papers in the Statistical Series. The additive gender effect has been evaluated by OLS [H-3] and random or fixed effects models in addition to separate models (fixed or random) were built for the two genders [H1]. Oaxaca-Blinder decomposition was calculated for these models. A very detailed analysis was using classical regression in [H2]. The results showed an adjusted gender pay gap (as absolute value) which decreased from 6,6% in 2008 down to 4,5% in 2016 as shown in [H3].

Comparisons and comments

The results of the present and previous studies are broadly comparable. The models include sets of variables which are overlapping to a large extent and their levels are rather similar or identical. *New* variables included in the present analysis concern the gender and age composition of occupation categories (occupations are classified according to the proportion of women employees and according to the proportion of employees older than 35), as well as several demographical (marital status, children of various ages) and employment attributes (type of company in terms of size and location, type of employment).

The present methods are more efficient at using data in addition to being more general, applicable to data with complex correlation structures. They provide results with *decreased uncertainty* of estimates (the Bayesian more than the frequentist versions) and *improved accuracy* of estimates but the present point estimates are not very far from the corresponding results from the most recent study which focused on classical regression models. The similarity is due to the fact that the most parsimonious models used in the present analysis contain fixed effects for most covariates and only random (intercept) effects of the cross-classifying variables. Models with random slopes have been tested but the gain in performance was much smaller than the loss in efficiency and with the danger of overfitting present they were therefore not adopted for the final estimates.

Details of modelling and results

Descriptive statistics

Detailed descriptive statistics for the whole data set is given in Table 7 and Table 8. Table 7 contains the description of all variables used in this study, with ranges, means / frequencies depending on their type. Table 8 shows the observed differences in characteristics, between women and men over the whole studied period.

Exploratory analysis

Before any modelling, the dataset is explored in terms of correlations and probability density distributions. We show here several of these results which are relevant for the next stage.

Main correlations

The main correlations between the main variables in the data set are useful as indicators of their *potential* usefulness in modelling the outcome.

Figure 1. Correlation of main variables in the data set.[11]

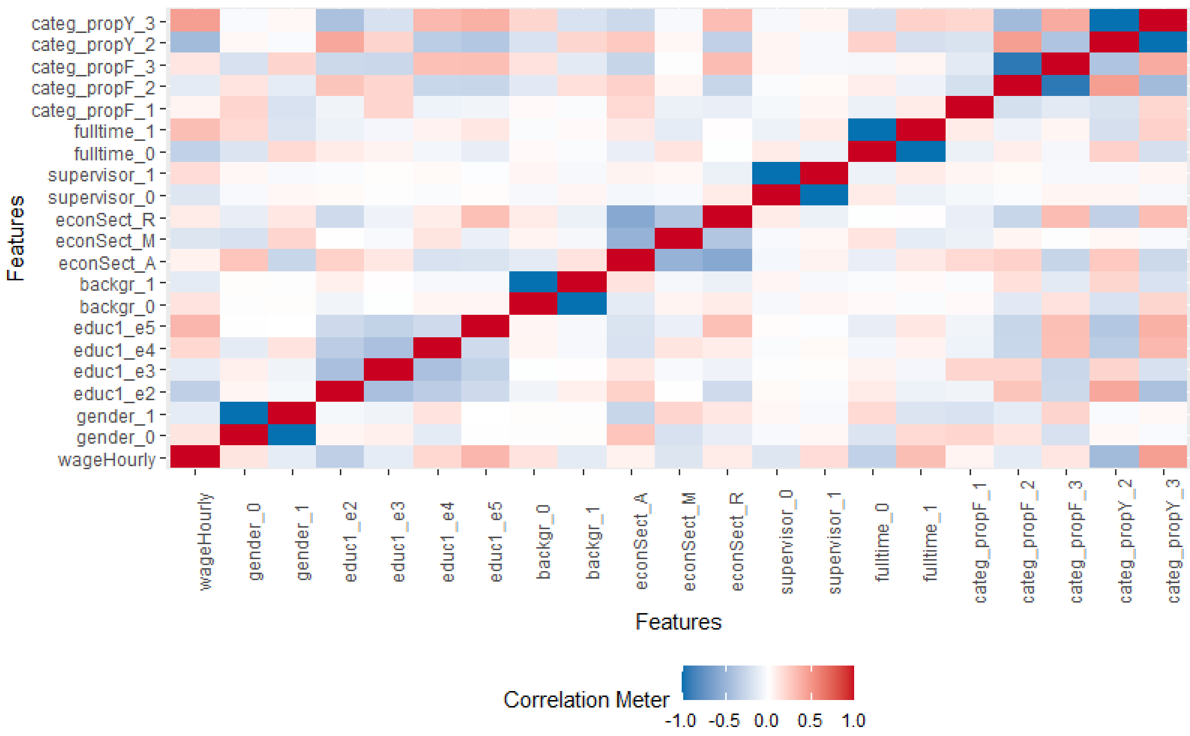


Figure 1 shows for instance that hourly wages are:

- positively correlated with: gender=0 (male employees), high education levels (e4, e5), and the A (private), R (state) economic sectors, occupations with high proportion of employees older than 35.
- negatively correlated with: gender=1 (female employees), although rather weakly, lower education levels (e2), not working full time, municipalities (M) sector.

One may also see that the higher education levels and the occupations with highest proportions of women are positively correlated, therefore, the more women, the more employees with higher education levels are in a given occupational group. The highest education levels are in addition positively correlated with the state (R) economic sector. Employees with non-Icelandic background are positively correlated (although mildly so) with lower levels of education, the private (A) economic

[11] Using the R-package DataExplorer, Boxuan Cui (2020). DataExplorer: Automate Data Exploration and Treatment. R package version 0.8.2. https://CRAN.R-project.org/package=DataExplorer

sector, occupations with medium proportions of women and medium proportions of employees over 35.

An important note is that, while hourly wages are sometimes rather strongly correlated with some of the other variables, those variables are most of the time only weakly correlated *to each other*, therefore the danger of multicollinearity in a model using most of them is rather low, as confirmed by our (VIF2) statistical tests.

Density distributions of wages

The comparison of density distributions of wages shows how different the number of men with any given wage is from the number of women with same wage.

Figure 2. Probability density of (log) hourly wages, for male and female employees and whole dataset.

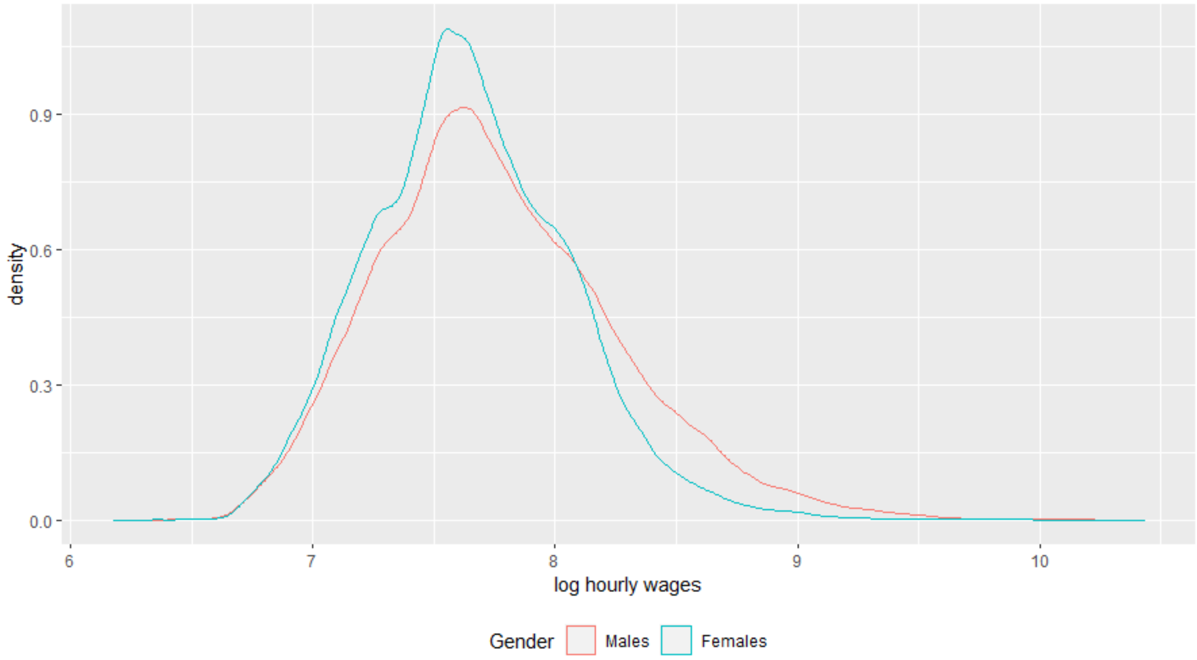
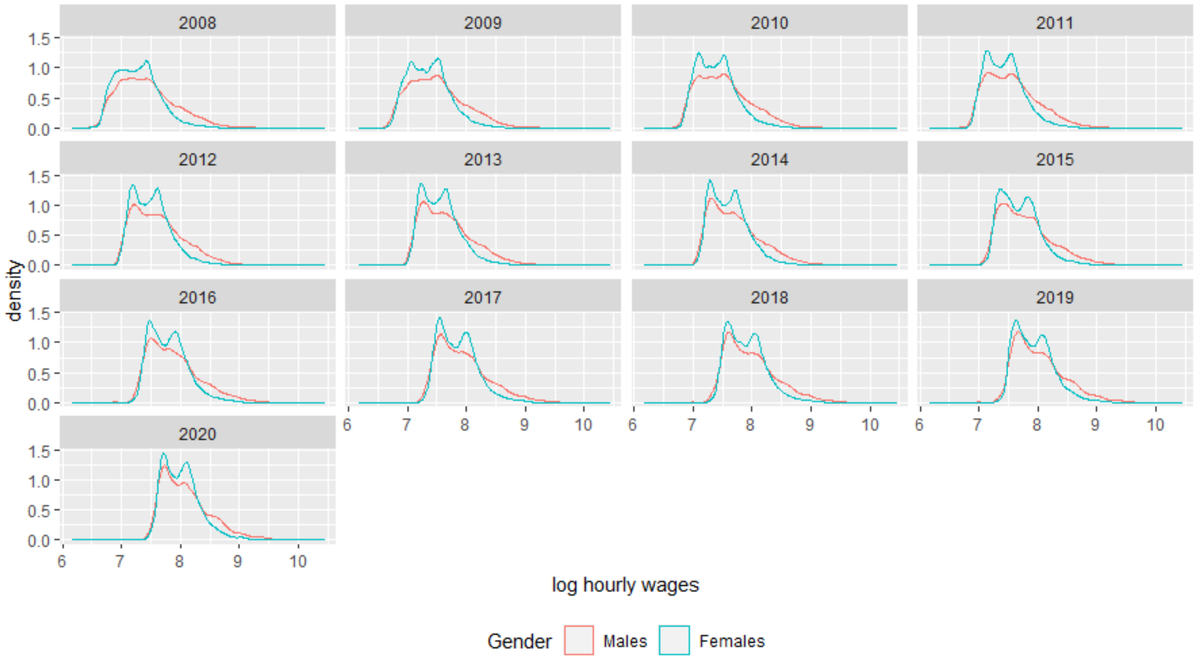


Figure 2 shows that for Gender=0 (male employees) the distribution of hourly wages is more skewed towards higher values of wages than for Gender=1 (female employees). This means that more men have higher salaries than women. At the same time it shows that more women than men have lower wages.

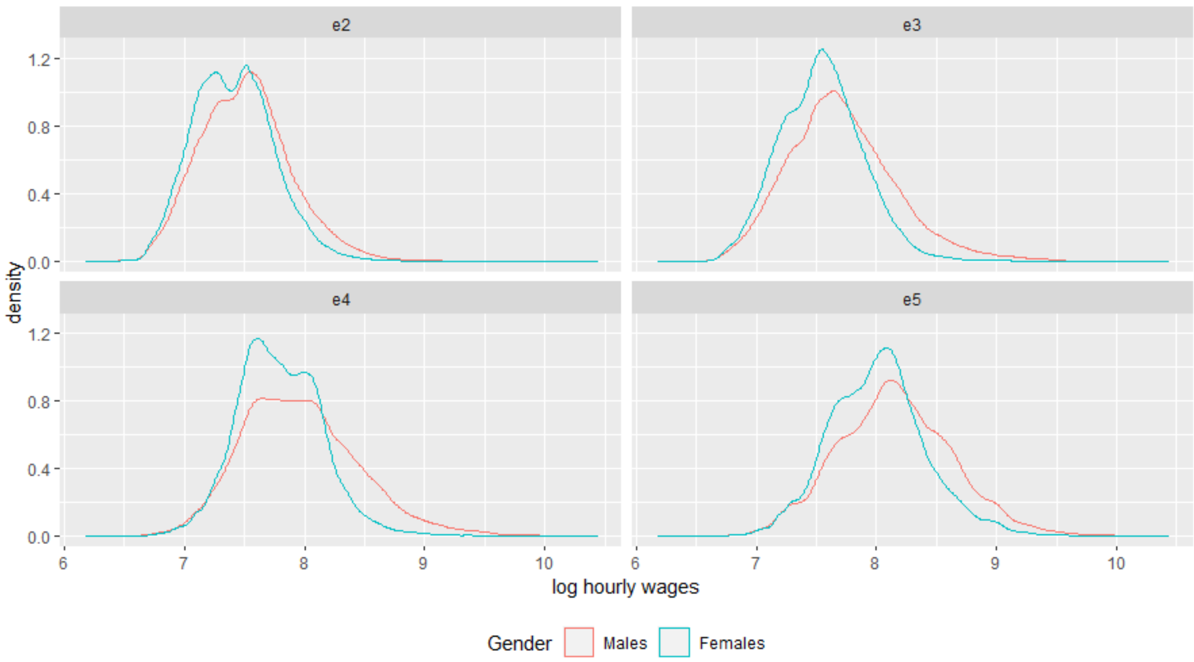
These density distributions evolve through time, although the main features are preserved to a large extent, at least when looking at the economy as a whole (see Figure 3).

Figure 3. Probability density of (log) hourly wages, by gender and for male and female employees.



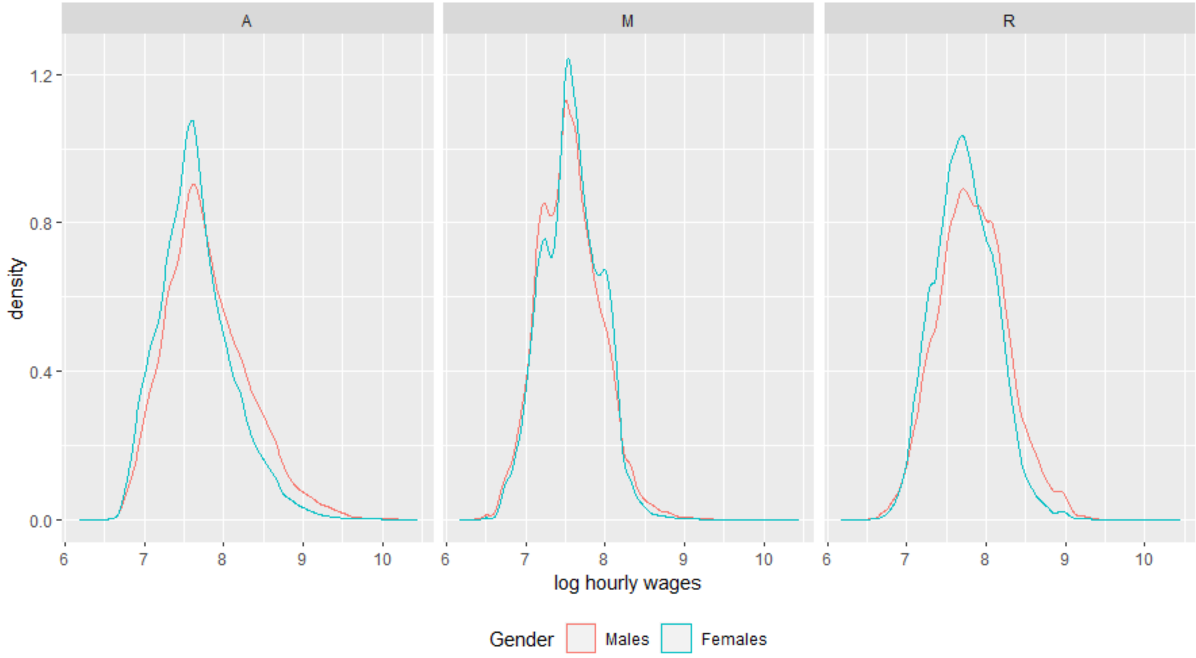
The wage distributions may be compared for employees with same education levels, as shown in Figure 4. The comparison still shows skewness towards higher salaries for men and some variation of disparities by education levels is apparent.

Figure 4. Density distributions by education levels (grouped as e2=10:29 ISCED levels, e3=30:49, e4=50:69, e5=70-89 ISCED levels), for male and female employees.



The results are rather different when comparing wages by economic sectors. A very equilibrated wage distribution is then seen for men and women employees in sector M (municipalities) while sectors A (private) and R (state) display the usual advantage for men albeit smaller for the state sector.

Figure 5. Density distributions by economic sector, for male and female employees.



Modelling and several cautionary methodological notes

Models

The multilevel models built in this study were **fitted** using:

- maximum likelihood, for the purpose of comparing the performance of several models
- restricted maximum likelihood, for estimation purposes.
- Bayesian framework, for best uncertainty estimates and for Bayesian variable selection and model averaging purposes.

The models have the following **structures**:

- null models, for testing the clustering of observations and for providing a baseline to more complex models: include only the random effects
- three level models, when modelling data of multiple years with a unique model.
  This was necessary especially when not sufficient observations were available for fitting the multilevel model for given fixed points in time. In this case, the intercept and slope depend on individual characteristics (level 2) and on company/economic activity/occupation attributes (level 3). Level 1 describes the time - autocorrelated observations within each individual. Interaction of (cross-) level attributes and significance of random gender slopes were tested. A typical model may be written as:
  $$y_{tjk} = \alpha_{000} + \beta_{100}t_{jk} + \beta_{010}X_{jk} + \beta_{001}Z_{jk} + (\beta_{110}X_{jk} + \beta_{101}Z_{jk} + \epsilon_{10k} + \epsilon_{1jk})t_{jk}$$
  $$+\epsilon_{00k} + \epsilon_{0jk} + \epsilon_{tjk} + \dots$$
  with and without interaction terms between the gender variable and several characteristics. In this formula (written in condensed form here but equivalent to the MLM usual layered equations as shown in [3] for instance) we denote by t_{jk} the time points of observing (log) wages of individual k in group j. The formula encodes products like $\beta_{010}X_{jk} = \sum_l \beta_{010}^l X_{jk}^l$, where X_{jk}^l are all model characteristics (l) of individuals (k), including gender (0,1), cross-

classified in groups (*j*) and where Z_{jk} are attributes of higher level structure (not changing between individuals of same group), i.e. of companies, economic activities or occupation groups. The structure of random effects is encoded by the variance-covariance matrix of $(\epsilon_{0jk}, \epsilon_{1jk})$ - individual level and $(\epsilon_{00k}, \epsilon_{10k})$ - group level. We converged to simple, diagonal matrices, after several exploratory model/computational tests. We omitted superscripts like „*l*" from terms like $\beta_{010}^{l} X_{jk}^{l}$ or $\beta_{001}^{l} Z_{jk}^{l}$ for simplifying the equations.

An added complexity of these models is that usually the variation of continuous individual attributes should be separated into within/between components $X_{tjk}^{within}, X_{jk}^{between}$, i.e. the part that changes/is constant in time for a given individual.

- two level models, when modelling data for fixed time values. A simple type may be written as:

$$y_{jk} = \alpha_{00} + \beta_{10} X_{jk} + \beta_{01} Z_{jk} + \epsilon_{0k} + \epsilon_{jk} + \ldots$$

with and without interaction terms between the gender variable and several characteristics. This solution was used for modelling the most recent years' data, since sufficiently detailed and high quality for fitting separate models for distinct years. It offers the advantage of easier interpretation of results, even for the interactive models and fast computation.

The variables included as *X, Z and random effects* are listed in Table 1 to Table 4 for the main models.

Interpretation

We remind the reader of the simple interpretation of modelling results like the ones included in the tables at the end of this report and how to extract information of interest from such tables.

In Table 4a, for example, one may see the model M, an **additive model**, fitted at several fixed time values. Let us look at year 2019, which is also found in Table 1.

It states that the mean value of the logarithm of the hourly wages, the so-called intercept of the model, given that the other variables included in the model are at their reference values, is 7.570. Since the reference value of the variable gender is gender=0, this also says that this is the mean value of log-wages for men. To this mean, the random effects add corresponding „noise" according to their variances (see below) and describing how the wage varies with the company, economic activity and occupation.

The coefficient of the *gender1* variable is -0.044, indicating that the mean log-wage of women is different from the value we saw for men and is 7.570-0.044=7.526. We emphasize here that these statements are for logarithm of wages. When transforming back to real wages, by using exponential function, we find that **the adjusted wage gap** is exp(-0.044)-1=-0.043 or 4.3%.

The next variable in the model has reference value zero: I(age-mean(age))=0, i.e. the reference age is the mean age of employees (since centred on this value). The coefficient is 0.003 and it indicates that, for every extra unit of this variable (in our case years) added to its reference value, the value of the outcome (*log-wage*) will change with this very 0.003 amount.

Another categorical variable like *educ1* (education) has a reference value educ1=e2 and the model tells us that all employees with education level higher than this will gain an extra 0.03 to their *log-wage* if they have level e3, or 0.08 if they have level e4, or 0.17 if their level is e5.

Similar interpretations can be given to the other continuous or categorical variables in the model.

The part of model summary concerning random effects (in this case only on intercepts) gives information about how much of the within versus between groups variability of the log-wages is explained by these variables. ICC=0.63 indicates that much of variability in wages is explained by cross-classifying observations (employees) according to occupation (>), economic activity (>) and company.

If we chose to analyse Table 2 or Table 3, we find models (M*) with **interaction** between gender and the other covariates. In Table 2 we show how fitting separate models is equivalent to a unique model with interaction included. Let us look at the last model in Table 3, which tells us not only **how much gender matters, in terms of wage, but also how this *depends on other* work related and demographic *attributes***.

The intercept gives us the mean log-wage of men, given that all other variables are kept at their reference values, i.e. average age, average length of employment in company, average number of total hours (scaled variable for computational reasons, by dividing the raw values with 365), not being a registered student or a registered apprentice, not being married, without children of any age, not in a labour union, of Icelandic background, not being a supervisor, …, working in a large company, having educ2- education level, working in economic sector A. To this mean, the random effects add the corresponding small variations due to clustering according to each of the three variables seen in the last part of the table (occupation, economic activity, company).

The coefficient of the *gender1*-variable tells us that, if all other variables are kept to their reference levels for both men and women, the difference in their mean log-wages will be -0.146. However, this is not very informative!

The coefficients of *I(age-mean(age))*, …, *econSector* indicate how the log-wages of *men* employees change when those attributes move away from their reference values. For example, men gain 0.033 to their log-wage if they are married, in comparison with men with exactly same other attributes but non-married. Or they add 0.170 to their log-wages if they are supervisors, in comparison with their colleagues with identical characteristics but who are not supervisors.

The coefficients of (interaction) terms like ***gender1::X***, where *X* is *I(age-mean(age))* or *marital1*, …, *econSector*, **indicate how different the effect of these characteristics is for women versus men**. For instance, the effect of being a supervisor (on the log-wage) is smaller for women than for men, with 0.033, therefore men gain 0.170 while women gain only 0.17-0.033=0.167. On the other hand, the interactions between the gender variable and the type of occupation in terms of proportion of women or employees older than 35 are not significant, showing that men and women gain very comparable advantages in these cases.

As we commented elsewhere, the significance of these differences (interactions) and main effects needs to be carefully assessed since multiple comparisons‘ issues may induce false conclusions. A simplistic rule of thumb could be to look at the errors of the estimates and check whether they are much smaller than the estimates but *joint significance tests* are recommended.

Additional remarks

The task of finding the most appropriate models, parsimonious but not over-simplifying and accounting for all structure in the data is a very complex one. We list here several issues that are also involved, in addition to model testing and fitting.

- Preliminary tests and choices:

Heteroskedasticity and heterogeneity tests were performed before building the main models.

Some of the variables may be regarded as both fixed or random effects, e.g. occupation and economic activity. Since the very detailed classifications of these variables was used and their numbers are huge, therefore the number of observations with given values of these two variables is most of the time very small or zero, we treated them as random effects (on intercept but tested for random gender slopes as well).

Some variables were rescaled and centred, in order to provide easier interpretation of results (example: effect of gender, when age is the mean age of individuals, not age zero) and to improve convergence of computations.

- Participation effect or the effect of self-selection.

This is manifested as a censoring effect and it is due to the sampling process which only includes *employed* individuals. If unemployment is in general small and not very different between genders, this effect is rather small and with little impact on averages of distributions, as is our case**.** In [2] it was also proven that, when including into the wage model factors which are also important in predicting participation (*marital status, number and age of children, income of partner* and their interaction with gender), the estimate of the inverse Mills may be shortcut while avoiding un-necessary multicollinearity issues. Our data set does not include information about the partner income but it includes rather detailed information known to have an impact on employment/participation.

- The quality of variables is not uniform and some are better than others, as has been described in [H1]-[H3].
- Models with more interactions and/or random effects on slopes have been examined although results were not included here.
- The *multiple comparisons* issues have been taken into account when testing for significance of effects in various models, by calculating adjusted p-values and confidence intervals (and equivalent measures) rather than univariate p-values.

Results

The main results are summarized in a set of tables which encode the conclusions of multiple level models fitted for testing the significance of various attributes, the interdependence between them or the time evolution of these effects. We also include a set of plots illustrating the uncertainty and estimates of fixed and random effects for relevant multilevel models (Figure 6 – Figure 8).

Table 1 proves that by adding more demographical, work, company and occupation related attributes one does improve the model: more variability is explained, intraclass correlation decreases, AIC decreases and fit improves (confirmed by residuals‘ plots as well).

Table 2 shows that the interaction between gender and many of the characteristics is significant. Therefore the effect of gender depends on their values and equivalently the effect of various characteristics on wages depends on the gender.

Table 3 illustrates the differential effects of characteristics depending on gender and changing with time, by including several models corresponding to several years. We marked red the (interaction) coefficients showing characteristics with a disadvantageous effect for women employees and green the coefficients showing advantages. The main effects are independent on gender.

Table 4 exemplifies how the so-called adjusted gender pay gap decreases with time, using two additive models (M and the more restricted M1), and for five different years.

The decomposition of the gender gap due to modelling, as total versus adjusted, for economic sectors A, R, M, models M and M1, years 2008-2020 as well as several versions of the decomposition into explained and un-explained wage gap for a fixed year and the whole economy, are given in Table 5a and 5b. Figures 9 and 10 complement this dual view of the wage gap.

Table 6 compares the influence of various characteristics on explaining the difference between wages of men and women for two different years, by fitting a set of *independent* models where, for each one, only a set of attributes is used as controls.

Details regarding the data set description and differences between genders' characteristics are given in Table 7 and Table 8.

Uncertainty and significance of effects

Performing Bayesian model averaging and variable selection methods [43], the set of main predictors used by the multilevel models was validated, by using the posterior probability of the coefficients in a BAS model.

The main multilevel models were fitted by both frequentist and Bayesian approaches, as discussed in the introduction, for best evaluation of uncertainty and for efficiency in terms of computing time. We include here several figures which exemplify the model building and the size of uncertainty of estimates. Estimates of model coefficients and standard errors are included in the tables mentioned above.

Figure 6. Maximal model with interaction, for fixed time value (year 2020), generated in the model building process.

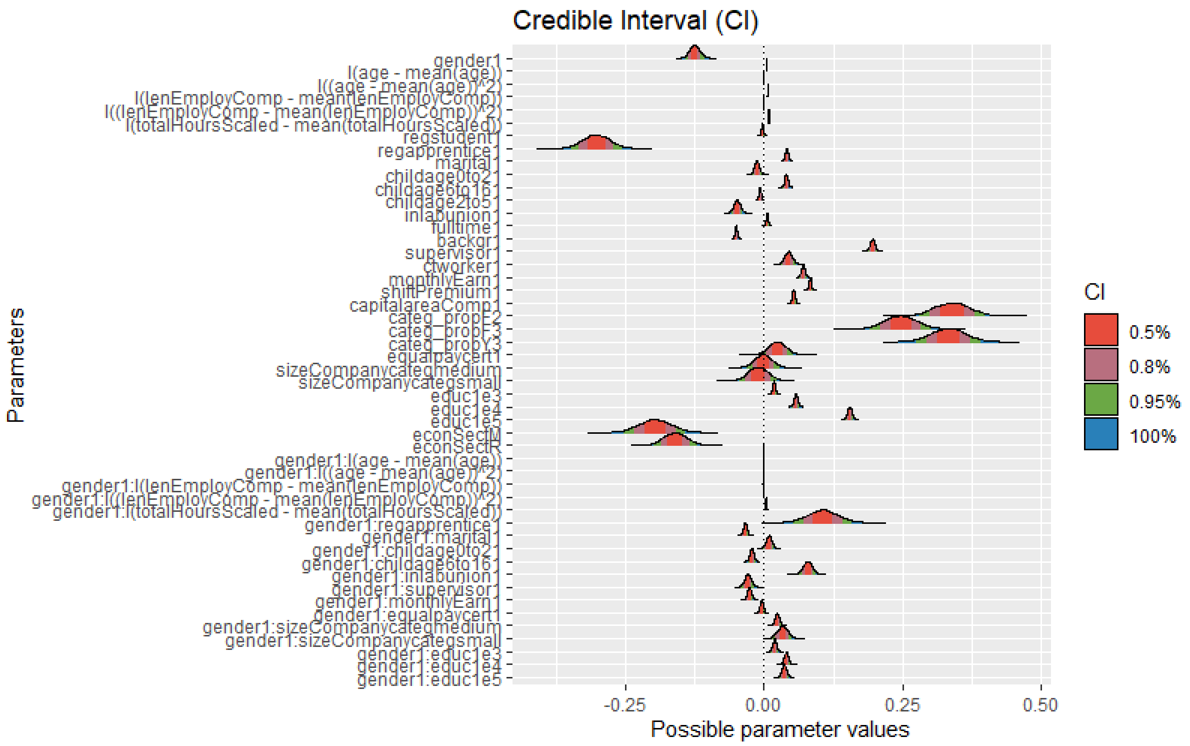


Figure 6 shows why several effects and interactions between gender and covariates, which are tested by fitting a maximal model, are not included in the final model. Credible intervals are practical and rather intuitive ways of assessing uncertainty. The estimate of a given effect for example, lies with 95%

(or other values like 80%, or 50%) probability in its credible interval. Large credible intervals indicate a large uncertainty in estimates.

Figure 7. Fixed effects of the optimum multilevel model with significant interactions (at fixed time, year 2020), showing that different characteristics have different effects depending on gender.

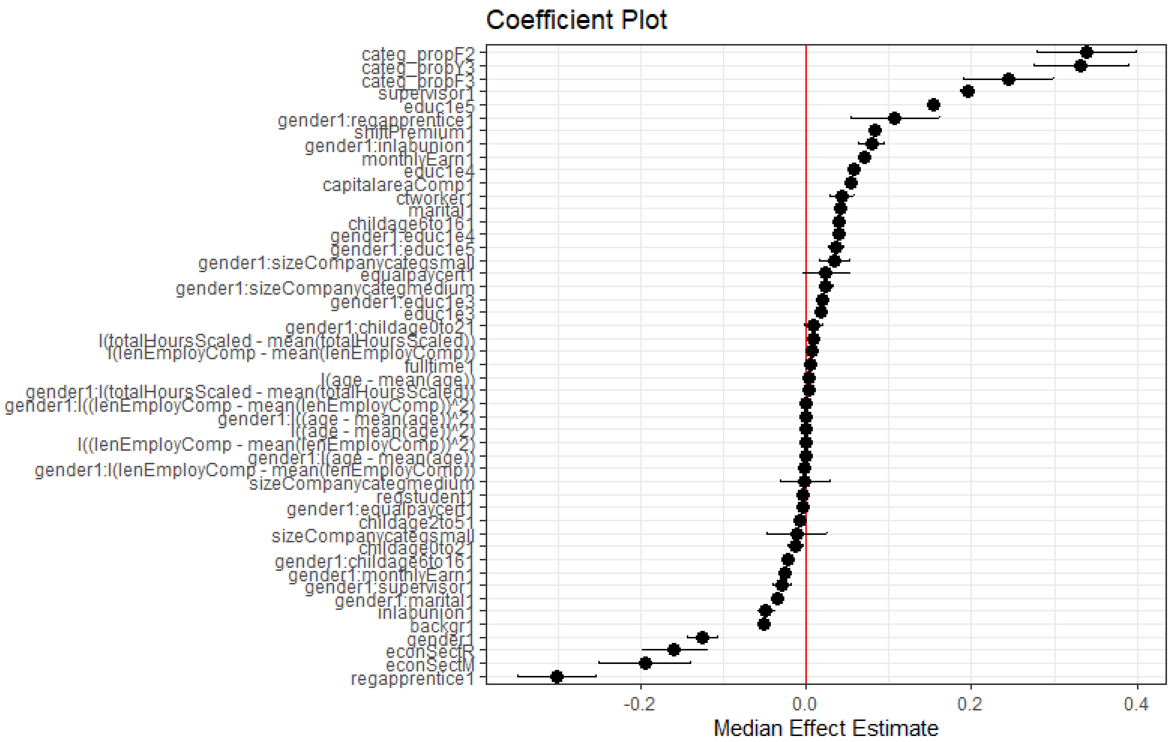
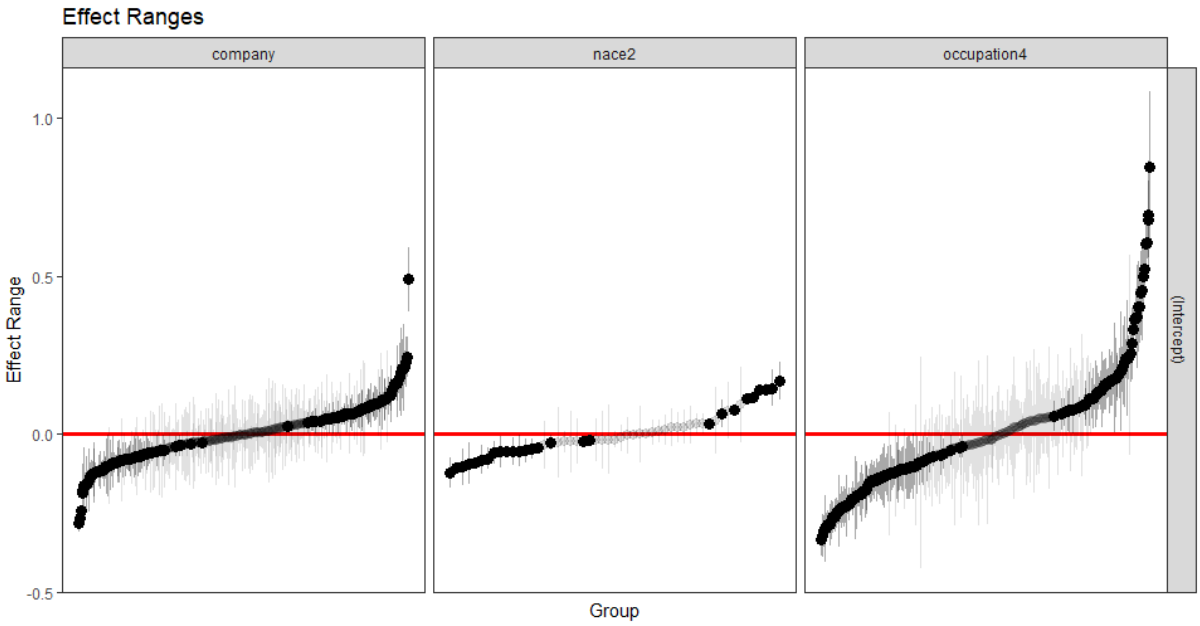


Figure 7 shows the fixed effects estimates and their 95% confidence intervals of a multilevel model (for easy interpretation, at fixed time) fitted using frequentist approach, using restricted maximum likelihood and containing both fixed and random effects.

Figure 8. The influence of grouping by occupation, economic activity and company (at fixed time) on wages (random intercept).



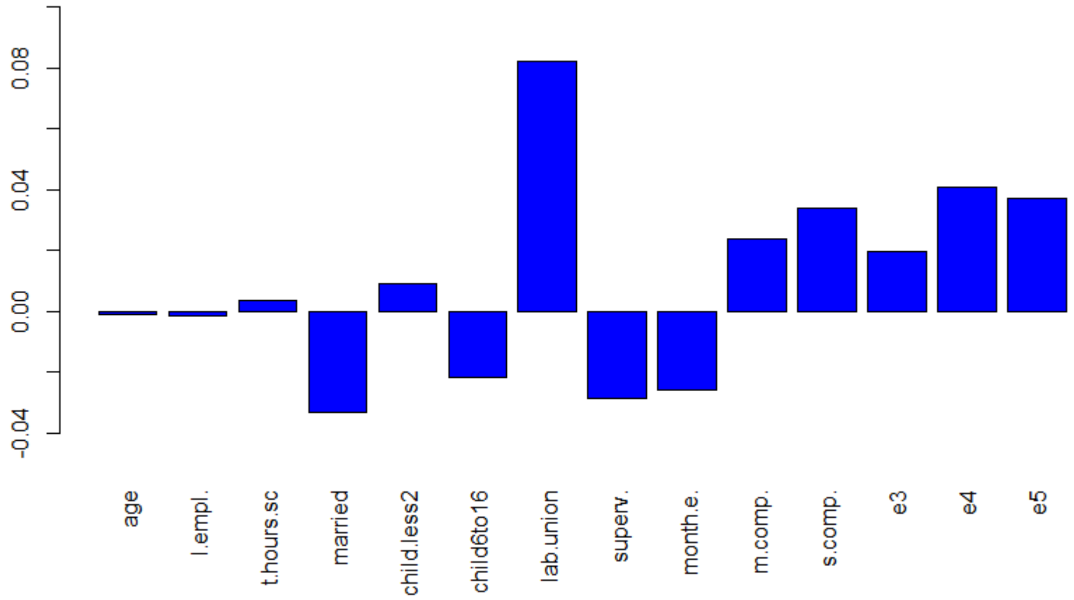
Notes: The effects are ordered by increasing size, for each of these grouping variables (on X-axis: company identifiers, nace2 - two digit codes, occupation – four digit codes).

Figure 8 shows the size of the random effects in the final M* model at fixed time, i.e. the effects of cross-classifying observations by these three criteria. Note that when time growth models are fitted, one more type of random effects is included (individual), due to the (time-) auto-correlation of observations within individuals.

Details of the wage gap decomposition

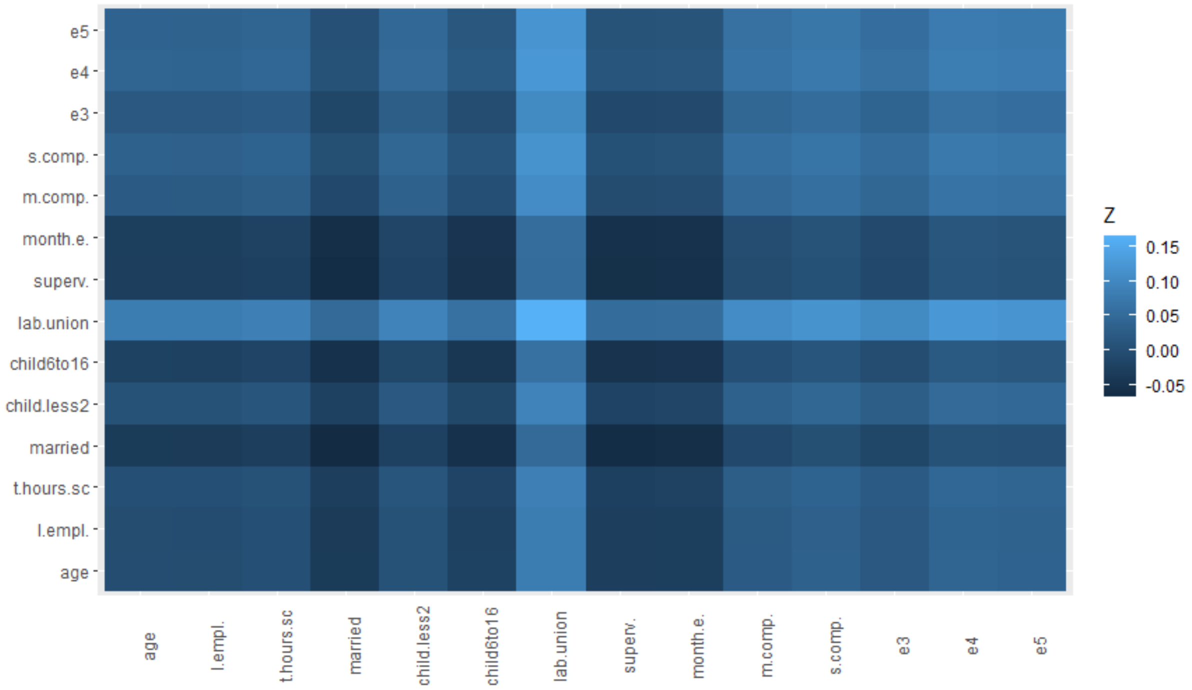
We show several details about adjusting (for covariates) and explaining the total differences between mean wages of women and men. Figure 9 (a and b) exemplifies the advantages/disadvantages of various characteristics on women employees' wages, both as independent and combined effects (of any two of them). This corresponds to the interaction models described above and detailed in Table 2 and 3, i.e. to the fact that wage models are different for men and women. The decomposition illustrated in Figure 10 (a, for model M and b, for model M1) corresponds to the additive models defining the adjusted gap as in Tables 1 and 4. It shows the influence of covariates on wages.

Figure 9a. Representation of differences in covariate effects (except for gender, which is still the largest effect), away from their mean/reference values, as: advantages/disadvantages for women versus men employees.



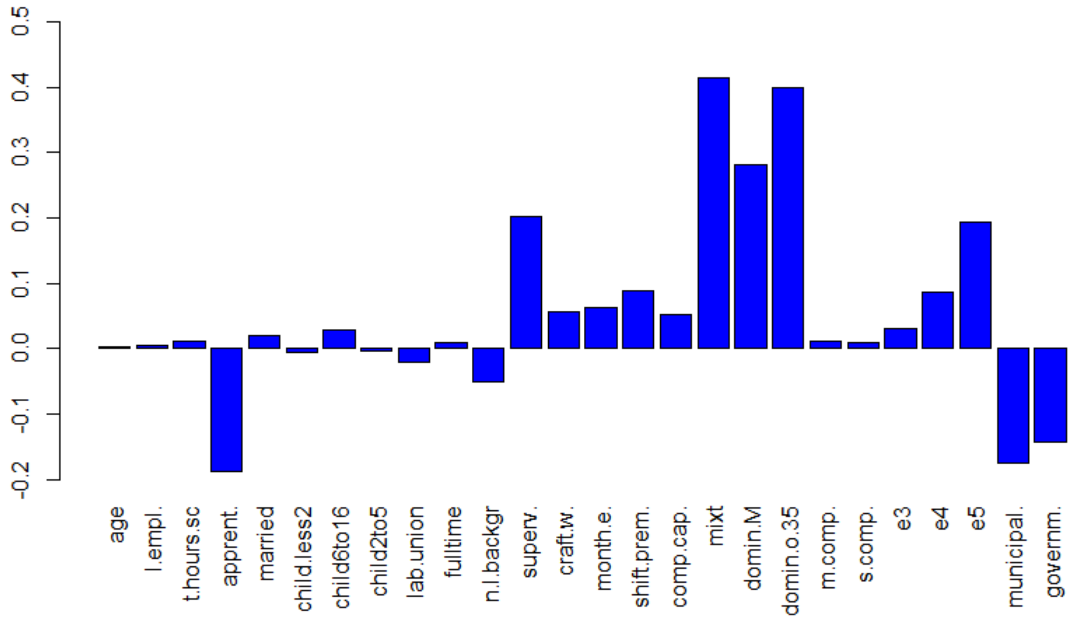
Notes: The following notations are used: age - the (centred) age, l.empl.- length of employment in company, t.hours.sc – total hours worked, scaled, married – marital status value 1, child.less2 – the employee is a parent of a child under two years old, child.6to16 – the employee is a parent of a child between 6 and 16, lab.union – in labout union, superv. – supervisor, month.e – monthly earnings, m.comp. – medium sized company, s.comp. – small company, e3, e4, e5- education levels as described in the main text.

Figure 9b. Representation of the combination of differences in any two covariate effects (except for gender, still the largest effect), away from their mean/reference values, as: advantages/disadvantages for women versus men employees.



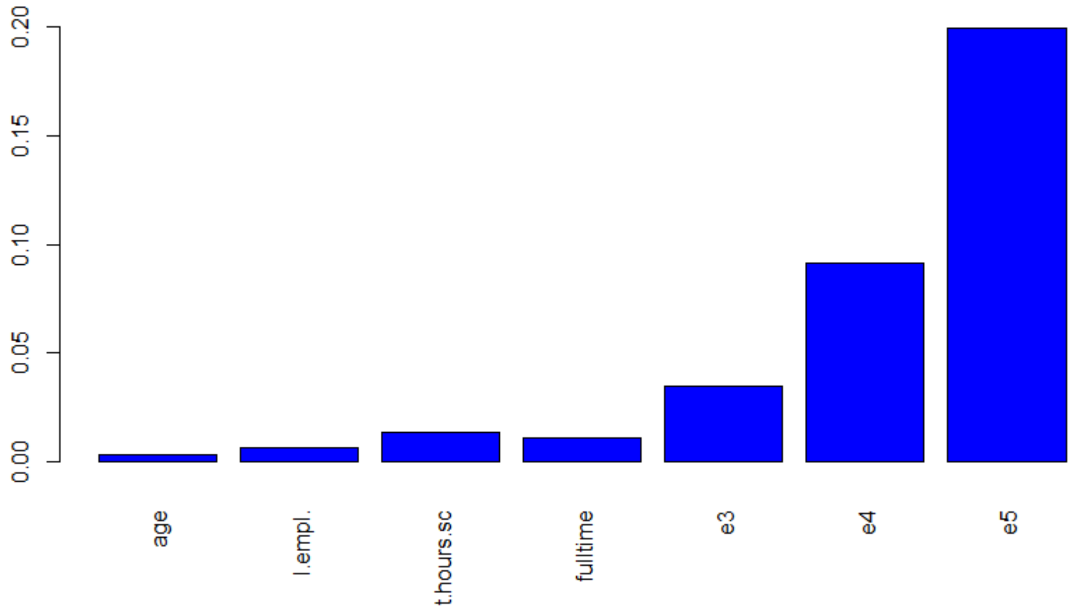
Notes: The X-axis has exactly same values as the Y-axis. The following notations are used: age - the (centred) age, l.empl.- length of employment in company, t.hours.sc – total hours worked, scaled, married – marital status value 1, child.less2 – the employee is a parent of a child under two years old, child.6to16 – the employee is a parent of a child between 6 and 16, lab.union – in labout union, superv. – supervisor, month.e – monthly earnings, m.comp. – medium sized company, s.comp. – small company, e3, e4, e5- education levels as described in the main text.

Figure 10a. Results showing how much various covariates (fixed effects) matter when adjusting the gap, based on an M-type of model.



Notes: The following notations are used: age - the (centred) age, l.empl.- length of employment in company, t.hours.sc – total hours worked scaled, apprent. – apprentice, married – marital status has value 1, child.less2 – the employee is a parent of a child under two years old, child.6to16 – the employee is a parent of a child between 6 and 16, child.2to5 – the employee is a parent of a child between 2 and 5 years old, lab.union – in labout union, fulltime – fulltime, superv. – supervisor, craft.w. – craft worker, month.e – monthly earnings, shift.prem.- shift premium, comp.cap. – company in capital area, mixt – occupation with a balanced mixture of men and women, as described in text, domin.M.- occupation dominated by female employees as defined in text, domin.o.35- occupation dominated by employees older than 35, m.comp. – medium sized company, s.comp. – small company, e3, e4, e5- education levels as described in the main text, municipal. – working for municipalities, governm. – working for government.

Figure 10b. Results showing how much various covariates (fixed effects) matter when adjusting the gap, based on an (more restricted) M1-type of model.



Notes: The following notations are used: age – the (centred) age, l.empl. .- length of employment in company, t.hours.sc – total hours worked scaled, fulltime- fulltime, e3, e4, e5- education levels as described in the main text.

Concluding remarks

We have analysed a large and complex dataset concerning wages in Iceland, between 2008 and 2020. New multilevel models were built in order to understand the influence of various attributes of employees and labour conditions on wages and whether this influence depends or not on gender.

We find of great interest to follow up this study with more detailed analysis. For example, the self-selection aspect is worth investigating more, i.e. testing whether it is indeed accounted for, by including variables with influence on the probability of being employed. Implementation of a time-dependent MLM decomposition of Oaxaca-Blinder type might also be worth pursuing, as well as a more focused study of the interaction between gender and occupation/economic activity.

Another direction of research could be to find the impact of this type of models on the wage index structure and behaviour. This entails a set of new questions and tests which should be developed in the future.

Computing details

The R-code used for the analysis is shared as open code, at: https://github.com/violetacln/GIW

As generated by running "report(SessionInfo(())", we conclude that:

analyses were conducted using the R Statistical language (version 4.0.3; R Core Team, 2020) on Windows 10 x64 (build 17763), using the packages arm (version 1.11.2; Andrew Gelman and Yu-Sung

Su, 2020), DescTools (version 0.99.41; Andri Signorell et mult. al., 2021), furniture (version 1.9.10; Barrett T, Brignone E, 2017), effectsize (version 0.4.4.1; Ben-Shachar M et al., 2020), Rcpp (version 1.0.6; Dirk Eddelbuettel and Romain Francois, 2011), Matrix (version 1.2.18; Douglas Bates and Martin Maechler, 2019), lme4 (version 1.1.26; Douglas Bates et al., 2015), sqldf (version 0.4.11; Grothendieck, 2017), gsubfn (version 0.7; Grothendieck, 2018), proto (version 1.0.0; Gabor Grothendieck, Louis Kates and Thomas Petzoldt, 2016), rstanarm (version 2.21.1; Goodrich B et al., 2020), ggplot2 (version 3.3.3; Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.), stringr (version 1.4.0; Hadley Wickham, 2019), tidyr (version 1.1.2; Hadley Wickham, 2020), forcats (version 0.5.1; Hadley Wickham, 2021), oaxaca (version0.1.4; Marek Hlavac, 2018) readr (version 1.4.0; Hadley Wickham and Jim Hester, 2020), dplyr (version 1.0.3; Hadley Wickham et al., 2021), stargazer (version 5.2.2; Hlavac, Marek, 2018), merTools (version 0.5.2; Jared Knowles and Carl Frederick, 2020), odbc (version 1.3.0; Jim Hester and Hadley Wickham, 2020), car (version 3.0.10; John Fox and Sanford Weisberg, 2019), carData (version 3.0.4; John Fox, Sanford Weisberg and Brad Price, 2020), tibble (version 3.0.6; Kirill Müller and Hadley Wickham, 2021), RSQLite (version 2.2.4; Kirill Müller et al., 2021), lmerTest (version 3.1.3; Kuznetsova A et al., 2017), purrr (version 0.3.4; Lionel Henry and Hadley Wickham, 2020), interactions (version 1.1.3; Long JA, 2019), sjmisc (version 2.8.6; Lüdecke D, 2018), parameters (version 0.13.0; Lüdecke D et al., 2020), insight (version 0.13.2; Lüdecke D et al., 2019), performance (version 0.7.1; Lüdecke et al., 2021), bayestestR (version 0.9.0; Makowski et al., 2019), report (version 0.3.0; Makowski et al., 2020), DBI (version 1.1.1; R Special Interest Group on Databases, R-SIG-DB), BAS (version 1.5.5; Merlise Clyde, 2020) , margins (version 0.3.26; Thomas Leeper, 2021), MASS (version 7.3.53.1; Venables et al., 2002) and tidyverse (version 1.3.0; Wickham et al., 2019).

References

[H1] Reports in the Statistical Series of Statistics Iceland (2010), https://hagstofa.is/media/43074/launamunur_kynjanna_lokaskyrsla.pdf and

https://hagstofa.is/media/43558/hag_100218.pdf

[H2] Sigurður Snævarr og velferðarráðuneytið (2015) Launamunur karla og kvenna, https://www.stjornarradid.is/media/velferdarraduneyti-media/media/rit-og-skyrslur-2015/Launamunur_karla_og_kvenna_19052015b.pdf

[H3] Analysis on Gender Pay Gap 2008–2016 (2018) Statistical series of Statistics Iceland, https://hagstofan.s3.amazonaws.com/media/public/4a70b304-09ff-4c6b-adcf-050ac2bef384/pub_doc_mqp71tr.pdf

[1] Meara, K., Pastore, F. & Webster, A. The gender pay gap in the USA: a matching study. *J Popul Econ* **33,** 271–305 (2020). https://doi.org/10.1007/s00148-019-00743-8.

[2] Development of econometric methods to evaluate the Gender pay gap using Structure of Earnings Survey data (2009) Luxembourg: Office for Official Publications of the European Communities.

[3] Jennifer Hill and Andrew Gelman (2012) Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge

[4] Andrew Gelman and Yu-Sung Su (2020). Arm: Data Analysis Using Regression and Multilevel/Hierarchical Models. R package version 1.11-2. https://CRAN.R-project.org/package=arm

[5] Andri Signorell et mult. Al. (2021). DescTools: Tools for descriptive statistics. R package version 0.99.41.

[6] Barrett T, Brignone E (2017). "Furniture for Quantitative Scientists." _The R Journal_, *9*(2),142-148. Doi: 10.32614/RJ-2017-037 (URL: https://doi.org/10.32614/RJ-2017-037), <URL:https://journal.r-project.org/archive/2017/RJ-2017-037/RJ-2017-037.pdf>.

[7] Ben-Shachar M, Lüdecke D, Makowski D (2020). Effectsize: Estimation of Effect Size Indices and Standardized Parameters. Journal of Open Source Software, 5(56), 2815. Doi: 10.21105/joss.02815

[8] Dirk Eddelbuettel and Romain Francois (2011). Rcpp: Seamless R and C++ Integration. Journal of Statistical Software, 40(8), 1-18. URL https://www.jstatsoft.org/v40/i08/.

[9] Douglas Bates and Martin Maechler (2019). Matrix: Sparse and Dense Matrix Classes and Methods. R package version 1.2-18. https://CRAN.R-project.org/package=Matrix

[10] Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. Doi:10.18637/jss.v067.i01.

[11] G. Grothendieck (2017). Sqldf: Manipulate R Data Frames Using SQL. R package version 0.4-11. https://CRAN.R-project.org/package=sqldf

[12] G. Grothendieck (2018). Gsubfn: Utilities for Strings and Function Arguments. R package version 0.7. https://CRAN.R-project.org/package=gsubfn

[13] Gabor Grothendieck, Louis Kates and Thomas Petzoldt (2016). Proto: Prototype Object-Based Programming. R package version 1.0.0. https://CRAN.R-project.org/package=proto

[14] Goodrich B, Gabry J, Ali I & Brilleman S. (2020). Rstanarm: Bayesian applied regression modeling via Stan. R package version 2.21.1 https://mc-stan.org/rstanarm.

[15] H. Wickham. Ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

[16] Hadley Wickham (2019). Stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0. https://CRAN.R-project.org/package=stringr

[17] Hadley Wickham (2020). Tidyr: Tidy Messy Data. R package version 1.1.2. https://CRAN.R-project.org/package=tidyr

[18] Hadley Wickham (2021). Forcats: Tools for Working with Categorical Variables (Factors). R package version 0.5.1. https://CRAN.R-project.org/package=forcats

[19] Hadley Wickham and Jim Hester (2020). Readr: Read Rectangular Text Data. R package version 1.4.0. https://CRAN.R-project.org/package=readr

[20] Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2021). Dplyr: A Grammar of Data Manipulation. R package version 1.0.3. https://CRAN.R-project.org/package=dplyr

[21] Hlavac, Marek (2018). Stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.1. https://CRAN.R-project.org/package=stargazer

[22] Jared E. Knowles and Carl Frederick (2020). merTools: Tools for Analyzing Mixed Effect Regression Models. R package version 0.5.2. https://CRAN.R-project.org/package=merTools

[23] Jim Hester and Hadley Wickham (2020). Odbc: Connect to ODBC Compatible Databases (using the DBI Interface). R package version 1.3.0. https://CRAN.R-project.org/package=odbc

[24] John Fox and Sanford Weisberg (2019). An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

[25] John Fox, Sanford Weisberg and Brad Price (2020). carData: Companion to Applied Regression Data Sets. R package version 3.0-4. https://CRAN.R-project.org/package=carData

[26] Kirill Müller and Hadley Wickham (2021). Tibble: Simple Data Frames. R package version 3.0.6. https://CRAN.R-project.org/package=tibble

[27] Kirill Müller, Hadley Wickham, David A. James and Seth Falcon (2021). RSQLite: ‚SQLite' Interface for R. R package version 2.2.4. https://CRAN.R-project.org/package=RSQLite

[28] Kuznetsova A, Brockhoff PB, Christensen RHB (2017). "lmerTest Package: Tests in Linear MixedEffects Models." _Journal of Statistical Software_, *82*(13), 1-26. Doi: 10.18637/jss.v082.i13(URL: https://doi.org/10.18637/jss.v082.i13).

[29] Lionel Henry and Hadley Wickham (2020). Purrr: Functional Programming Tools. R package version 0.3.4. https://CRAN.R-project.org/package=purrr

[30] Long JA (2019). _interactions: Comprehensive, User-Friendly Toolkit for Probing Interactions_. Rpackage version 1.1.0, <URL: https://cran.r-project.org/package=interactions>.

[31] Lüdecke D (2018). “sjmisc: Data and Variable Transformation Functions.” _Journal of Open SourceSoftware_, *3*(26), 754. Doi: 10.21105/joss.00754 (URL: https://doi.org/10.21105/joss.00754).

[32] Lüdecke D, Ben-Shachar M, Patil I, Makowski D (2020). “parameters: Extracting, Computing andExploring the Parameters of Statistical Models using R.” _Journal of Open Source Software_,*5*(53), 2445. Doi: 10.21105/joss.02445 (URL: https://doi.org/10.21105/joss.02445).

[33] Lüdecke D, Waggoner P, Makowski D (2019). “insight: A Unified Interface to Access Information fromModel Objects in R.” _Journal of Open Source Software_, *4*(38), 1412. Doi: 10.21105/joss.01412(URL: https://doi.org/10.21105/joss.01412).

[34] Lüdecke et al., (2021). Assessment, Testing and Comparison of Statistical Models using R. Journal of Open Source Software, 6(59), 3112. https://doi.org/10.31234/osf.io/vtq8f

[35] Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. Journal of Open Source Software, 4(40), 1541. Doi:10.21105/joss.01541

[36] Makowski, D., Ben-Shachar, M.S., Patil, I. & Lüdecke, D. (2020). Automated Results Reporting as a Practical Tool to Improve Reproducibility and Methodological Best Practices Adoption. CRAN. Available from https://github.com/easystats/report. Doi: .

[37] R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[38] R Special Interest Group on Databases (R-SIG-DB), Hadley Wickham and Kirill Müller (2021). DBI: R Database Interface. R package version 1.1.1. https://CRAN.R-project.org/package=DBI

[39] Thomas J. Leeper (2021). Margins: Marginal Effects for Model Objects. R package version 0.3.26.

[40] Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

[41] Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

[42] Hlavac, Marek (2018). oaxaca: Blinder-Oaxaca Decomposition in R. R package version 0.1.4. https://CRAN.R-project.org/package=oaxaca

[43] Clyde M (2020). *BAS: Bayesian Variable Selection and Model Averaging using Bayesian Adaptive Sampling*. R package version 1.5.5.

[44] Bunel, M, Guironnet, J-P (2017) Income inequalities for recently graduated French workers: a mu ltilevel modelling approach, Empir Econ (2017) 53:755-778.

[45] Correll, S. (2001). Gender and the Career Choice Process: The Role of Biased Self-Assessments. *Americ an Journal of Sociology, 106*(6), 1691-1730. doi:10.1086/321299.

Table 1. A set of models, at fixed time (year=2019) showing that, by adding more demographical, work, company and occupation related attributes/structure, one does improve the model: more variability is explained, intraclass correlation decreases, AIC decreases and fit improves (confirmed by residuals' plots as well). Note that models M1a and M1 differ only by (not-) including the effect of clustering of observations by company (i.e. of correlation of wages inside each company). This has the undesired effect of exaggerating the significance of the fixed effects and introducing a small bias in the estimates of M1a, that is why one should prefer M1 if using a restricted set of covariates.

	M1a		**M1**		**M2**		**M3**		**M**	
Predictors	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*
(Intercept)	8.053	0.026	8.038	0.023	8.016	0.022	7.967	0.0232	7.570	0.041
gender1	**-0.045**	**0.002**	**-0.047**	**0.001**	**-0.048**	**0.001**	**-0.045**	**0.001**	**-0.044**	**0.001**
I(age - mean(age))	0.004	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003	0.000
I((age - mean(age))^2)	-0.000	0.000	-0.000	0.000	-0.000	0.000	-0.000	0.000	-0.000	0.000
I(lenEmployComp - mean(lenEmployComp))	0.006	0.000	0.007	0.000	0.006	0.000	0.005	0.000	0.005	0.000
I((lenEmployComp - mean(lenEmployComp))^2)	-0.000	0.000	-0.000	0.000	-0.000	0.000	-0.000	0.000	-0.000	0.000
I(totalHoursScaled - mean(totalHoursScaled))	0.013	0.000	0.013	0.000	0.013	0.000	0.010	0.000	0.010	0.000
fulltime1	0.012	0.002	0.011	0.002	0.014	0.002	0.008	0.002	0.008	0.002
educ1e3	0.035	0.002	0.034	0.002	0.036	0.002	0.032	0.002	0.031	0.002
educ1e4	0.088	0.002	0.088	0.002	0.091	0.002	0.085	0.002	0.083	0.002
educ1e5	0.181	0.003	0.182	0.003	0.185	0.003	0.180	0.003	0.177	0.003
marital1					0.021	0.002	0.020	0.002	0.020	0.002
childage0to2_1					-0.003	0.003	-0.005	0.003	-0.005	0.003
childage6to16_1					0.027	0.002	0.028	0.002	0.028	0.002
childage2to5_1					-0.007	0.002	-0.005	0.002	-0.004	0.002
backgr1					-0.052	0.002	-0.049	0.002	-0.051	0.002
regstudent1							-0.001	0.002	-0.002	0.002
regapprentice1							-0.206	0.011	-0.209	0.011
inlabunion1							-0.022	0.004	-0.020	0.004
supervisor1							0.184	0.003	0.184	0.003
ctworker1							0.049	0.005	0.055	0.007
monthlyEarn1							0.061	0.003	0.060	0.003
shiftPremium1							0.083	0.002	0.084	0.002

Table 1. *Cont.*

	M1a		M1		M2		M3		M	
Predictors	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*
capitalareaComp1									0.052	0.003
categ_propF2									0.347	0.031
categ_propF3									0.248	0.028
categ_propY3									0.337	0.030
equalpaycert1									0.023	0.015
sizeCompanycategmedium									0.011	0.015
sizeCompanycategsmall									0.008	0.017
econSectM									-0.193	0.028
econSectR									-0.156	0.019
Random Effects										
σ^2	0.03		0.03		0.03		0.03		0.03	
τ_{00}	0.06 $_{occupation4}$		0.01 $_{company}$		0.01 $_{company}$		0.01 $_{company}$		0.01 $_{company}$	
	0.02 $_{nace2}$		0.06 $_{occupation4}$		0.05 $_{occupation4}$		0.06 $_{occupation4}$		0.03 $_{occupation4}$	
			0.01 $_{nace2}$		0.01 $_{nace2}$		0.01 $_{nace2}$		0.01 $_{nace2}$	
ICC	0.71		0.72		0.71		0.74		0.63	
N	53 $_{nace2}$		319 $_{company}$		319 $_{company}$		319 $_{company}$		319 $_{company}$	
	273 $_{occupation4}$		53 $_{nace2}$		53 $_{nace2}$		53 $_{nace2}$		53 $_{nace2}$	
			273 $_{occupation4}$		273 $_{occupation4}$		273 $_{occupation4}$		273 $_{occupation4}$	
Observations	86496		86496		86496		86496		86496	
Marginal R^2 / Conditional R^2	0.121 / 0.747		0.128 / 0.754		0.149 / 0.755		0.161 / 0.778		0.406 / 0.782	
AIC	-46.378.840		-52.352.846		-53.629.817		-59.799.560		-60.401.651	

Table 2. A set of models (additive: M-model, separate models for women: w-model and men: m-model, interactive: M*-model) at fixed time (year 2019) showing that the interaction between gender and many characteristics is significant therefore the effect of gender depends on their values and equivalently that the effect of various characteristics depends on the gender.

	Additive (M)		**f**		**m**		**Interaction (M*)**	
Predictors	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*
(Intercept)	7.570	0.041	7.475	0.051	7.623	0.044	7.617	0.041
gender1	-0.044	0.001					-0.125	0.009
I(age - mean(age))	0.003	0.000	0.003	0.000	0.004	0.000	0.004	0.000
I((age - mean(age))^2)	-0.000	0.000	0.000	0.000	-0.000	0.000	-0.000	0.000
I(lenEmployComp - mean(lenEmployComp))	0.005	0.000	0.004	0.000	0.007	0.000	0.006	0.000
I((lenEmployComp - mean(lenEmployComp))^2)	-0.000	0.000	-0.000	0.000	-0.000	0.000	-0.000	0.000
I(totalHoursScaled - mean(totalHoursScaled))	0.010	0.000	0.012	0.001	0.008	0.001	0.008	0.001
regstudent1	-0.002	0.002	-0.002	0.002	-0.006	0.003	-0.003	0.002
regapprentice1	-0.209	0.011	-0.199	0.011	-0.343	0.027	-0.303	0.024
marital1	0.020	0.002	0.008	0.002	0.039	0.003	0.041	0.003
childage0to2_1	-0.005	0.003	-0.002	0.003	-0.013	0.004	-0.013	0.004
childage6to16_1	0.028	0.002	0.017	0.002	0.039	0.003	0.040	0.003
childage2to5_1	-0.004	0.002	-0.009	0.002	-0.003	0.004	-0.007	0.002
inlabunion1	-0.020	0.004	0.042	0.007	-0.048	0.006	-0.048	0.005
fulltime1	0.008	0.002	0.001	0.002	0.013	0.003	0.005	0.002
backgr1	-0.051	0.002	-0.048	0.002	-0.053	0.003	-0.050	0.002
supervisor1	0.184	0.003	0.166	0.004	0.193	0.005	0.196	0.004
ctworker1	0.055	0.007	0.070	0.016	0.027	0.009	0.044	0.007
monthlyEarn1	0.060	0.003	0.036	0.003	0.079	0.004	0.071	0.003
shiftPremium1	0.084	0.002	0.089	0.002	0.073	0.004	0.083	0.002
capitalareaComp1	0.052	0.003	0.053	0.003	0.051	0.004	0.053	0.003
categ_propF2	0.347	0.031	0.359	0.045	0.333	0.033	0.339	0.031
categ_propF3	0.248	0.028	0.238	0.042	0.215	0.030	0.246	0.028
categ_propY3	0.337	0.030	0.387	0.030	0.343	0.032	0.332	0.030
equalpaycert1	0.023	0.015	0.003	0.013	0.028	0.017	0.024	0.015
sizeCompanycategmedium	0.011	0.015	0.021	0.014	0.007	0.017	-0.001	0.015
sizeCompanycategsmall	0.008	0.018	0.026	0.017	-0.008	0.021	-0.010	0.018
educ1e3	0.031	0.002	0.035	0.002	0.019	0.003	0.018	0.002
educ1e4	0.083	0.002	0.092	0.003	0.062	0.004	0.058	0.003
educ1e5	0.177	0.003	0.185	0.003	0.154	0.004	0.154	0.004

Table 2. *Cont.*

	Additive (M)		f		m		Interaction (M*)	
Predictors	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*
econSectM	-0.193	0.028	-0.207	0.028	-0.199	0.032	-0.196	0.028
econSectR	-0.155	0.020	-0.159	0.022	-0.200	0.023	-0.160	0.020
gender1:I(age - mean(age))							-0.001	0.000
gender1:I((age - mean(age))^2)							0.000	0.000
gender1:I(lenEmployComp - mean(lenEmployComp))							-0.002	0.000
gender1:I((lenEmployComp - mean(lenEmployComp))^2)							0.000	0.000
gender1:I(totalHoursScaled - mean(totalHoursScaled))							0.003	0.001
gender1:regapprentice1							0.106	0.026
gender1:marital1							-0.034	0.003
gender1:childage0to2_1							0.009	0.005
gender1:childage6to16_1							-0.022	0.003
gender1:inlabunion1							0.079	0.008
gender1:supervisor1							-0.029	0.006
gender1:monthlyEarn1							-0.026	0.004
gender1:equalpaycert1							-0.004	0.003
gender1:sizeCompanycategmedium							0.024	0.004
gender1:sizeCompanycategsmall							0.034	0.009
gender1:educ1e3							0.019	0.003
gender1:educ1e4							0.040	0.004
gender1:educ1e5							0.036	0.004
Random Effects								
σ^2	0.03		0.02		0.03		0.03	
τ_{00}	0.01 $_{company}$		0.01 $_{company}$		0.01 $_{company}$		0.01 $_{company}$	
	0.03 $_{occupation4}$		0.03 $_{occupation4}$		0.04 $_{occupation4}$		0.03 $_{occupation4}$	
	0.01 $_{nace2}$		0.01 $_{nace2}$		0.01 $_{nace2}$		0.01 $_{nace2}$	
ICC	0.64		0.65		0.61		0.63	
N	319 $_{company}$		309 $_{company}$		309 $_{company}$		319 $_{company}$	
	53 $_{nace2}$		53 $_{nace2}$		53 $_{nace2}$		53 $_{nace2}$	
	273 $_{occupation4}$		241 $_{occupation4}$		263 $_{occupation4}$		273 $_{occupation4}$	
Observations	86496		49644		36852		86496	
Marginal R^2 / Conditional R^2	0.404 / 0.783		0.449 / 0.805		0.412 / 0.771		0.412 / 0.785	
AIC	-60.067.118		-44.105.407		-18.643.218		-61.007.042	

Table 3. A set of identical models (with significant interaction between gender and other attributes) fitted for several years (2016-2020), showing the differential effects of characteristics depending on gender and changing with time. We marked red the (interaction) coefficients showing characteristics with a disadvantageous effect for women employees and green the coefficients showing advantages. The main effects are of course independent on gender.

	2016		**2017**		**2018**		**2019**		**2020**	
Predictors	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*
(Intercept)	7.492	0.040	7.464	0.041	7.529	0.040	7.621	0.041	7.641	0.042
gender1	-0.154	0.009	-0.141	0.009	-0.134	0.009	-0.136	0.009	-0.146	0.009
I(age – mean(age))	0.004	0.000	0.004	0.000	0.004	0.000	0.004	0.000	0.004	0.000
I((age – mean(age))^2)	-0.000	0.000	-0.000	0.000	-0.000	0.000	-0.000	0.000	-0.000	0.000
I(lenEmployComp – mean(lenEmployComp))	0.007	0.000	0.007	0.000	0.007	0.000	0.007	0.000	0.005	0.000
I((lenEmployComp – mean(lenEmployComp))^2)	-0.000	0.000	-0.000	0.000	-0.000	0.000	-0.000	0.000	-0.000	0.000
I(totalHoursScaled – mean(totalHoursScaled))	0.007	0.001	0.008	0.001	0.007	0.001	0.008	0.001	0.009	0.001
regstudent1	-0.002	0.002	-0.004	0.002	-0.005	0.002	-0.004	0.002	0.000	0.002
regapprentice1	-0.280	0.047	-0.299	0.030	-0.284	0.024	-0.307	0.024	-0.245	0.026
marital1	0.039	0.003	0.040	0.003	0.042	0.003	0.040	0.003	0.033	0.003
childage0to2_1	-0.021	0.004	-0.022	0.004	-0.015	0.004	-0.013	0.004	-0.020	0.004
childage6to16_1	0.040	0.003	0.044	0.003	0.043	0.003	0.040	0.003	0.044	0.003
childage2to5_1	-0.009	0.002	-0.008	0.002	-0.009	0.002	-0.007	0.002	-0.012	0.002
inlabunion1	-0.027	0.006	-0.030	0.006	-0.035	0.006	-0.047	0.005	-0.055	0.005
fulltime1	0.007	0.002	0.008	0.002	0.007	0.002	0.005	0.002	-0.002	0.002
backgr1	-0.041	0.002	-0.044	0.002	-0.047	0.002	-0.051	0.002	-0.049	0.002
supervisor1	0.182	0.005	0.193	0.004	0.187	0.004	0.193	0.004	0.179	0.004
ctworker1	0.012	0.008	0.028	0.007	0.057	0.007	0.040	0.007	0.033	0.007
monthlyEarn1	0.087	0.004	0.085	0.003	0.084	0.003	0.076	0.003	0.061	0.003
shiftPremium1	0.084	0.002	0.086	0.002	0.089	0.002	0.083	0.002	0.076	0.002
capitalareaComp1	0.049	0.003	0.041	0.003	0.045	0.003	0.053	0.003	0.065	0.003
categ_propF2	0.213	0.024	0.283	0.025	0.358	0.030	0.336	0.031	0.390	0.031
categ_propF3	0.277	0.031	0.331	0.030	0.259	0.027	0.245	0.028	0.244	0.027
categ_propY3	0.260	0.023	0.300	0.023	0.351	0.029	0.330	0.030	0.382	0.030
equalpaycert1							0.021	0.015	0.017	0.014
sizeCompanycategmedium	-0.027	0.016	-0.002	0.015	0.011	0.015	-0.001	0.015	0.006	0.014
sizeCompanycategsmall	-0.017	0.019	-0.026	0.018	-0.007	0.018	-0.011	0.018	0.008	0.019

Table 3. *Cont.*

	2016		2017		2018		2019		2020	
Predictors	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*
educ1e3	0.014	0.003	0.014	0.003	0.019	0.002	0.020	0.002	0.019	0.002
educ1e4	0.067	0.003	0.065	0.003	0.063	0.003	0.061	0.003	0.061	0.003
educ1e5	0.179	0.004	0.170	0.004	0.165	0.004	0.161	0.004	0.154	0.004
econSectM	-0.176	0.033	-0.218	0.031	-0.221	0.030	-0.212	0.028	-0.186	0.028
econSectR	-0.137	0.021	-0.177	0.020	-0.163	0.020	-0.172	0.020	-0.181	0.020
gender1:I(age – mean(age))	-0.001	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000	-0.001	0.000
gender1:I((age – mean(age))^2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
gender1:I(lenEmployComp – mean(lenEmployComp))	-0.002	0.000	-0.002	0.000	-0.002	0.000	-0.002	0.000	-0.001	0.000
gender1:I((lenEmployComp – mean(lenEmployComp))^2)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
gender1:I(totalHoursScaled – mean(totalHoursScaled))	0.002	0.001	0.002	0.001	0.003	0.001	0.004	0.001	0.003	0.001
gender1:regapprentice1	0.130	0.049	0.138	0.032	0.103	0.026	0.109	0.026	0.088	0.028
gender1:marital1	-0.037	0.003	-0.037	0.003	-0.034	0.003	-0.033	0.003	-0.029	0.003
gender1:childage0to2_1	0.018	0.005	0.016	0.005	0.015	0.005	0.010	0.005	0.013	0.005
gender1:childage6to16_1	-0.019	0.004	-0.025	0.004	-0.025	0.003	-0.022	0.003	-0.024	0.003
gender1:inlabunion1	0.106	0.008	0.096	0.008	0.081	0.008	0.083	0.008	0.100	0.008
gender1:supervisor1	-0.031	0.006	-0.039	0.006	-0.019	0.006	-0.025	0.006	-0.033	0.006
gender1:monthlyEarn1	-0.048	0.004	-0.045	0.004	-0.034	0.004	-0.038	0.004	-0.038	0.004
gender1:equalpaycert1							0.001	0.003	-0.001	0.004
gender1:sizeCompanycategmedium	0.024	0.005	0.024	0.004	0.018	0.004	0.022	0.004	0.018	0.004
gender1:sizeCompanycategsmall	0.025	0.011	0.038	0.011	0.034	0.010	0.033	0.009	0.033	0.010
gender1:educ1e3	0.022	0.004	0.020	0.003	0.016	0.003	0.016	0.003	0.016	0.003
gender1:educ1e4	0.033	0.004	0.029	0.004	0.032	0.004	0.033	0.004	0.029	0.004
gender1:educ1e5	0.017	0.005	0.017	0.005	0.020	0.005	0.026	0.004	0.026	0.004
gender1:econSectM	0.045	0.004	0.046	0.004	0.039	0.004	0.038	0.004	0.036	0.003
gender1:econSectR	0.040	0.004	0.039	0.004	0.031	0.004	0.035	0.004	0.034	0.004

Table 3. *Cont.*

	2016		2017		2018		2019		2020	
Predictors	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*
Random Effects										
σ^2	0.03		0.03		0.03		0.03		0.03	
τ_{00}	0.01 company		0.01 company		0.01 company		0.01 company		0.01 company	
	0.03 occupation4		0.03 occupation4		0.03 occupation4		0.03 occupation4		0.03 occupation4	
	0.01 nace2		0.01 nace2		0.01 nace2		0.01 nace2		0.01 nace2	
ICC	0.65		0.62		0.62		0.63		0.62	
N	279 company		298 company		306 company		319 company		311 company	
	52 nace2		52 nace2		52 nace2		53 nace2		53 nace2	
	274 occupation4		275 occupation4		270 occupation4		273 occupation4		267 occupation4	
Observations	75534		79807		83370		86496		83597	
Marginal R^2 / Conditional R^2	0.427 / 0.798		0.462 / 0.797		0.426 / 0.785		0.412 / 0.785		0.438 / 0.786	
AIC	-51.710.150		-53.796.925		-58.098.358		-61.129.053		-63.713.281	

Table 4a. A set of identical, additive-gender M-models (as defined in Table 1), fitted for several time values (years 2016-2020) and showing how the so-called adjusted gender pay gap decreases with time.

	2016		2017		2018		2019		2020	
Predictors	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*
(Intercept)	7.447	0.041	7.423	0.041	7.482	0.040	7.570	0.041	7.586	0.043
gender1	-0.051	0.002	-0.050	0.002	-0.049	0.001	-0.044	0.001	-0.042	0.001
I(age – mean(age))	0.003	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003	0.000
I((age – mean(age))^2)	-0.000	0.000	-0.000	0.000	-0.000	0.000	-0.000	0.000	-0.000	0.000
I(lenEmployComp – mean(lenEmployComp))	0.005	0.000	0.006	0.000	0.006	0.000	0.005	0.000	0.004	0.000
I((lenEmployComp – mean(lenEmployComp))^2)	-0.000	0.000	-0.000	0.000	-0.000	0.000	-0.000	0.000	-0.000	0.000
I(totalHoursScaled – mean(totalHoursScaled))	0.009	0.000	0.009	0.000	0.008	0.000	0.010	0.000	0.011	0.000
regstudent1	-0.000	0.002	-0.002	0.002	-0.003	0.002	-0.002	0.002	0.002	0.002
regapprentice1	-0.140	0.011	-0.164	0.011	-0.187	0.011	-0.209	0.011	-0.158	0.010
marital1	0.017	0.002	0.018	0.002	0.021	0.002	0.020	0.002	0.016	0.001
childage0to2_1	-0.006	0.003	-0.009	0.003	-0.003	0.003	-0.005	0.003	-0.009	0.002
childage6to16_1	0.030	0.002	0.030	0.002	0.028	0.002	0.028	0.002	0.030	0.002
childage2to5_1	-0.005	0.002	-0.005	0.002	-0.006	0.002	-0.004	0.002	-0.009	0.002
inlabunion1	0.006	0.005	-0.002	0.005	-0.009	0.005	-0.020	0.004	-0.020	0.004
fulltime1	0.010	0.002	0.010	0.002	0.009	0.002	0.008	0.002	0.001	0.002
backgr1	-0.041	0.002	-0.045	0.002	-0.048	0.002	-0.051	0.002	-0.049	0.002
supervisor1	0.170	0.003	0.177	0.003	0.181	0.003	0.184	0.003	0.165	0.003
ctworker1	0.025	0.008	0.041	0.007	0.072	0.007	0.055	0.007	0.048	0.007
monthlyEarn1	0.067	0.003	0.068	0.003	0.070	0.003	0.060	0.003	0.044	0.002
shiftPremium1	0.085	0.002	0.087	0.002	0.090	0.002	0.084	0.002	0.077	0.002
capitalareaComp1	0.049	0.003	0.040	0.003	0.044	0.003	0.052	0.003	0.064	0.003
categ_propF2	0.219	0.025	0.289	0.025	0.369	0.031	0.347	0.031	0.402	0.032
categ_propF3	0.279	0.031	0.334	0.030	0.263	0.028	0.248	0.028	0.249	0.027
categ_propY3	0.262	0.023	0.304	0.023	0.358	0.029	0.337	0.030	0.390	0.030
equalpaycert1	_	_	_	_	_	_	0.023	0.015	0.016	0.014
sizeCompanycategmedium	-0.016	0.016	0.010	0.015	0.020	0.015	0.011	0.015	0.016	0.014
sizeCompanycategsmall	-0.004	0.019	-0.006	0.018	0.012	0.017	0.008	0.018	0.027	0.019

Table 4a. *Cont.*

	2016		2017		2018		2019		2020	
Predictors	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*
educ1e3	0.029	0.002	0.027	0.002	0.030	0.002	0.031	0.002	0.030	0.002
educ1e4	0.089	0.002	0.084	0.002	0.084	0.002	0.083	0.002	0.080	0.002
educ1e5	0.192	0.003	0.182	0.003	0.178	0.003	0.177	0.003	0.171	0.003
econSectM	-0.153	0.034	-0.197	0.032	-0.204	0.031	-0.193	0.028	-0.169	0.028
econSectR	-0.117	0.021	-0.160	0.020	-0.151	0.020	-0.155	0.020	-0.165	0.020
Random Effects										
σ^2	0.03		0.03		0.03		0.03		0.03	
τ_{00}	0.01 $_{company}$		0.01 $_{company}$		0.01 $_{company}$		0.01 $_{company}$		0.01 $_{company}$	
	0.04 $_{occupation4}$		0.03 $_{occupation4}$		0.03 $_{occupation4}$		0.03 $_{occupation4}$		0.03 $_{occupation4}$	
	0.01 $_{nace2}$		0.01 $_{nace2}$		0.01 $_{nace2}$		0.01 $_{nace2}$		0.01 $_{nace2}$	
ICC	0.65		0.63		0.63		0.64		0.62	
N	279 $_{company}$		298 $_{company}$		306 $_{company}$		319 $_{company}$		311 $_{company}$	
	52 $_{nace2}$		52 $_{nace2}$		52 $_{nace2}$		53 $_{nace2}$		53 $_{nace2}$	
	274 $_{occupation4}$		275 $_{occupation4}$		270 $_{occupation4}$		273 $_{occupation4}$		267 $_{occupation4}$	
Observations	75534		79807		83370		86496		83597	
Marginal R^2 / Conditional R^2	0.419 / 0.796		0.454 / 0.796		0.419 / 0.782		0.404 / 0.783		0.431 / 0.784	
AIC	-50.570.652		-52.621.768		-56.967.026		-60.067.118		-62.631.910	

Table 4b. A set of identical, additive (gender) M1-models (less variables, defined as in Table 1), fitted for several time values (years 2016-2020) and showing how the so-called adjusted gender pay gap decreases with time. Comparing Table 4a and 4b one may see that the additional variables in model M (as compared to M1) give a correction of one order of magnitude smaller than the adjusted wage gap. It also shows how the random effects capture more of the variability in wages since that is not explained by other covariates as in M.

	2016		**2017**		**2018**		**2019**		**2020**	
Predictors	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*
(Intercept)	7.877	0.023	7.924	0.022	7.971	0.021	8.016	0.022	8.095	0.023
gender1	-0.053	0.002	-0.053	0.002	-0.052	0.002	-0.048	0.001	-0.044	0.001
I(age - mean(age))	0.003	0.000	0.003	0.000	0.003	0.000	0.003	0.000	0.003	0.000
I((age - mean(age))^2)	-0.000	0.000	-0.000	0.000	-0.000	0.000	-0.000	0.000	-0.000	0.000
I(lenEmployComp - mean(lenEmployComp))	0.006	0.000	0.006	0.000	0.006	0.000	0.006	0.000	0.005	0.000
I((lenEmployComp - mean(lenEmployComp))^2)	-0.000	0.000	-0.000	0.000	-0.000	0.000	-0.000	0.000	-0.000	0.000
I(totalHoursScaled - mean(totalHoursScaled))	0.012	0.000	0.012	0.000	0.012	0.000	0.013	0.000	0.015	0.000
marital1	0.018	0.002	0.020	0.002	0.022	0.002	0.021	0.002	0.017	0.002
childage0to2_1	-0.005	0.003	-0.007	0.003	0.000	0.003	-0.003	0.003	-0.007	0.003
childage6to16_1	0.029	0.002	0.029	0.002	0.027	0.002	0.027	0.002	0.028	0.002
childage2to5_1	-0.008	0.002	-0.007	0.002	-0.008	0.002	-0.007	0.002	-0.011	0.002
fulltime1	0.015	0.002	0.016	0.002	0.015	0.002	0.014	0.002	0.004	0.002
backgr1	-0.044	0.002	-0.047	0.002	-0.050	0.002	-0.052	0.002	-0.050	0.002
educ1e3	0.033	0.002	0.031	0.002	0.035	0.002	0.036	0.002	0.035	0.002
educ1e4	0.097	0.002	0.090	0.002	0.092	0.002	0.091	0.002	0.088	0.002
educ1e5	0.200	0.003	0.188	0.003	0.186	0.003	0.185	0.003	0.177	0.003

Table 4b. *Cont.*

	2016		2017		2018		2019		2020	
Predictors	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*	*Estimates*	*SE*
Random Effects										
σ^2	0.03		0.03		0.03		0.03		0.03	
τ_{00}	0.01 company		0.01 company		0.01 company		0.01 company		0.01 company	
	0.06 occupation4		0.06 occupation4		0.05 occupation4		0.05 occupation4		0.05 occupation4	
	0.01 nace2		0.01 nace2		0.01 nace2		0.01 nace2		0.01 nace2	
ICC	0.72		0.72		0.70		0.71		0.71	
N	279 company		298 company		306 company		319 company		311 company	
	52 nace2		52 nace2		52 nace2		53 nace2		53 nace2	
	274 occupation4		275 occupation4		270 occupation4		273 occupation4		267 occupation4	
Observations	75534		79807		83370		86496		83597	
Marginal R^2 / Conditional R^2	0.139 / 0.762		0.142 / 0.756		0.153 / 0.746		0.149 / 0.755		0.135 / 0.752	
AIC	-45.390.527		-46.824.525		-50.148.676		-53.423.162		-57.114.107	

Table 5a. Adjusted wage gap values for all economic sectors, total, A (private), R (government), M (municipalities), years2008-2020, model M and model M1

Year	Total wage-gap	Adjusted wage gap Model M	Adjusted wage gap Model M1	Economic Sector
2008	-0.203	-0.064	-0.069	total
2009	-0.183	-0.063	-0.068	total
2010	-0.178	-0.062	-0.067	total
2011	-0.175	-0.06	-0.065	total
2012	-0.172	-0.059	-0.064	total
2013	-0.165	-0.058	-0.063	total
2014	-0.153	-0.058	-0.061	total
2015	-0.149	-0.057	-0.06	total
2016	-0.139	-0.049	-0.052	total
2017	-0.131	-0.049	-0.052	total
2018	-0.126	-0.048	-0.051	total
2019	-0.129	-0.043	-0.047	total
2020	-0.117	-0.041	-0.043	total
2008	-0.209	-0.067	-0.076	A
2009	-0.213	-0.066	-0.075	A
2010	-0.207	-0.066	-0.075	A
2011	-0.21	-0.066	-0.074	A
2012	-0.205	-0.066	-0.073	A
2013	-0.198	-0.066	-0.073	A
2014	-0.192	-0.066	-0.072	A
2015	-0.175	-0.065	-0.071	A
2016	-0.166	-0.061	-0.062	A
2017	-0.165	-0.06	-0.061	A
2018	-0.164	-0.058	-0.062	A
2019	-0.164	-0.054	-0.058	A
2020	-0.157	-0.056	-0.058	A

Year	Total wage-gap	Adjusted wage gap Model M	Adjusted wage gap Model M1	Economic Sector
2008	-0.185	-0.049	-0.047	R
2009	-0.17	-0.048	-0.046	R
2010	-0.154	-0.048	-0.045	R
2011	-0.15	-0.047	-0.044	R
2012	-0.154	-0.047	-0.043	R
2013	-0.145	-0.046	-0.042	R
2014	-0.132	-0.045	-0.041	R
2015	-0.141	-0.045	-0.04	R
2016	-0.143	-0.042	-0.038	R
2017	-0.134	-0.04	-0.037	R
2018	-0.131	-0.04	-0.037	R
2019	-0.119	-0.034	-0.031	R
2020	-0.1	-0.033	-0.03	R
2008	-0.081	-0.063	-0.069	M
2009	-0.052	-0.061	-0.067	M
2010	-0.043	-0.06	-0.065	M
2011	-0.041	-0.058	-0.064	M
2012	-0.033	-0.056	-0.062	M
2013	-0.031	-0.054	-0.06	M
2014	-0.024	-0.053	-0.058	M
2015	-0.015	-0.051	-0.056	M
2016	-0.007	-0.032	-0.022	M
2017	-0.01	-0.034	-0.027	M
2018	-0.003	-0.032	-0.024	M
2019	0.001	-0.031	-0.024	M
2020	0.006	-0.027	-0.017	M

Table 5b. Example of wage gap Oaxaca-Blinder type of decomposition for model M, year 2019, total economic sector and *all possible choices of a reference model used when defining the decomposition.*

The choice refers to the fact that the *explained* part of a gap is defined as a function of both the differences between average characteristics of men and women *and* the coefficients corresponding to these characteristics in a model fitted as a *reference* model. The unexplained part depends on the differences between the (coefficients of the) models fitted for men (or women) *and the (coefficients of the) reference model* as well as on the corresponding average characteristics. The standard errors are obtained by bootstrapping.[12]

Group	Expl-Coeff (S.E.)	Unexpl-Coeff (S.E)
0	0.032 (0.0028)	0.061 (0.0030)
1	0.045 (0.0027)	0.049 (0.0021)
0.5	0.038 (0.0024)	0.055 (0.0022)
0.42	0.037 (0.0024)	0.056 (0.0021)
-1	0.046 (0.0021)	0.047 (0.0017)
-2	0.036 (0.0020)	0.057 (0.0020)

In Table 5b, Group = 0 means that the chosen reference is the model fitted for women employees, Group=1 means that the model fitted for men data is chosen as a reference, Group=0.5 indicates that equally weighted average of both genders' is used as reference, Group=0.42 shows that an average of the two groups weighted by their number of observations is used, Group=-1 denotes the Neumark reference (pooled regression without group indicator variable) and Group=-2 means that coefficients of pooled regression including group indicator are chosen for reference.

[12] similar to https://cran.r-project.org/web/packages/oaxaca/vignettes/oaxaca.pdf

Table 6. Comparing the influence of various characteristics on the difference between wages of men and women for two different years, by fitting a set of *independent* models where, for each one, only a set of attributes is used as controls, as follows:

M00_company, a model with only random effects due to grouping in companies;

M00_occupation4, a model with random effects due to grouping by occupation (detailed classification, 4 digits);

M00_nace2, a model including only the random effects due to grouping by economic activity as classified by NACE codes, 2 digits;

M00_educ1, a model where only education variable is used as control;

M00_demography, a model where only demographical attributes are included (age, marital status, having children of various ages) and

M00_labour, a model where only labour related attributes are included (working fulltime, the length of employment in the same company and the total number of hours worked, scaled variable)

This illustrates the fact that occupation and labour attributes, followed by economic activity, explain most of the differences between salaries. It also shows that the role of grouping by company and the effects of education and demographical attributes in explaining the differences in wages has decreased between 2010 and 2019. As a cautionary remark, note that one should not add the values in the „explained wage gap" column, since those are obtained according to independent models. Instead, they should be used as indicators about the impact of the included attributes.

Note that two sub-models of the type M00_labour were also tested, one with only the length of employment as a covariate and the other only containing the variables concerning fulltime work and total number of hours worked. The later shows a bigger explanatory impact than the former.

Year	Total Gap	Model	Coefficient of adjusted gap	se(coefficient)	Adjusted Wage Gap	Model-Explained Wage Gap
2019	-0.129					
		M00_company	-0.08	0.003	-0.077	-0.052
		M00_occupation4	-0.051	0.002	-0.05	**-0.079**
		M00_nace2	-0.082	0.003	-0.078	-0.051
		M00_educ1	-0.141	0.002	-0.131	0.002
		M00_demography	-0.115	0.002	-0.109	-0.02
		M00_labour	-0.07	0.002	-0.067	**-0.062**
2010	-0.178					
		M00_company	-0.112	0.003	-0.106	-0.072
		M00_occupation4	-0.076	0.002	-0.073	**-0.105**
		M00_nace2	-0.128	0.003	-0.12	-0.058
		M00_educ1	-0.148	0.003	-0.137	-0.041
		M00_demography	-0.157	0.003	-0.145	-0.033
		M00_labour	-0.092	0.003	-0.088	**-0.09**

Table 7. Descriptive statistics of the data set used for the analysis

Characteristic		N = 998,340[1]
wageHourly		7.66 (7.39, 7.98)
time		0, … , 12
gender		
	0	425,082 (43%)
	1	573,258 (57%)
educ1		
e2		255,945 (26%)
e3		347,004 (35%)
e4		248,601 (25%)
e5		146,790 (15%)
age		40 (28, 52)
lenEmployComp		5 (2, 11)
totalHoursScaled		4.56 (1.99, 5.73)
fulltime		
	0	386,732 (39%)
	1	611,608 (61%)
inlabunion		
	0	30,371 (3.0%)
	1	967,969 (97%)

Characteristic		N = 998,340[1]
regstudent		
	0	804,047 (81%)
	1	194,293 (19%)
regapprentice		
	0	995,446 (100%)
	1	2,894 (0.3%)
backgr		
	0	890,513 (89%)
	1	107,827 (11%)
supervisor		
	0	954,012 (96%)
	1	44,328 (4.4%)
ctworker		
	0	961,241 (96%)
	1	37,099 (3.7%)
monthlyEarn		
	0	169,300 (17%)
	1	829,040 (83%)
shiftPremium		
	0	753,070 (75%)
	1	245,270 (25%)

Characteristic		N = 998,340[1]
marital		
	0	452,278 (45%)
	1	546,062 (55%)
childage0to2		
	0	928,086 (93%)
	1	70,254 (7.0%)
childage6to16		
	0	726,666 (73%)
	1	271,674 (27%)
childage2to5		
	0	862,948 (86%)
	1	135,392 (14%)
sizeCompanycateg		
high		851,226 (85%)
medium		127,854 (13%)
small		19,260 (1.9%)
capitalareaComp		
	0	301,448 (30%)
	1	696,892 (70%)

Characteristic		N = 998,340[1]
equalpaycert		
	0	866,926 (87%)
	1	131,414 (13%)
categ_propY		
	2	536,454 (54%)
	3	461,886 (46%)
categ_propF		
	1	30,381 (3.0%)
	2	579,840 (58%)
	3	388,119 (39%)

1 Median (IQR); n (%)

nace2 and occupation4 not shown (too many categories)

Table 8. Descriptive statistics of the data set, by gender.

Characteristic		0, N = 425,0821	1, N = 573,2581
wageHourly		7.71 (7.42, 8.07)	7.63 (7.37, 7.92)
time		6 (3, 10)	6 (3, 10)
educ1			
e2		119,298 (28%)	136,647 (24%)
e3		165,050 (39%)	181,954 (32%)
e4		77,818 (18%)	170,783 (30%)
e5		62,916 (15%)	83,874 (15%)
age		38 (27, 52)	40 (29, 52)
lenEmployComp		4 (1, 10)	5 (2, 12)
totalHoursScaled		5.20 (2.05, 6.03)	4.19 (1.96, 5.70)
fulltime			
	0	122,318 (29%)	264,414 (46%)
	1	302,764 (71%)	308,844 (54%)
inlabunion			
	0	18,639 (4.4%)	11,732 (2.0%)
	1	406,443 (96%)	561,526 (98%)
regstudent			
	0	349,280 (82%)	454,767 (79%)
	1	75,802 (18%)	118,491 (21%)
regapprentice			
	0	424,787 (100%)	570,659 (100%)
	1	295 (<0.1%)	2,599 (0.5%)
backgr			
	0	377,551 (89%)	512,962 (89%)
	1	47,531 (11%)	60,296 (11%)

Characteristic		0, N = 425,0821	1, N = 573,2581
supervisor			
	0	402,442 (95%)	551,570 (96%)
	1	22,640 (5.3%)	21,688 (3.8%)
ctworker			
	0	389,577 (92%)	571,664 (100%)
	1	35,505 (8.4%)	1,594 (0.3%)
monthlyEarn			
	0	89,407 (21%)	79,893 (14%)
	1	335,675 (79%)	493,365 (86%)
shiftPremium			
	0	335,477 (79%)	417,593 (73%)
	1	89,605 (21%)	155,665 (27%)
marital			
	0	203,035 (48%)	249,243 (43%)
	1	222,047 (52%)	324,015 (57%)
childage0to2			
	0	396,903 (93%)	531,183 (93%)
	1	28,179 (6.6%)	42,075 (7.3%)
childage6to16			
	0	329,789 (78%)	396,877 (69%)
	1	95,293 (22%)	176,381 (31%)
childage2to5			
	0	375,022 (88%)	487,926 (85%)
	1	50,060 (12%)	85,332 (15%)

Characteristic		0, N = 425,0821	1, N = 573,2581
econSect			
A		255,943 (60%)	187,068 (33%)
M		65,661 (15%)	187,712 (33%)
R		103,478 (24%)	198,478 (35%)
sizeCompanycateg			
high		345,976 (81%)	505,250 (88%)
medium		69,228 (16%)	58,626 (10%)
small		9,878 (2.3%)	9,382 (1.6%)
capitalareaComp			
	0	126,206 (30%)	175,242 (31%)
	1	298,876 (70%)	398,016 (69%)
equalpaycert			
	0	370,395 (87%)	496,531 (87%)
	1	54,687 (13%)	76,727 (13%)
categ_propY			
	2	236,000 (56%)	300,454 (52%)
	3	189,082 (44%)	272,804 (48%)
categ_propF			
	1	29,434 (6.9%)	947 (0.2%)
	2	278,183 (65%)	301,657 (53%)
	3	117,465 (28%)	270,654 (47%)

nace2 and occupation4 not shown (too many categories)

Table 9. Overview of the variables in the analysis and explanation.

Variable	Variable name in analysis	Notes		Coding
Intercept	**(Intercept)**	Intercept in model	Intercept in model	
Gender	**gender1**	Gender	Gender	0 (male), 1 (female)
Age	**I(age – mean(age))**	Age: Centred	Age: Scaled	Numerical variable
	I((age – mean(age))^2)	Age squared: scaled	Age squared: scaled	Numerical variable
Lenght of employ-ment	**I(lenEmployComp – mean(lenEmployComp))**	Length of service (employment) in enterprise	length of service (employment) in enterprise	Numerical variable
	I((lenEmployComp – mean(lenEmployComp))^2)	Length of service (employment) in enterprise squared	length of service (employment) in enterprise squared	Numerical variable
Hours worked	**I(totalHoursScaled – mean(totalHoursScaled))**	Total hours worked - scaled	Total hours worked - scaled	Numerical variable
Full time worker	**fulltime1**	Full time worker	Full time worker	0,1 (fulltime)
Education	**educ1e1**	Less than primary education (ISCED <1)	Less than primary education	0, 1 (ISCED <10)
	educ1e2	Basic education (ISCED 1,2)	Basic education (ISCED 1,2)	0, 1 (ISCED 10:29)
	educ1e3	Upper secondary education (ISCED 3,4)	Upper secondary education (ISCED 3,4)	0, 1 (ISCED 30:49)
	educ1e4	Teritery education bachelor or diploma (ISCED 5, 6)	Teritery education bachelor or diploma (ISCED 5, 6)	0, 1 (ISCED 50:69)
	educ1e5	Teritery education master/doctor (ISCED 7,8)	Teritery education master/doctor (ISCED 7,8)	0, 1 (ISCED 70:89)
Marital status	**marital1**	Marital status: Married	Marital status: Married	0, 1 (married)
Children	**childage0to21**	Has children younger then 2 years	Has children younger then 2 years	0, 1 (one or more children in the age range)
	childage6to161	Has children that are between 2 years old and younger then 6 years old	Has children that are between 2 years old and younger then 6 years old	0, 1 (one or more children in the age range)
	childage2to51	Has children that are between 6 years old and younger then 16 years old	Has children that are between 6 years old and younger then 16 years old	0, 1 (one or more children in the age range)

Table 9. Overview of the variables in the analysis and explanation (*cont.*)

Variable	Variable name in analysis	Notes		Coding
Background	**backgr1**	Has a foreign background	Has a foreign background	0, 1 (married)
Student	**regstudent1**	Student in student register	Student in student register	0, 1 (student)
Apprentice	**regapprentice1**	Is registerd as student in wage survey - apprentice	Is registerd as student in wage survey - apprentice	0, 1 (apprentice)
In labour union	**inlabunion1**	Member of a union	Member of a union	0, 1 (in labour union)
Supervisor	**supervisor1**	Supervisor/management	Supervisor/management	0, 1 (supervisor)
Craft/trade worker	**ctworker1**	Craft/trade worker	Craft/trade worker	0, 1 (craft/trade worker)
Monthly earnings	**monthlyEarn1**	Monthly earnings (Monthly or hourly earnings)	Monthly earnings (Monthly or hourly earnings)	0, 1 (monthly earnings)
Shift Premium	**shiftPremium1**	Shift premium	Shift premium	0,1 (shift premium)
Capital area	**capitalareaComp1**	Enterprise in captial area	Enterprise in captial area	0,1 (company in capital area)
	categ_propF1	Proportion of women in a particular occupation <33.3%	Proportion of women in a particular occupation <33.3%	0, 1 (individual in occupation which has the particular gender proportion)
Proportion of male employees	**categ_propF2**	Proportion of women in a particular occupation between 33.3 and 66.6%	Proportion of women in a particular occupation between 33.3 and 66.6%	0, 1 (individual in occupation which has the particular gender proportion)
	categ_propF3	Proportion of women in a particular occupation >66.6%	Proportion of women in a particular occupation >66.6%	0, 1 (individual in occupation which has the particular gender proportion)

Table 9. Overview of the variables in the analysis and explanation (*cont.*)

Variable	Variable name in analysis	Notes		Coding
Proportion of employees older than 35	**categ_propY1**	Proportion of employees older than 35: <33.3%	Proportion of employees older than 35: <33.3%	0, 1 (Individual in an occupation which has the particular proportion of employees older than 35)
	categ_propY2	Proportion of employees older than 35: between 33.3 and 66.6%	Proportion of employees older than 35: between 33.3 and 66.6%	0, 1 (Individual in an occupation which has the particular proportion of employees older than 35)
	categ_propY3	Proportion of employees older than 35: >66.6%	Proportion of employees older than 35: >66.6%	0, 1 (Individual in an occupation which has the particular proportion of employees older than 35)
Equal pay certificate	**equalpaycert1**	Equal pay official certification	Equal pay official certification	0, 1 (certified)
Company size	**sizeCompanycateglarge**	Size of the company: 250 or more employees	Size of the company: 250 or more employees	0, 1 (large sized company)
	sizeCompanycategmedium	Size of the company: between 50 and 250	Size of the company: between 50 and 250	0, 1 (medium sized company)
	sizeCompanycategsmall	Size of the company: less than 50	Size of the company: less than 50	0, 1 (small sized company)
Economic sector	**econSectA**	Economic sector: Private sector	Economic sector: Private sector	0, 1 (Private sector employee employee) - default group
	econSectM	Economic sector: Municipalities	Economic sector: Municipalities	0, 1 (Municipality employee)
	econSectR	Economic sector: Government	Economic sector: Government	0, 1 (Government employee)

Statistical Series *Working papers*

Vol. 106 • Issue 6 7 September 2021

ISSN 1670-4770 Supervision: Violeta Calian, chief methodologist, Unit of Administration and Cooperation

• violeta.calian@statice.is

Contributors: Experts from the Unit of Wages, income and education

• wages@statice.is

www.statice.is

Telephone (+354) 528 1000

Fax (+354) 528 1099