

# Multilevel modelling for gender wage gap analysis

Violeta Calian, Statistics Iceland, [violeta.calian@hagstofa.is](mailto:violeta.calian@hagstofa.is)

Kristín Arnórsdóttir, Statistics Iceland, [kristin.arnorsdottir@hagstofa.is](mailto:kristin.arnorsdottir@hagstofa.is)

## Abstract

*The purpose of this paper is to estimate the difference in hourly wages between female and male employees and its time evolution in Iceland, while accounting for a set of measured individual and employment characteristics. These include work experience in a given company, several 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.*

*We find, by using multilevel models of wages (with interaction and both frequentist and Bayesian estimates), that the observed, total gap is explained by (i) the differences in average **characteristics** of men and women, and (ii) the differences in the **effects** of these characteristics on wages for the two genders. Certain covariates have a statistically significant advantageous effect on the wages of female/male employees and in addition this effect evolves with time.*

*For instance, women gain less than men do 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 than men do, by being highly educated, 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 or by men, and/or working for a company with equal pay certification. Differences of occupational age composition effects are non-significant in Iceland, unlike differences observed in other countries where significant advantages are reported for men.*

**Keywords:** multilevel models, gender pay gap, open R code

## 1. Introduction

In this paper we reformulate the problem of analysing the gender wage gap structure and its time evolution in terms of statistical hypotheses testing. We provide a rigorous solution and the associated quantitative estimates as well as the corresponding uncertainty measures.

The wage gap is defined as the relative difference between the *mean* hourly wages of women and men. Its estimate can be reported as either raw-data based or as an adjusted value when controlling for the influence of other variables. These are

individual and employment characteristics such as work experience in a given company, several demographic attributes of employees (age, marital status, having children), education, occupation, economic activity of employer, female/male proportion of employees in the occupation category, economic sector and activity, size and location of company. The difference between wages of female and male employees is actually distributed according to non-standard laws which could be better described by their higher cumulants or quantiles.

The main research questions we tested in this study are the following:

- (i) which characteristics have a statistically significant influence on the hourly wages and did this influence change over the period 2008-2020?
- (ii) do these influences depend on gender and if so, do they evolve with time?
- (iii) what is the estimated wage gap and how did it change during the past decade, when accounting for the differences between the characteristics of male and female employees?

The first research question is answered by performing joint significance tests for the predictors included in a wage model, at fixed values of time or for time dependent models as described in the following section. The second question is equivalent to performing joint significance tests of interaction between the gender variable and other predictors included in the model, while taking into account the dynamical aspects as well. This is also a test of whether two significantly different models can be fitted for the wages of men and women. The last research question is concerned with giving estimates and uncertainty measures for the gap, based on best additive wage models.

The solution provided by our analysis uses multilevel/mixed/hierarchical (MLM) models of wages as functions of individual and group characteristics of employees as defined in the next section. The groups/clusters of correlated observations are defined by wages of employees with common occupation, economic activity or company. In addition, employees themselves define wage clusters when considering time dependent models and longitudinal data with autocorrelated observations.

## **2. Data and methods**

## 2.1. Data

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<sup>1</sup>. The quality of variables is not uniform and some are better than others, as described in [H1]-[H3].

The following set of variables has been used for modelling:

- the outcome of interest (*wageHourly*): the (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 of the *ISCED* classification 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 centred 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<sup>2</sup> 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: over 250 employees), capital area location of company (*capitalareaComp*: 0 or 1), equal-pay certificate (*equalpay*: 0 or 1);

- occupational attributes:

---

<sup>1</sup> 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.

<sup>2</sup> This is a generic term, which does include private sector companies but also employers from the state economic sector and municipalities.

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).

## 2.2. Model choice and model fitting

The main advantages of MLM models are improved accuracy and decreased uncertainty, when compared with models which do not exploit the data structure, as in the case of OLS, fixed effects or random effects models. These advantages are accompanied by an optimum use of data, due to strength borrowing by modelling similar observations. MLM provide better explanation of variability in the outcome as well as group specific predictions and estimates of model parameters.

Models with interactions and/or random effects on intercepts and slopes have been examined although results were not all included here. A data exploratory analysis stage preceded any model fitting and is shown in detail in [H0]. It confirms the set of variables correlated with wages and shows distributional differences by various factors. Heteroskedasticity and heterogeneity tests were also performed. Finally, several models were fitted by using:

- (i) maximum likelihood, for the purpose of comparing the performance of different models
- (ii) restricted maximum likelihood, for fast and more accurate estimation
- (iii) Bayesian framework, for best uncertainty estimates and Bayesian variable selection as well as model averaging purposes.

The models we built<sup>3</sup> for testing/estimating have the following structures:

- (i) Null models, for testing the significance of clustering of observations and for providing a baseline to more complex models. They include only random effects and show for instance that intra-class correlation (ICC) for company, economic activity, and especially occupation variables are significant, thus wages are more different between classes defined by such grouping variables than within.
- (ii) Two-level models, when modelling data subsets defined by fixed time values. A simple type may be written as a composite model (see [H0]) or equivalently, displaying the hierarchical model structure:

---

<sup>3</sup> See our shared R-code at <https://github.com/violetacln/GIW>

$$y_{jk} = a_j + \sum_m b_j^m X_{jk}^m + inter + \epsilon_{jk}$$

$$a_j = a^0 + \sum_l a^{1(l)} Z_j^l + inter + \epsilon_j^a$$

$$b_j = b^0 + \sum_l b^{1(l)} Z_j^l + inter + \epsilon_j^b$$

$$(a^0, b^0, a^1, b^1)' \sim N((0,0,0', 0')', \Sigma)$$

$$\epsilon_{jk} \sim N(0, \sigma^2)$$

$$(\epsilon_j^a, \epsilon_j^b)' \sim N((0,0)', \Sigma_{ab})$$

with and without interaction terms (*inter*) between the gender variable and/or several individual or group characteristics. Notations like  $0'$  indicate a whole vector.

This solution was used for modelling data of years between 2016 and 2020, since sufficiently detailed and high quality for fitting separate models for distinct years. It offers the advantage of fast computation and easier interpretation of results, even for models with interactions.

(iii) Three-level models, when using data of multiple years and 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 a composite one ([H0]) or, by hierarchical levels, as a (conditional) growth model (linear or more complex, e.g. using spline functions):

$$y_{ijk} = a_{ij} + b_{ij} t_{ijk} + \epsilon_{ijk}$$

$$a_{ij} = a_i + \sum_l a^{X(l)} X_{ij}^l + inter + \epsilon_{ij}^a$$

$$b_{ij} = b_i + \sum_l b^{X(l)} X_{ij}^l + inter + \epsilon_{ij}^b$$

$$a_i = a + \sum_l a_i^{Z(l)} Z_i^l + inter + \epsilon_i^a$$

$$\begin{aligned}
 b_i &= b + \sum_l b_i^{Z(l)} Z_i^l + inter + \epsilon_i^b \\
 a_i^X &= a^X + \sum_l a_i^{XZ(l)} Z_i^l + inter + \epsilon_i^{a0} \\
 b_i^X &= b^X + \sum_l b_i^{XZ(l)} Z_i^l + inter + \epsilon_i^{b0}
 \end{aligned}$$

$$\begin{aligned}
 (\epsilon_{ij}^a, \epsilon_{ij}^b)' &\sim N((0,0)', \Sigma) \\
 \Sigma_{aa} &= \sigma_a^2, \Sigma_{bb} = \sigma_b^2, \Sigma_{ab} = \rho\sigma_a\sigma_b \\
 (\epsilon_i^a, \epsilon_i^b, \epsilon_i^{a0}, \epsilon_i^{b0})' &\sim N((0,0,0',0')', \Sigma^0) \\
 \epsilon_{ijk} &\sim N(0, \sigma^2)
 \end{aligned}$$

with and without interaction terms (inter) between the gender variable and several characteristics. Note that time dependence could be also modelled by spline functions, higher polynomials or other functions which describe the wage growth with time, if appropriate.

In this formula we denote by  $t_{ijk}$  the time points ( $k$ ) of observing (log) wages of individual  $i$  in group  $j$ ,  $X_{ij}^l$  are model characteristics ( $l$ ) of individuals ( $i$ ), including gender, cross-classified in groups ( $j$ ) while  $Z_{jk}^l$  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  $\Sigma$  - individual level and  $\Sigma^0$  - at group level. We chose simple, diagonal matrices, after several exploratory model/computational tests. We included random slope components for testing purposes but not for estimation purposes.

Note that 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. This was the case for characteristics such as age, length of employment which vary during a given time between employees but also within each employee during the 2008-2020 interval. Other characteristics, like gender or education did not change for a given employee or had a stepwise change. The explicit form of the models allows us

to evaluate the significance of effects and their time trends as explained in the following section.

### **3. Results**

In order to illustrate the way statistical hypothesis testing offers the solution to the gender wage gap analysis, we refer to the Figure 1 in the Appendix. It shows the Bayesian credible intervals (CI) of a (fixed time) model which includes random effects for the intercepts but not for slopes as well as a large set of predictors and their interaction with the gender variable. Such a model describes the fact that:

- (i) wages depend on the variables with CI not including zero (at various probability levels) such as gender, marital status, having children, location of company, proportion of women in the occupation of the employee, education, having or not an Icelandic background, being a supervisor.
- (ii) the effect of some predictors (with CI of corresponding interaction terms not including zero at some probability level) on wages depend on gender, or equivalently the effect of gender on wages depend on some predictors. For instance, being a supervisor or being married and a man gives a bigger gain in wages than being a woman with otherwise equal characteristics. The wage gap between genders still grows with age and employment length.

We also notice that the influence on wages of having an Icelandic background, having an occupation with a balanced mixture of men and women or the proportion of young employees in one's occupation do not vary by gender. The effect of the economic sector on wages is also both significant and different for the two genders. We exemplify the consequences of this testing process by the model given in Table 1 of the Appendix.

These conclusions are related to the complementary results regarding the adjusted wage gap, i.e. the gender effect in additive models including all other significant predictors. Such measures show a difference in the decay slopes between economic sectors, with the municipalities' sector basically closing the wage gap in recent years, as shown in Table 2 of the Appendix while the private sector still decreasing much more slowly.



The standard decomposition of the wage gap into “explained” versus “not explained” parts needs to be appropriately generalised for (time dependent) MLM models since in its original form it is based on linear, OLS estimated. This is the object of work in progress and is aimed at separating, for these more general problems:

- (i) the contribution to the wage gap of differences in *average characteristics*, between gender groups
- (ii) the contribution to the wage gap of the differences in *effects of characteristics*, between gender groups

at both fixed time values and as changes between years.

#### **4. Discussion**

Our analysis is easy to apply to new data sets of similar structures. We emphasize the need of careful interpretation of modelling results as well as the importance of using model testing as a statistically sound method for explaining the structure and evolution of wages and gender wage gap.

A limitation of the present study is that the participation effect / self-selection has not been independently measured for our data, although data does mirror employment distributions over economic sectors and enterprise type/size distributions. 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 shown 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 influence employment/participation therefore it is quite likely that the participation effect is not significant.

#### ***Acknowledgement***



The authors are grateful to the experts from the Unit of Wages, income and education of Statistics Iceland for preparing the data set used in this analysis.

## 5. References

- [H0] Calian, V. (2021) Icelandic gender pay gap analysis 2008-2020, Statistical Series of Statistics Iceland, Working Papers series,  
<http://hagstofan.s3.amazonaws.com/media/public/2021/2b7b1b12-47b5-4dfd-b1af-97437a20c243.pdf>
- [H1] Indriðadóttir M. K., Sigurðsson E. (2010) Gender wage differential in the private sector 2000–2007, Statistical Series of Statistics Iceland,  
[https://hagstofa.is/media/43558/hag\\_100218.pdf](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](https://www.stjornarradid.is/media/velferdarraduneyti-media/media/rit-og-skyrslur-2015/Launamunur_karla_og_kvenna_19052015b.pdf)
- [H3] Gylfadóttir M. V., Indriðadóttir M. K., Karlsson A. Ö. (2018) Statistical series of Statistics Iceland (2018), Analysis on Gender Pay Gap 2008–2016,  
[https://hagstofan.s3.amazonaws.com/media/public/4a70b304-09ff-4c6b-adcf-050ac2bef384/pub\\_doc\\_mqp71tr.pdf](https://hagstofan.s3.amazonaws.com/media/public/4a70b304-09ff-4c6b-adcf-050ac2bef384/pub_doc_mqp71tr.pdf)
- [1] Meara, K., Pastore, F. & Webster, A. (2020) 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] 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.

- [5] 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>.
- [6] Jared E. Knowles and Carl Frederick (2020). merTools: Tools for Analyzing Mixed Effect Regression Models. R package version 0.5.2. (URL: <https://CRAN.R-project.org/package=merTools>)
- [7] 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> ).
- [8] Lüdtke 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>
- [9] Makowski, D., Ben-Shachar, M., & Lüdtke, 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
- [10] Makowski, D., Ben-Shachar, M.S., Patil, I. & Lüdtke, 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: .
- [11] 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/>.
- [12] Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>
- [13] Bunel, M, Guironnet, J-P (2017) Income inequalities for recently graduated French workers: a multilevel modelling approach, *Empir Econ* (2017) 53:755-778.
- [14] Correll, S. (2001). Gender and the Career Choice Process: The Role of Biased Self-Assessments. *American Journal of Sociology*, 106(6), 1691-1730. doi:10.1086/321299.

## 6. Appendices

### Appendix 6.a.

**Figure 1.** Maximal model with interaction, for fixed time value (year 2020), generated in the model building process.

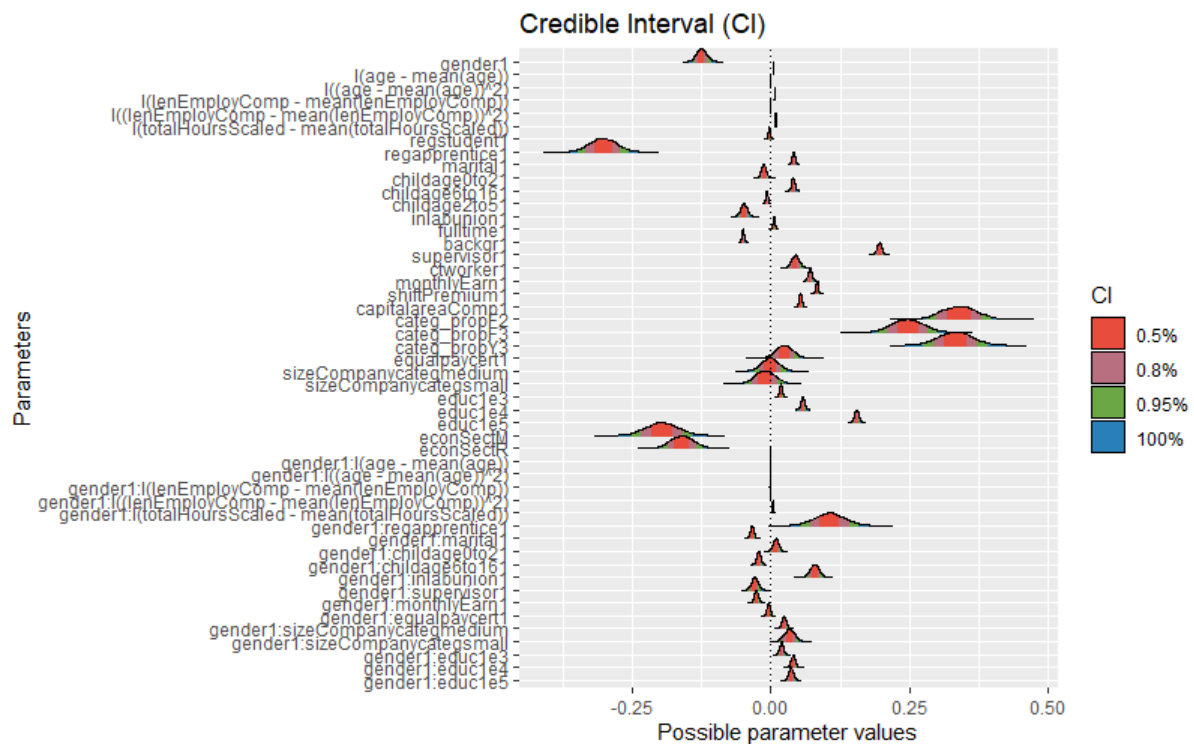


Figure 1 shows why several effects and interactions between gender and covariates, which are tested by fitting a maximal model, are not included in the final model. Credible intervals are practical and rather intuitive ways of assessing uncertainty. The estimate of a given effect for example, lies with 95% (or other values like 80%, or 50%) probability in its credible interval. Large credible intervals indicate a large uncertainty in estimates.

**Appendix 6.b.**

**Table 1.** A set of identical models (with significant interaction between gender and other attributes) fitted for several years (2016 and 2019, 2020), showing the differential effects of characteristics (negative coefficients show disadvantage for women) depending on gender and changing with time.

<i>Predictors</i>	2016		2019		2020	
	<i>Estimates</i>	<i>SE</i>	<i>Estimates</i>	<i>SE</i>	<i>Estimates</i>	<i>SE</i>
(Intercept)	7.492	0.040	7.621	0.041	7.641	0.042
gender1	-0.154	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
I((age – mean(age))^2)	-0.000	0.000	-0.000	0.000	-0.000	0.000
I(lenEmployComp – mean(lenEmployComp))	0.007	0.000	0.007	0.000	0.005	0.000
I((lenEmployComp – mean(lenEmployComp))^2)	-0.000	0.000	-0.000	0.000	-0.000	0.000
I(totalHoursScaled – mean(totalHoursScaled))	0.007	0.001	0.008	0.001	0.009	0.001
regstudent1	-0.002	0.002	-0.004	0.002	0.000	0.002
regapprentice1	-0.280	0.047	-0.307	0.024	-0.245	0.026
marital1	0.039	0.003	0.040	0.003	0.033	0.003
childage0to2_1	-0.021	0.004	-0.013	0.004	-0.020	0.004
childage6to16_1	0.040	0.003	0.040	0.003	0.044	0.003
childage2to5_1	-0.009	0.002	-0.007	0.002	-0.012	0.002
inlabunion1	-0.027	0.006	-0.047	0.005	-0.055	0.005
fulltime1	0.007	0.002	0.005	0.002	-0.002	0.002
backgr1	-0.041	0.002	-0.051	0.002	-0.049	0.002
supervisor1	0.182	0.005	0.193	0.004	0.179	0.004
ctworker1	0.012	0.008	0.040	0.007	0.033	0.007
monthlyEarn1	0.087	0.004	0.076	0.003	0.061	0.003
shiftPremium1	0.084	0.002	0.083	0.002	0.076	0.002
capitalareaComp1	0.049	0.003	0.053	0.003	0.065	0.003

categ_propF2	0.213	0.024	0.336	0.031	0.390	0.031
categ_propF3	0.277	0.031	0.245	0.028	0.244	0.027
categ_propY3	0.260	0.023	0.330	0.030	0.382	0.030
equalpaycert1			0.021	0.015	0.017	0.014
sizeCompanycategmedium	-0.027	0.016	-0.001	0.015	0.006	0.014
sizeCompanycategsmall	-0.017	0.019	-0.011	0.018	0.008	0.019
educ1e3	0.014	0.003	0.020	0.002	0.019	0.002
educ1e4	0.067	0.003	0.061	0.003	0.061	0.003
educ1e5	0.179	0.004	0.161	0.004	0.154	0.004
econSectM	-0.176	0.033	-0.212	0.028	-0.186	0.028
econSectR	-0.137	0.021	-0.172	0.020	-0.181	0.020
gender1:(age – mean(age))	-0.001	0.000	-0.001	0.000	-0.001	0.000
gender1:((age – mean(age))^2)	0.000	0.000	0.000	0.000	0.000	0.000
gender1:(lenEmployComp – mean(lenEmployComp))	-0.002	0.000	-0.002	0.000	-0.001	0.000
gender1:((lenEmployComp – mean(lenEmployComp))^2)	0.000	0.000	0.000	0.000	0.000	0.000
gender1:(totalHoursScaled – mean(totalHoursScaled))	0.002	0.001	0.004	0.001	0.003	0.001
gender1:regapprentice1	0.130	0.049	0.109	0.026	0.088	0.028
gender1:marital1	-0.037	0.003	-0.033	0.003	-0.029	0.003
gender1:childage0to2_1	0.018	0.005	0.010	0.005	0.013	0.005
gender1:childage6to16_1	-0.019	0.004	-0.022	0.003	-0.024	0.003
gender1:inlabunion1	0.106	0.008	0.083	0.008	0.100	0.008
gender1:supervisor1	-0.031	0.006	-0.025	0.006	-0.033	0.006
gender1:monthlyEarn1	-0.048	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	0.022	0.004	0.018	0.004
gender1:sizeCompanycategsmall	0.025	0.011	0.033	0.009	0.033	0.010
gender1:educ1e3	0.022	0.004	0.016	0.003	0.016	0.003
gender1:educ1e4	0.033	0.004	0.033	0.004	0.029	0.004
gender1:educ1e5	0.017	0.005	0.026	0.004	0.026	0.004
gender1:econSectM	0.045	0.004	0.038	0.004	0.036	0.003
gender1:econSectR	0.040	0.004	0.035	0.004	0.034	0.004

Random effects (Table 1 - continued)

$\sigma^2$	0.03		0.03		0.03
T00	0.01 <sub>company</sub>		0.01 <sub>company</sub>		0.01 <sub>company</sub>
	0.03 <sub>occupation4</sub>		0.03 <sub>occupation4</sub>		0.03 <sub>occupation4</sub>
	0.01 <sub>nace2</sub>		0.01 <sub>nace2</sub>		0.01 <sub>nace2</sub>
ICC	0.65		0.63		0.62
N	279 <sub>company</sub>		319 <sub>company</sub>		311 <sub>company</sub>
	52 <sub>nace2</sub>		53 <sub>nace2</sub>		53 <sub>nace2</sub>
	274 <sub>occupation4</sub>		273 <sub>occupation4</sub>		267 <sub>occupation4</sub>
Observations	75534		86496		83597
Marginal R2 / Conditional R2	0.427 / 0.798		0.412 / 0.785		0.438 / 0.786
AIC	- 51.710.150		- 61.129.053		- 63.713.281

**Appendix 6.c.**

**Table 2.** Adjusted wage gap values for all economic sectors, total, A (private), R (government), M (municipalities), years 2008-2020, additive model.

Year	Total wage-gap	Adjusted wage gap Additive M	Economic Sector
2008	-0.203	-0.069	total
2009	-0.183	-0.068	total
2010	-0.178	-0.067	total
2011	-0.175	-0.065	total
2012	-0.172	-0.064	total
2013	-0.165	-0.063	total
2014	-0.153	-0.061	total
2015	-0.149	-0.06	total
2016	-0.139	-0.052	total
2017	-0.131	-0.052	total
2018	-0.126	-0.051	total
2019	-0.129	-0.047	total
2020	-0.117	-0.043	total
2008	-0.209	-0.076	A
2009	-0.213	-0.075	A

<b>2010</b>	-0.207	-0.075	A
<b>2011</b>	-0.21	-0.074	A
<b>2012</b>	-0.205	-0.073	A
<b>2013</b>	-0.198	-0.073	A
<b>2014</b>	-0.192	-0.072	A
<b>2015</b>	-0.175	-0.071	A
<b>2016</b>	-0.166	-0.062	A
<b>2017</b>	-0.165	-0.061	A
<b>2018</b>	-0.164	-0.062	A
<b>2019</b>	-0.164	-0.058	A
<b>2020</b>	-0.157	-0.058	A
<b>2008</b>	-0.185	-0.047	R
<b>2009</b>	-0.17	-0.046	R
<b>2010</b>	-0.154	-0.045	R
<b>2011</b>	-0.15	-0.044	R
<b>2012</b>	-0.154	-0.043	R
<b>2013</b>	-0.145	-0.042	R
<b>2014</b>	-0.132	-0.041	R
<b>2015</b>	-0.141	-0.04	R
<b>2016</b>	-0.143	-0.038	R
<b>2017</b>	-0.134	-0.037	R
<b>2018</b>	-0.131	-0.037	R
<b>2019</b>	-0.119	-0.031	R
<b>2020</b>	-0.1	-0.03	R
<b>2008</b>	-0.081	-0.069	M
<b>2009</b>	-0.052	-0.067	M
<b>2010</b>	-0.043	-0.065	M
<b>2011</b>	-0.041	-0.064	M
<b>2012</b>	-0.033	-0.062	M
<b>2013</b>	-0.031	-0.06	M
<b>2014</b>	-0.024	-0.058	M
<b>2015</b>	-0.015	-0.056	M
<b>2016</b>	-0.007	-0.022	M
<b>2017</b>	-0.01	-0.027	M
<b>2018</b>	-0.003	-0.024	M
<b>2019</b>	0.001	-0.024	M
<b>2020</b>	0.006	-0.017	M