



EUROPEAN COMMISSION
EUROSTAT

Directorate F: Social statistics

Doc. DSS/2018/March/4.3

Item 4.3

Adjusted gender pay gap

MEETING OF THE EUROPEAN DIRECTORS OF SOCIAL STATISTICS

LUXEMBOURG, 1-2 MARCH 2018

BECH BUILDING, ROOM AMPÈRE

The purpose of this document is to inform DSS members about Eurostat's proposal to adjust the gender pay gap using the Structure of Earnings Survey data. Firstly, the gender policy context as well as the gender pay gap indicator and the adjustment (decomposition) method are introduced. Then, the results are shown, including estimations for the adjusted gender pay gap which is also called the "unexplained gender pay gap". Finally, the outcomes of the discussion in LAMAS Working Group in October 2017 and DSS Board in December 2017 are presented.

1. Policy background

Reducing the gender pay gap (GPG) is one of the key priorities of gender policies at the EU and national levels. At the EU level, the European Commission prioritised "reducing the gender pay, earnings and pension gaps and thus fighting poverty among women" as one of the key areas in its "Strategic engagement for gender equality 2016-2019".

One of the European Commission's actions in this area is the European Equal Pay Day (in early November) with Europe-wide information activities to raise awareness of EU citizens on possible discriminations between men and women. The unadjusted gender pay gap (unadjusted GPG) gets high visibility in this context as the key indicator to monitor progress in the field of equal earnings for men and women. The importance of the unadjusted GPG indicator is now reinforced by its inclusion in the Indicators Scoreboard of the European Pillar of Social Rights.

2. The gender pay gap

2.1 The unadjusted GPG

The unadjusted GPG is calculated as the difference between the average hourly earnings of women and men expressed as a percentage of the average hourly earnings of men. It thus provides a simple indicator of wage inequalities which explains its wide use by policy makers.

However, the unadjusted GPG entangles in its measurement both possible discrimination between men and women, in terms of "unequal pay for equal work", as well as the impact of differences in the average characteristics of men and women in the labour market.

2.2 The unexplained (adjusted) and explained GPGs

Against this backdrop, Eurostat has developed a methodology to decompose the unadjusted GPG using the Structure of Earnings Survey (SES) microdata (detailed methodology provided in annexes 1 and 2). The methodology is based on the Blinder-Oaxaca decomposition. The SES microdata provide information on the earnings of individual employees as well as on some personal, job and enterprise characteristics.

Within the decomposition approach, it must be decided which earnings structure constitutes the non-discriminatory benchmark against which to decompose the difference between hourly earnings of men and women. It is assumed, in accordance with the definition of the unadjusted gender pay gap, that the male earnings structure constitutes this benchmark.

Eurostat's methodology provides a decomposition of the difference between the means of log hourly earnings of men and women into explained and unexplained parts. The explained part is the difference between male and female earnings which is due to the differences in the average characteristics (sector of activity, age, occupation, etc.) of male and female employees. The unexplained part measures the difference between the financial returns to men and women with the same characteristics.

Finally, the explained GPG and the adjusted GPG (which is also called the "unexplained GPG") can be calculated by applying, respectively, a percentage of the explained part and the unexplained part of the difference between the means of log hourly earnings of men and women to the unadjusted GPG. Details on the methodology are presented in annexes 1 and 2.

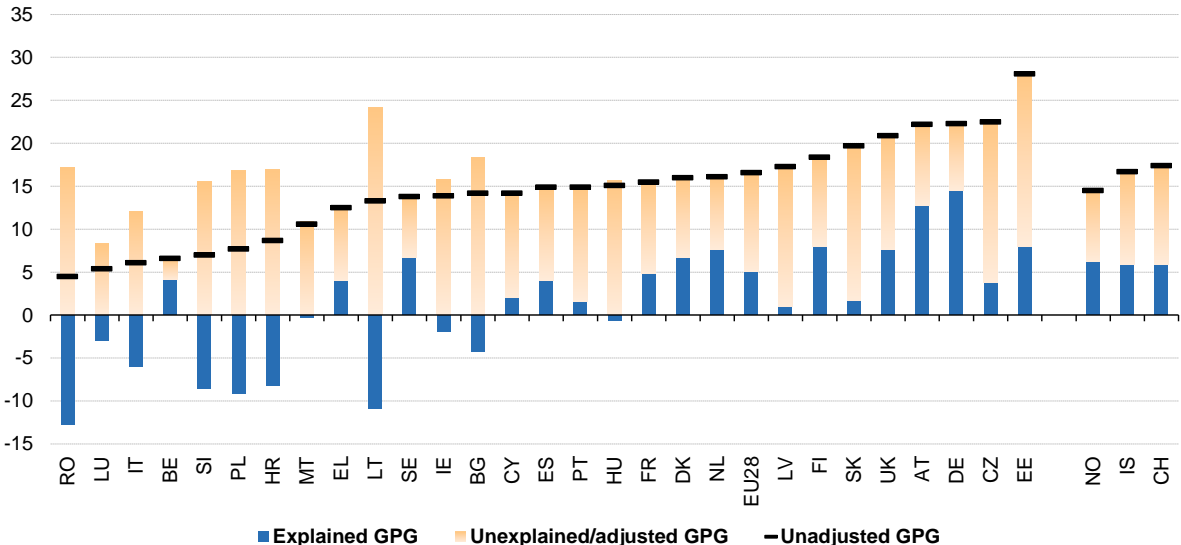
3. Results

Figure 1 presents the results of a decomposition of the unadjusted GPG into the explained GPG and the unexplained (adjusted) GPG. At the EU level, the overall explained GPG is 5.1% against 16.6% for the unadjusted GPG. This means that women are expected to earn 5.1% less than men due to better, on average, characteristics of men compared to women in the European labour markets.

Across the EU Member States, the overall explained GPG varies from -12.7% in Romania to 14.5% in Germany. A negative gap of 12.7% in Romania means that women are expected to earn 12.7% more than men due to better, on average, characteristics of women compared to men in the labour market.

The overall explained gap is negative in eleven Member States: Bulgaria, Croatia, Hungary, Ireland, Italy, Lithuania, Luxembourg, Malta, Poland, Romania and Slovenia, and positive in seventeen Member States. For the countries with the negative explained gap, the unexplained (adjusted) GPG is higher than the unadjusted figure.

Figure 1: Decomposition of the unadjusted GPG (% of male hourly earnings)



For the EU as a whole, the unexplained (adjusted) GPG was 11.5% in 2014 against 16.6% for the unadjusted GPG. The unadjusted GPG varies from 4.5% in Romania to 28.1% in Estonia, whereas the unexplained (adjusted) GPG varies from 2.5% in Belgium to 24.2% in Lithuania. When comparing the ranking of EU countries (and arranged from smallest to largest GPG) before/after adjustment we observe the most significant impact for Romania, Lithuania, Croatia, Poland, Slovenia, Bulgaria and Italy moving by at least 10 positions downwards in the ranking as well as for Germany, Austria, the Netherlands Finland and Denmark moving by at least 10 positions upwards. More details on the results are presented in annex 1.

4. Outcome of the LAMAS and DSS Board's consultations and next steps

In July 2017, Eurostat consulted LAMAS as well as the contact persons in the NSIs responsible for providing the annual unadjusted gender pay gap, on a proposal to adjust the GPG. This proposal, that included a methodology and test results, was also discussed at the LAMAS meeting of 10-11 October 2017. Eurostat proposed, in particular, that the adjusted GPG indicator become the new headline indicator with the unadjusted GPG as the background indicator.

The consultations showed that a majority of LAMAS delegates were not in favour of making the adjusted GPG the new headline indicator instead of the unadjusted one.

LAMAS broadly agreed on the decomposition method proposed to calculate the adjusted GPG for countries and the EU as a whole. On the one hand, a group of Member States agreed with the Oaxaca decomposition method proposed by Eurostat (some of them already use the same method at the national level, e.g. Belgium, Finland and Germany). On the other hand, there was a number of Member States that were in favour of a more thorough discussion on the Oaxaca decomposition method pointing at some other possible options (e.g. using different earnings structures as non-discriminatory benchmarks) or some limitations in that method.

LAMAS agreed that Eurostat would publish the results as a methodological paper in "Statistics Explained" along with references to national adjusted GPG's, to make users aware that results depend on the data used and the assumptions made. LAMAS also agreed that Eurostat would consult LAMAS in written on a layout to possibly publish the adjustment factors, alongside the adjusted GPG, as experimental statistics. Eurostat was also asked to propose alternative labels for the "adjusted GPG". Eurostat considers to use - as already in this document - the term "unexplained GPG".

The proposal was also discussed at the meeting of the DSS Board on 4-5 December 2017. The majority of the DSS board (DSSB) was in favour to keep the unadjusted GPG as the headline indicator. The DSSB agreed that Eurostat would publish the methodology and results for the adjusted GPG in 'Statistical working paper'.

Eurostat hopes that the methodology proposed and the results obtained will stimulate further discussion within the European Statistical System on a common method to adjust the GPG indicator.

5. Conclusions

Whereas the unadjusted GPG provides an overall measure of the gap between the hourly earnings of men and women, it cannot be interpreted as either an approximation or even an upper bound of possible discrimination in the sense of unequal pay for equal work which is the main concern of EU gender policy.

A part of the unadjusted GPG can be explained by the differences in the average characteristics of male and female workers. This explained part can vary significantly across the countries. In most of the countries men have, on average, better characteristics than women in the labour market which results in the explained (adjusted) GPG being lower than the unadjusted figure. However, in some countries, the reverse situation holds which results in the unexplained (adjusted) GPG being higher than the unadjusted figure. This is notably the case in some countries where the employment rate of women is lower than men which may create a selection bias in the population of the employed women.

For these reasons, the unexplained (adjusted) GPG indicator can give a better approximation of possible discrimination than the unadjusted figures. However, even the unexplained (adjusted) GPG should be interpreted with caution as other explanatory variables not included in the regression (e.g. because they are unobserved) may change the results of the decomposition.

Furthermore, the decomposition method allows measuring and comparing gender segregation effects in the European labour markets. These segregation effects can be analysed on the basis of the overall differences in average characteristics of men and women, e.g. in economic activities or occupations.

Taken together, the unadjusted GPG and the unexplained (adjusted) GPG with the adjustment factors can provide policy makers with better indicators to analyse the underlying causes of the differences recorded between the average earnings of men and women.

The members of DSS are invited:

- **To take note of the decomposition method of the unadjusted GPG and the results presented;**
- **To take note that Eurostat intends to publish the decomposition methodology and results as 'Statistical working paper';**
- **To discuss about a possible dissemination of the explained GPG with its components presented in annex 3 as experimental statistics.**

ANNEX 1: Methodology and results

1. Introduction

Reducing the gender pay gap is one of the key priorities of gender policies at the EU and national levels. At the EU level, the European Commission prioritised "reducing the gender pay, earnings and pension gaps and thus fighting poverty among women" as one of the key areas in its "Strategic engagement for gender equality 2016-2019". One of the European Commission's actions in this area is to continue to mark the European Equal Pay Day (in early November) with Europe-wide information activities to raise awareness of EU citizens on possible discriminations between men and women. The unadjusted gender pay gap (unadjusted GPG) gets high visibility in this context as the key indicator to monitor progress in the field of equal earnings for men and women. The importance of the unadjusted GPG indicator is now reinforced by its inclusion in the Indicators Scoreboard of the European Pillar of Social Rights.

2. The unadjusted GPG

The unadjusted GPG is published annually by Eurostat in cooperation with the National Statistical Institutes (NSIs) of the EU Member States and the EFTA countries. It is based on the methodology of the Structure of Earnings Survey (SES) which is carried out with a four-yearly periodicity and on national sources for the years between the SES years (from reference year 2007 onwards). The scope and coverage of the unadjusted GPG are as follows: NACE sections B to S without O, enterprises with 10 employees or more, no restrictions for age and hours worked, and part-timers included. The definition of the unadjusted GPG, expressed in percentage, is as follows:

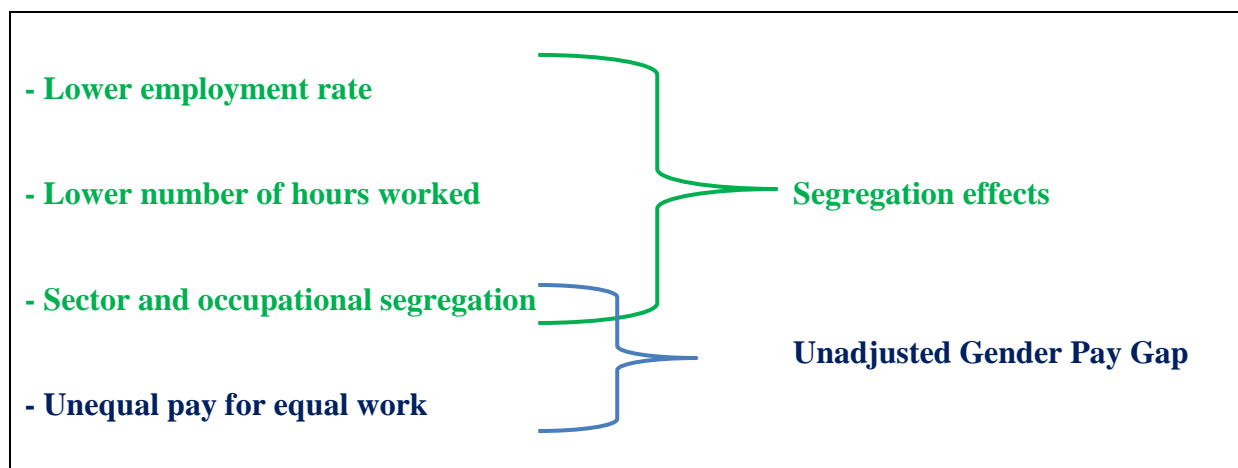
$$\frac{\text{Mean hourly earnings of men} - \text{Mean hourly earnings of women}}{\text{Mean hourly earnings of men}}$$

As an unadjusted indicator, the GPG gives an overall picture of the differences between men and women in terms of pay and measures a concept which is broader than the concept of equal pay for equal work. A part of the difference in earnings of men and women can be explained, among others, by sectoral and occupational gender segregations. Sectoral gender segregation means that women tend to be clustered in low paying sectors whereas men in high pay sectors. Occupational gender segregation is a result of men being more often promoted to supervisory and management positions compared to women. As a consequence, the unadjusted gender pay gap is rather a complex indicator which entangles in its measurement both the possible discrimination between men and women in the terms of "unequal pay for equal work" as well as the consequences of sectoral and occupational gender segregations in the labour market on the earnings of women compared with men.

It must be noted that the unadjusted GPG does not capture all segregation effects between men and women in the labour market. In particular, a higher proportion of women is 'inactive' (i.e. not seeking or not available to work) as illustrated by a lower employment rate. Moreover, the average number of hours worked by (active) women is lower than for men, which is not captured by the GPG which is calculated on an hourly basis.

Diagram 1 summarizes the main effects that play a role on the overall gap between the average earnings of all women of working age compared with men:

Diagram 1: decomposition of the 'overall' gender earnings gap



In view of this, the European Commission's DG for Justice and Consumers expressed a strong interest in the decomposition of the unadjusted GPG into the share which is due to composition effects (different jobs held by men and women) and the remaining part (unexplained GPG).

Against this backdrop, Eurostat developed a methodology to adjust the gender pay gap which was discussed at the LAMAS.

3. The unexplained (adjusted GPG) and adjustment factors

Eurostat used the Oaxaca decomposition (also called Blinder-Oaxaca decomposition) to estimate the unexplained (adjusted) GPG with the Structure of Earnings Survey (SES) as the data source. The main part of the document will only outline this method. More detailed information can be found in the annex.

The Oaxaca decomposition is carried out in two stages which are a regression analysis and a decomposition analysis. The regression analysis is carried out separately for the earnings structure of men and the earnings structure of women. The regression equations provide insights into the male and female earnings structures by showing the relationship between log hourly earnings and characteristics (observed personal, job and enterprise characteristics) for men and women, separately. Compared to the presentation made at the LAMAS and DSS Board meetings, the method was slightly modified by using ISCO 2-digit level for occupation in the regression analysis.

In a second step, the difference between the means of log hourly earnings of men and women is decomposed into two parts: an unexplained part and an explained part. Within the decomposition approach, it must be decided which earnings structure constitutes the non-discriminatory benchmark against which to decompose the difference between the means of hourly earnings of men and women. We assume, in accordance with the definition of the unadjusted gender pay gap, that the male earnings structure constitutes this benchmark.

The final decomposition equation for the difference between the means of log hourly earnings of men (M) and women (W) is as follows:

$$\Delta = \overline{\ln y^M} - \overline{\ln y^W} = \underbrace{(\hat{\beta}_0^M - \hat{\beta}_0^W) + \sum_{k=1}^K \bar{x}_k^W (\hat{\beta}_k^M - \hat{\beta}_k^W)}_{\text{Unexplained part of } \Delta} + \underbrace{\sum_{k=1}^K \hat{\beta}_k^M (\bar{x}_k^M - \bar{x}_k^W)}_{\text{Explained part of } \Delta}$$

where:

- $\overline{\ln y}$ represents the natural log of mean of hourly earnings of men (M) and women (W),
- \bar{x}_k , from $k=1$ to $k=K$, are the means of the variables covering the observed personal, job and enterprise characteristics for men (M) and women (W),
- β_0 is a constant and β_k , from $k=1$ to $k=K$, are the parameters for the corresponding variables covering the observed characteristics for men (M) and women (W).

The first part of the equation, which is marked as *Unexplained*, measures the difference between financial returns to men and women. In other words, we calculate what a female worker with the average characteristics would have earned if she had been treated in the same way as a typical male worker, and compare these earnings with what she actually earns. A possible interpretation of this part as discrimination requires some caution as other explanatory variables not included in the regression (e.g. because they are unobserved), such as career breaks, may change the results of the decomposition.

The second part of the equation, which is marked as *Explained*, measures the part of the difference between the means of log hourly earnings of men and women which is due to the differences in average characteristics for men and women

Each of the components, *Explained* and *Unexplained*, can be expressed as a proportion of the overall difference Δ . We can also express each subcomponent of *Explained* as a proportion of the overall difference Δ . This allows to estimate a magnitude of the effects of the specific characteristics explaining the overall difference Δ . Those subcomponents can be called adjustment factors.

Finally, the unexplained (adjusted) GPG can be calculated by applying a percentage of the overall unexplained part to the unadjusted GPG.

4. Results and analysis

The 2014 data of the Structure of Earnings Survey (SES) are used in the decomposition and estimations of the unexplained (adjusted) GPG. Table 1 shows the overall explained part (*Explained*) and its subcomponents (adjustment factors) as well as the unexplained part (*Unexplained*) expressed as a percentage of the difference between log hourly earnings of men and women (Δ). The positive explained part means that men had, on average, better characteristics compared to women in the labour market, whereas the negative explained part means that women had, on average, better characteristics compared to men in the labour market. In the latter case, the unexplained (adjusted) GPG is higher than the unadjusted figures.

Table 1: The overall explained part with its components (adjustment factors) and the unexplained part expressed as a percentage of the difference between log hourly earnings of men and women, 2014*

	Overall explained part	Explained part									Unexplained part
		Personal and job characteristics					Enterprise characteristics				
		Age	Education	Occupation	Job experience	Employment contract	Working time	Economic activity	Enterprise size	Enterprise control	
EU28	31	0	-7	-3	1	-1	13	32	-5	0	69
Belgium	63	1	20	-7	2	0	20	30	-6	2	37
Bulgaria	-30	-3	-32	-15	-17	1	1	15	12	10	130
Czech Republic	17	-3	0	-2	0	1	1	21	0	-1	83
Denmark	42	0	-3	6	-1	-3	-1	30	-5	20	58
Germany	65	-1	4	6	2	0	30	26	-3	0	35
Estonia	29	3	-6	10	-2	0	3	24	-2	-1	71
Ireland	-14	2	-4	8	5	0	1	-7	-14	-5	114
Greece	32	10	-14	-11	-1	0	12	36	-2	1	68
Spain	27	2	-8	3	7	1	4	27	-9	-1	73
France	31	1	-6	-10	1	0	-2	51	-3	-1	69
Croatia	-94	-11	-92	-14	-9	-1	-4	56	-1	-17	194
Italy	-99	-16	-51	-153	5	-2	58	106	-15	-31	199
Cyprus	14	-5	-5	29	-7	0	2	11	-11	0	86
Latvia	6	5	-31	5	-16	1	0	69	-25	-3	94
Lithuania	-82	-1	-29	-71	-31	0	6	69	-25	0	182
Luxembourg	-55	32	-15	-20	8	-3	13	0	-30	-41	155
Hungary	-4	-5	-27	-26	-8	1	-21	82	-21	22	104
Malta	-3	14	-18	11	9	0	3	-3	-17	-1	103
Netherlands	47	11	-2	14	3	0	19	-29	-6	37	53
Austria	58	0	1	15	8	-5	17	22	-1	0	42
Poland	-118	-15	-87	-105	-12	-3	3	95	21	-16	218
Portugal	11	1	-32	14	1	-1	-1	40	-10	-2	89
Romania	-283	-1	-149	-182	-43	0	10	217	-118	-17	383
Slovenia	-121	-10	-87	-102	-25	-9	2	137	-6	-23	221
Slovakia	9	-3	-5	-11	-3	0	4	22	1	4	91
Finland	43	-2	-3	14	1	1	-1	32	-5	7	57
Sweden	49	-2	-9	-6	-1	:	2	58	-9	15	51
United Kingdom	37	0	0	21	1	-1	4	21	-4	-5	63
Iceland	35	-7	-14	-34	-4	-1	10	40	5	39	65
Norway	43	0	-8	-15	1	0	9	58	-5	3	57
Switzerland	34	4	11	5	4	0	-5	17	-1	1	66

* The figures for age are the sums of the results for the "age" and "age squared" variables; the figures of age are the sums of the results for the "job experience" and "job experience squared" variables; for Romania, employees whose weights belong to 5th highest percentile are excluded.

At the EU level, the overall explained part - calculated as the weighted average¹ of the explained parts in EU Member States - is 31%. This means that 31% of the difference between log hourly earnings of men and women can be attributed to better, on average, characteristics of men compared to women in the European labour markets. The EU explained part is mostly driven by economic activity and working time which attribute 32% and 13%, respectively, to the difference between log hourly earnings of men and women. Job experience explains only 1% of the difference whereas age and enterprise control have no explanatory effect. The positive adjustment factors are partially cancelled out by the negative adjustment factors for education (-7%), enterprise size (-5%), occupation (-3%) and employment contract (-1%). At the EU level, the effect of occupation is rather smaller than expected due to its heterogeneous effect across EU Member States as a positive or negative adjustment factor.

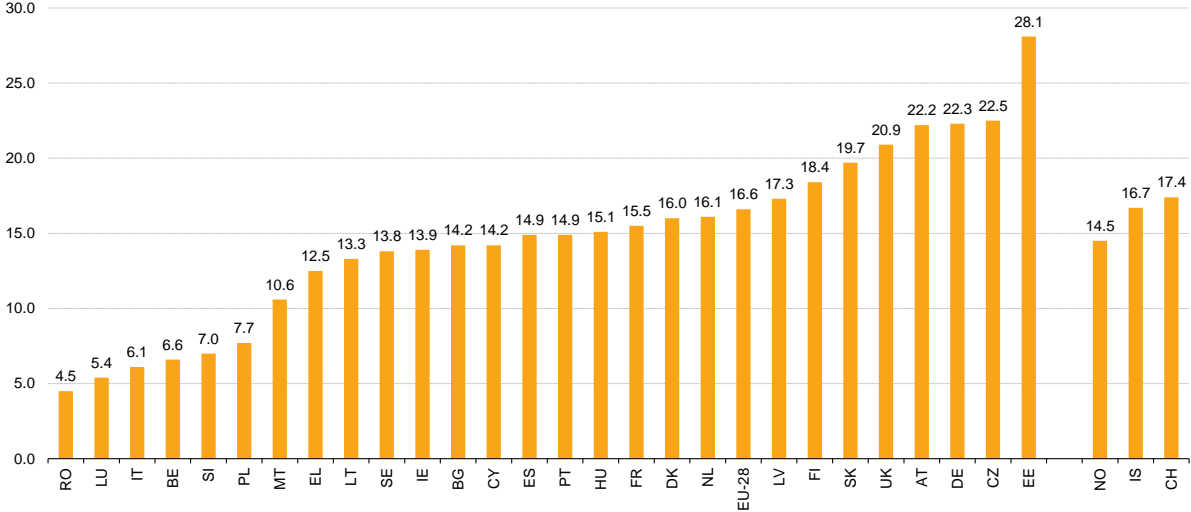
¹ A weight for each EU Member State is the number of employees in that Member States multiplied by the unadjusted gender pay gap of that Member States that is divided by the unadjusted GPG of the EU.

Across EU Member States, the overall explained part varies from -283% in Romania² to 65% in Germany.

As economic activity, occupation and enterprise size are categorical variables, the difference in average characteristics for these variables can be interpreted as gender segregation. For example, a positive adjustment factor for economic activity means that men worked, on average, in better paid NACE sections compared to women, leading to a lower unexplained (adjusted) GPG compared with the unadjusted value. Conversely, a negative component means that women worked, on average, in better paid NACE sections compared to men, leading to an upward correction of the unadjusted GPG.

Figure 2 shows the estimates of the unexplained (adjusted) GPG which are the results of applying a percentage of the overall unexplained part to the unadjusted GPG presented in Figure 1.

Figure 1: Unadjusted GPG (%), 2014



The unexplained (adjusted) GPG for the EU has been calculated using the same method as for the EU unadjusted GPG, as the average of the unexplained (adjusted) GPGs in EU Member States weighted by the corresponding numbers of employees. The EU unexplained GPG, calculated for the whole economy except public administration, is 11.5% against 16.6% for the unadjusted GPG in 2014. This means that women are expected to earn 11.5% less than men due to higher financial returns to men than women having the same average characteristics.

Across EU Member States, the unadjusted GPG varies from 4.5% in Romania to 28.1% in Estonia, whereas the unexplained (adjusted) GPG varies from 2.5% in Belgium to 24.2% in Lithuania. The unexplained GPG can be split further into the part measuring the difference between financial returns to men and women for all SES variables and the part being the residual. Eurostat, however, does not analyse further the components of the unexplained GPG due to the methodological reasons explained in annex 2 (p.17).

² For Romania, employees whose weights belong to 5th highest percentile are excluded.

Figure 2: Unexplained (adjusted) GPG (%), 2014

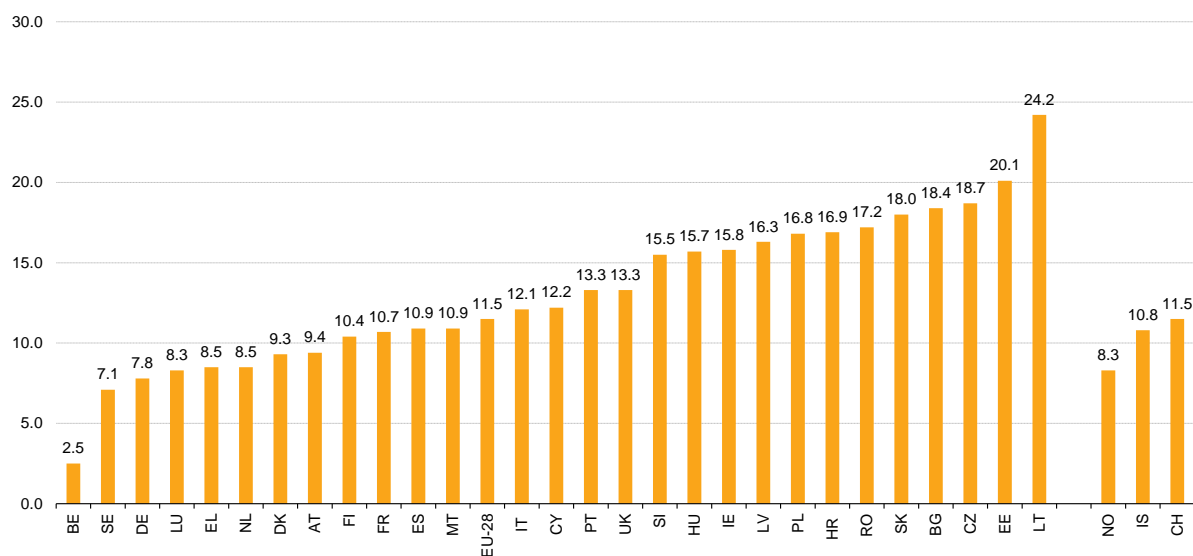


Figure 3 compares the unexplained (adjusted) and unadjusted GPGs. When comparing the ranking of EU countries (and arranged from smallest to largest GPG) before/after adjustment we observe the most significant impact for Romania, Lithuania, Croatia, Poland, Slovenia, Bulgaria and Italy moving by at least 10 positions downwards in the ranking as well as for Germany, Austria, the Netherlands Finland and Denmark moving by at least 10 positions upwards.

Figure 3: Unexplained (adjusted) and unadjusted GPGs (%), 2014

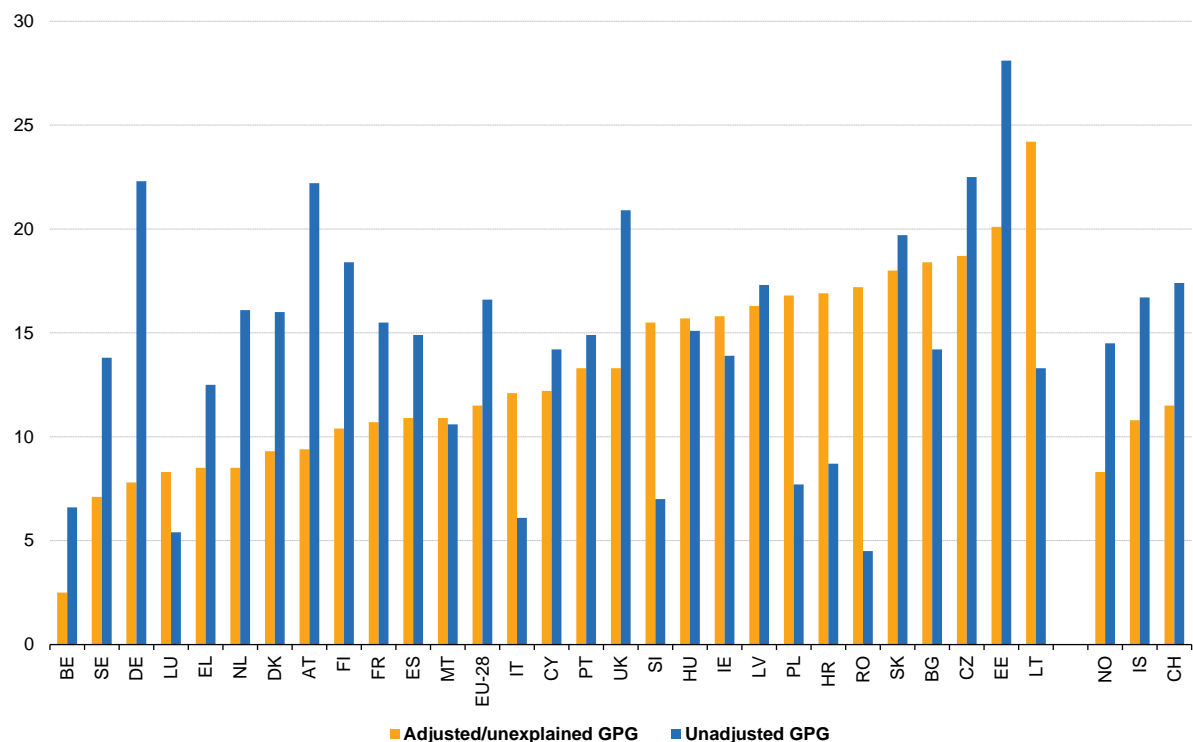


Table 2 and Figure 4 show a decomposition of the explained GPG which is the result of applying the adjustment factors displayed in Table 1 to the unadjusted GPG provided in Figure 1. At the EU level, the overall explained GPG - calculated as the weighted average of the explained gaps in EU Member States - is 5.1%. This means that women are expected to earn 5.1% less than men due to better, on average, characteristics of men compared to women in the European labour markets.

Table 2: Decomposition of the unadjusted GPG (in % of male hourly earnings), 2014

	Unadjusted GPG	Explained GPG										Unexplained/adjusted GPG
		Overall explained gap	Personal and job characteristics						Enterprise characteristics			
			Age	Education	Occupation	Job experience	Employment contract	Working time	Economic activity	Enterprise size	Enterprise control	
EU28	16.6	5.1	-0.1	-1.2	-0.4	0.1	-0.1	2.1	5.4	-0.8	0.0	11.5
Belgium	6.6	4.1	0.1	1.3	-0.5	0.1	0.0	1.4	2.0	-0.4	0.1	2.5
Bulgaria	14.2	-4.2	-0.5	-4.6	-2.2	-2.4	0.1	0.1	2.2	1.7	1.4	18.4
Czech Republic	22.5	3.8	-0.6	-0.1	-0.4	0.0	0.3	0.2	4.8	0.0	-0.3	18.7
Denmark	16	6.7	0.0	-0.5	0.9	-0.2	-0.4	-0.2	4.8	-0.9	3.2	9.3
Germany	22.3	14.5	-0.2	1.0	1.3	0.5	-0.1	6.8	5.7	-0.6	0.1	7.8
Estonia	28.1	8.0	0.7	-1.5	2.9	-0.6	0.0	0.8	6.6	-0.6	-0.2	20.1
Ireland	13.9	-1.9	0.3	-0.6	1.2	0.7	0.0	0.2	-0.9	-2.0	-0.7	15.8
Greece	12.5	4.0	1.3	-1.7	-1.3	-0.2	0.0	1.5	4.6	-0.2	0.1	8.5
Spain	14.9	4.0	0.3	-1.2	0.5	1.1	0.1	0.6	4.1	-1.3	-0.2	10.9
France	15.5	4.8	0.2	-0.9	-1.6	0.2	0.0	-0.3	7.9	-0.4	-0.2	10.7
Croatia	8.7	-8.2	-1.0	-8.0	-1.2	-0.8	-0.1	-0.4	4.9	0.0	-1.5	16.9
Italy	6.1	-6.0	-0.9	-3.1	-9.3	0.3	-0.1	3.6	6.5	-0.9	-1.9	12.1
Cyprus	14.2	2.0	-0.7	-0.7	4.1	-0.9	0.0	0.2	1.5	-1.5	0.0	12.2
Latvia	17.3	1.0	0.9	-5.3	0.9	-2.7	0.2	0.0	11.9	-4.3	-0.6	16.3
Lithuania	13.3	-10.9	-0.1	-3.8	-9.5	-4.1	-0.1	0.8	9.2	-3.4	0.0	24.2
Luxembourg	5.4	-2.9	1.8	-0.8	-1.1	0.4	-0.2	0.7	0.0	-1.6	-2.2	8.3
Hungary	15.1	-0.6	-0.8	-4.0	-3.9	-1.2	0.1	-3.2	12.4	-3.2	3.3	15.7
Malta	10.6	-0.3	1.5	-1.9	1.1	0.9	0.0	0.3	-0.4	-1.8	-0.1	10.9
Netherlands	16.1	7.6	1.8	-0.4	2.2	0.5	0.1	3.1	-4.6	-0.9	5.9	8.5
Austria	22.2	12.8	0.0	0.3	3.4	1.8	-1.2	3.8	5.0	-0.2	0.0	9.4
Poland	7.7	-9.1	-1.2	-6.7	-8.1	-0.9	-0.2	0.3	7.3	1.6	-1.2	16.8
Portugal	14.9	1.6	0.1	-4.8	2.1	0.2	-0.2	-0.1	6.0	-1.5	-0.3	13.3
Romania	4.5	-12.7	0.0	-6.7	-8.2	-1.9	0.0	0.4	9.7	-5.3	-0.8	17.2
Slovenia	7	-8.5	-0.7	-6.1	-7.2	-1.7	-0.6	0.2	9.6	-0.4	-1.6	15.5
Slovakia	19.7	1.7	-0.6	-1.0	-2.2	-0.5	0.0	0.8	4.3	0.2	0.8	18.0
Finland	18.4	8.0	-0.4	-0.6	2.5	0.1	0.3	-0.2	5.9	-0.9	1.4	10.4
Sweden	13.8	6.7	-0.2	-1.3	-0.8	-0.2	0.3	0.3	8.0	-1.2	2.1	7.1
United Kingdom	20.9	7.6	0.0	-0.1	4.4	0.3	-0.2	0.8	4.4	-0.8	-1.1	13.3
Iceland	16.7	5.9	-1.1	-2.3	-5.7	-0.6	-0.2	1.7	6.7	0.8	6.5	10.8
Norway	14.5	6.2	0.0	-1.2	-2.2	0.2	0.0	1.3	8.4	-0.8	0.4	8.3
Switzerland	17.4	5.9	0.7	1.9	0.8	0.6	-0.1	-0.9	3.0	-0.2	0.1	11.5

* The figures for age are the sums of the results for the "age" and "age squared" variables; the figures of age are the sums of the results for the "job experience" and "job experience squared" variables; for Romania, employees whose weights belong to 5th highest percentile are excluded.

The EU explained GPG is mostly driven by economic activity (5.4% gap) and working time (2.1% gap). A positive explained gap (0.1%) is also recorded for job experience. The positive explained gaps for those characteristics are cancelled out partially by the negative gaps for education (-1.2%), enterprises size (-0.8%) occupation (-0.4%), age and employment contract (both -0.1%). A negative gap means that women are expected to earn more than men due to possessing on average higher levels of a given characteristic than men. For example, a negative gap of 1.2% for education means that women are expected to earn 1.2% more than men due to possessing on average higher levels of education in the European labour markets. Finally, there is no explained gap for enterprise control.

Across the EU Member States, the overall explained GPG varies from -12.7% in Romania³³ to 14.5% in Germany. A negative gap of 12.7% in Romania means that women are expected to earn 12.7% more than men due to better, on average, characteristics of women compared to men in the labour market. The overall explained gap is negative in eleven Member States: Bulgaria, Croatia, Hungary, Ireland, Italy, Lithuania, Luxembourg, Malta, Poland, Romania and Slovenia, and positive in seventeen Member States.

The overall explained GPG is mostly driven by three characteristics: economic activity, occupation and education. An explained gap of at least 1% (irrespective of the sign) is recorded for economic activity in twenty-five Member States, for occupation in twenty-two, and for education in eighteen. Among those three characteristics, economic activity and education have the most homogenous effect across EU Member States. This means that gaps for those characteristics are either positive in the majority of the countries or negative in the majority of the countries.

An explained gender pay gap is positive for economic activity in all EU Member States, except, Ireland, Luxembourg, Malta and the Netherlands which is the only country with an explained gap below -1%. On the other hand, in that country, the highest explained gap is recorded for enterprise control meaning that the higher share of men compared to that of women work in the private sector where earnings are higher on average than in the public sector.

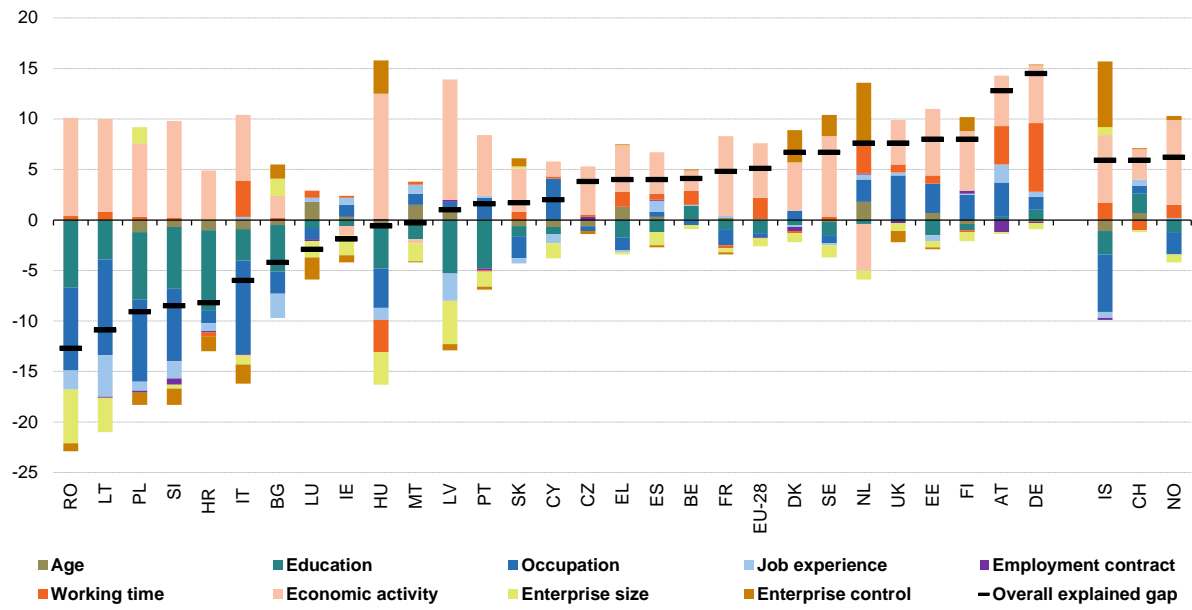
In contrary, an explained gender pay gap is negative for education in all EU Member States, except Austria, Belgium, the Czech Republic, Germany and the United Kingdom. Nevertheless, only Belgium and Germany recorded an explained gap of at least 1% for that characteristic. This means that women have a higher average level of education than men in most European labour markets.

A more mixed picture can be observed for occupation as the countries are split almost equally between those recording positive or negative gaps. The explained gap for occupation varies from -9.5% in Lithuania to 4.4% in the United Kingdom. Nevertheless quite a clear pattern can be observed for the countries with a negative overall explained gender pay gap. The countries such as Italy, Lithuania, Poland, Romania and Slovenia with an overall explained gap below -5% recorded also an explained gap for occupation which is below the same threshold. Furthermore, those countries recorded also an explained gap below -5% for education.

It can be also noted that working time plays a significant role in explaining the unadjusted gender pay gap in Germany which is the only country with an explained gap above 5%.

³³ For Romania, employees whose weights belong to 5th highest percentile are excluded.

Figure 4: Decomposition of the explained GPG (in % of male hourly earnings), 2014



ANNEX 2: Oaxaca decomposition

Oaxaca decomposition, sometimes called Blinder-Oaxaca decomposition, is carried out in two stages which are a regression analysis and a decomposition analysis of the structure of earnings. In the first stage, a regression analysis is used to estimate the earnings equations separately for men (M) and women (W) as detailed in the following equations:

$$\ln y_i^M = \beta_0^M + \sum_{k=1}^K x_{ki}^M \beta_k^M + u_i^M$$

$$\ln y_i^W = \beta_0^W + \sum_{k=1}^K x_{ki}^W \beta_k^W + u_i^W$$

where:

- $\ln y_i$ represents the natural log of hourly earnings of individual i ,
- x_{ki} , from $k=1$ to $k=K$, are variables covering the observed personal, job and enterprise characteristics that may impact on the log hourly earnings of individual i ,
- β_0 is a constant and β_k , from $k=1$ to $k=K$, are the parameters for the corresponding variables covering the observed characteristics,
- u_i is a disturbance term for observation i .

Eurostat uses the observed characteristics which are available in the SES microdata provided by the countries (see Table 1).

Table 1: Observed characteristics covered by the variables x_k

Observed characteristics	Values	Codes of the corresponding SES variables
Personal and job characteristics		
Age	Age, age squared	Variable 2.2
Education	ISCED levels 0+1+2, 3+4, 5+6, 7+8 *	Variable 2.5
Occupation	2-digit ISCO-08 occupation levels * ⁴	Variable 2.3
Job experience (in the current enterprise)	Job experience in years, job experience in years squared	Variable 2.6
Employment contract	Indefinite duration, temporary duration, apprentice *	Variable 2.8
Working time	Full time, part time *	Variable 2.7
Enterprise characteristics		
Principal economic activity	All NACE rev. 2 sections except section O *	Variable 1.3
Enterprise size	Enterprises with 10-49, 50-249, 250-499, 500-999, 1000+ employees *	Variable 1.2
Enterprise control	Public, private *	Variable 1.4

* For the categorical characteristics, all categories without one (omitted category) are expressed as dummy variables.

⁴ As suggested by some Member States, Eurostat used ISCO-08 at 2 digits instead of 1-digit level for the "occupation" variable.

After fitting separate regressions for men and women the following equalities hold for either male or female individual i :

$$\widehat{\ln y}_i^M = \hat{\beta}_0^M + \sum_{k=1}^K x_{ki}^M \hat{\beta}_k^M$$

$$\widehat{\ln y}_i^W = \hat{\beta}_0^W + \sum_{k=1}^K x_{ki}^W \hat{\beta}_k^W$$

These regression equations provide insights into the male and female earnings structures by showing the relationship between log hourly earnings and characteristics (observed personal, job and enterprise characteristics) for men and women. In this relationship, the estimated constants and coefficients measure the financial returns to male or female characteristics. Figure 1 shows this relationship for the bivariate regression with one explanatory variable.

In the second stage, we carry out a decomposition analysis of the difference between the means of log hourly earnings of men and women:

$$\Delta = \overline{\ln y}^M - \overline{\ln y}^W$$

The Oaxaca decomposition uses the following regression property for the means of log hourly earnings of men and women:

$$\overline{\ln y}^M = \hat{\beta}_0^M + \sum_{k=1}^K \bar{x}_k^M \hat{\beta}_k^M$$

$$\overline{\ln y}^W = \hat{\beta}_0^W + \sum_{k=1}^K \bar{x}_k^W \hat{\beta}_k^W$$

Within the decomposition approach, it must be decided which earnings structure constitutes the non-discriminatory benchmark against which to decompose the difference, Δ , between the means of log hourly earnings of men and women. We assume, in accordance with the definition of the unadjusted gender pay gap, that the male earnings structure constitutes this benchmark⁵. The constant and the coefficients in the men's equation are treated as the non-discriminatory benchmarks for the financial returns to characteristics of employees. In view of this, a *counterfactual* equation is constructed where the constant and coefficients in the women's equation are replaced by those of the men's equation:

$$\overline{\ln y}^{W*} = \hat{\beta}_0^M + \sum_{k=1}^K \bar{x}_k^W \hat{\beta}_k^M$$

This equation can be interpreted as what the average female worker would have earned if she had been paid on the same basis as an equivalent male worker (see also Figure 1 showing the Oaxaca decomposition for a single explanatory variable). The difference between the means of log hourly earnings of men and women can, then, be decomposed as follows:

$$\Delta = U + E$$

⁵ The following other options are possible:

- the female earnings structure constitutes a non-discriminatory benchmark,
- both the female and male earnings structures constitute non-discriminatory benchmarks with some weighted average applied ,
- the whole population earnings structure constitutes a non-discriminatory benchmark.

where $U = \overline{\ln y}^{W*} - \overline{\ln y}^W$ and $E = \overline{\ln y}^M - \overline{\ln y}^{W*}$

The first part of the equation, which is marked as U , is the difference between the *counterfactual* and actual means of log hourly earnings of women. We calculate what a female worker with the average characteristics would have earned if she had been treated in the same way as a typical male worker, and compare these earnings with what she actually earns. After manipulation, this comparison, $\overline{\ln y}^{W*} - \overline{\ln y}^W$, can be expressed as:

$$U = (\hat{\beta}_0^M - \hat{\beta}_0^W) + \sum_{k=1}^K \bar{x}_k^W (\hat{\beta}_k^M - \hat{\beta}_k^W)$$

It measures the part of Δ which is due to the differences in the estimated constants plus the difference in the coefficients for men and women weighted by the averages of female characteristics. The difference between the constants and the difference between the coefficients measure the difference between financial returns to men and women. Figure 1 shows this difference for the bivariate regression with one variable. In the decomposition, the part U can be interpreted as the "unexplained" part of the difference in earnings between men and women (Δ). This "unexplained" part could be considered as a better approximation of potential discrimination, i.e., how the labour market rewards men and women in an unequal way for the same work. Nevertheless, the interpretation of the U part as discrimination requires some caution as other explanatory variables not included in the regression (e.g. because they are unobserved) may change the results of the decomposition. This is why the term "better approximation" is used.

The second part of the equation, which is marked as E , is the difference between the actual mean of the log hourly earnings of men and the *counterfactual* mean of the log hourly earnings of women. This can be expressed as:

$$E = \sum_{k=1}^K \hat{\beta}_k^M (\bar{x}_k^M - \bar{x}_k^W)$$

It measures the part of Δ which is due to the differences in average characteristics of men and women weighted by the male coefficients. This represents the "explained" part (E) of the difference in earnings between men and women (Δ). Figure 1 shows this "explained" part for the bivariate regression with one explanatory variable.

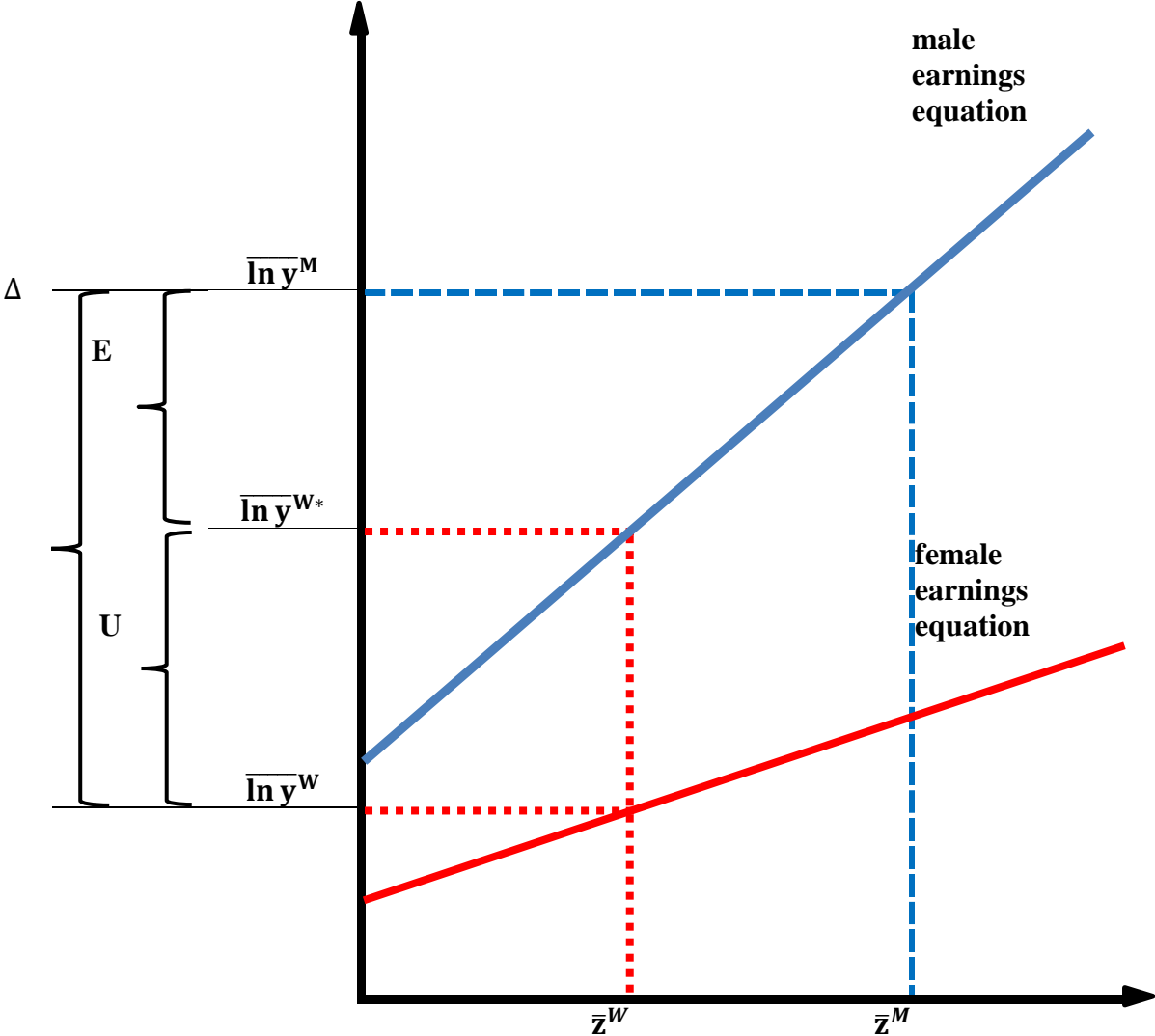
The final decomposition equation for the difference between the means of log hourly earnings of men and women is as follows:

$$\overline{\ln y}^M - \overline{\ln y}^W = \underbrace{(\hat{\beta}_0^M - \hat{\beta}_0^W) + \sum_{k=1}^K \bar{x}_k^W (\hat{\beta}_k^M - \hat{\beta}_k^W)}_{\text{Unexplained}} + \underbrace{\sum_{k=1}^K \hat{\beta}_k^M (\bar{x}_k^M - \bar{x}_k^W)}_{\text{Explained}}$$

Each of the components, E and U , can be expressed as a proportion of the overall difference Δ . We can also express each subcomponent of E , i.e., $E_k = \hat{\beta}_k^M (\bar{x}_k^M - \bar{x}_k^W)$ as a proportion of the overall difference Δ . This allows to estimate a magnitude of the effects of the specific characteristics explaining the overall difference Δ . However, the interpretation of the estimated coefficients can be misleading, when the explanatory variables are categorical with more than two categories. This is due to the coefficient estimates for the categories of a categorical variable being sensitive to a choice of the reference category. In the explained part

E , nevertheless, the overall effect of the categorical variable (e.g., the overall effect of economic activity) in the decomposition does not depend on the omitted category. It is, however, not the case for the unexplained part U where the overall effect of a categorical variable depends on the omitted category. For this reasons only the overall results for the unexplained part U can be taken into account and the results of the detailed decomposition for the unexplained component will not be interpreted.

*Figure 1: The Oaxaca decomposition for a single explanatory variable**



* Bivariate regression: $\overline{\ln y}_i = \alpha_0 + \alpha_1 z_i + v_i$

ANNEX 3: Decomposition of the explained gender pay gap (in % of male hourly earnings), 2014

	Overall explained gap	Personal and job characteristics						Enterprise characteristics		
		Age	Education	Occupation	Job experience	Employment contract	Working time	Economic activity	Enterprise size	Enterprise control
EU28	5.1	-0.1	-1.2	-0.4	0.1	-0.1	2.1	5.4	-0.8	0.0
Belgium	4.1	0.1	1.3	-0.5	0.1	0.0	1.4	2.0	-0.4	0.1
Bulgaria	-4.2	-0.5	-4.6	-2.2	-2.4	0.1	0.1	2.2	1.7	1.4
Czech Republic	3.8	-0.6	-0.1	-0.4	0.0	0.3	0.2	4.8	0.0	-0.3
Denmark	6.7	0.0	-0.5	0.9	-0.2	-0.4	-0.2	4.8	-0.9	3.2
Germany	14.5	-0.2	1.0	1.3	0.5	-0.1	6.8	5.7	-0.6	0.1
Estonia	8.0	0.7	-1.5	2.9	-0.6	0.0	0.8	6.6	-0.6	-0.2
Ireland	-1.9	0.3	-0.6	1.2	0.7	0.0	0.2	-0.9	-2.0	-0.7
Greece	4.0	1.3	-1.7	-1.3	-0.2	0.0	1.5	4.6	-0.2	0.1
Spain	4.0	0.3	-1.2	0.5	1.1	0.1	0.6	4.1	-1.3	-0.2
France	4.8	0.2	-0.9	-1.6	0.2	0.0	-0.3	7.9	-0.4	-0.2
Croatia	-8.2	-1.0	-8.0	-1.2	-0.8	-0.1	-0.4	4.9	0.0	-1.5
Italy	-6.0	-0.9	-3.1	-9.3	0.3	-0.1	3.6	6.5	-0.9	-1.9
Cyprus	2.0	-0.7	-0.7	4.1	-0.9	0.0	0.2	1.5	-1.5	0.0
Latvia	1.0	0.9	-5.3	0.9	-2.7	0.2	0.0	11.9	-4.3	-0.6
Lithuania	-10.9	-0.1	-3.8	-9.5	-4.1	-0.1	0.8	9.2	-3.4	0.0
Luxembourg	-2.9	1.8	-0.8	-1.1	0.4	-0.2	0.7	0.0	-1.6	-2.2
Hungary	-0.6	-0.8	-4.0	-3.9	-1.2	0.1	-3.2	12.4	-3.2	3.3
Malta	-0.3	1.5	-1.9	1.1	0.9	0.0	0.3	-0.4	-1.8	-0.1
Netherlands	7.6	1.8	-0.4	2.2	0.5	0.1	3.1	-4.6	-0.9	5.9
Austria	12.8	0.0	0.3	3.4	1.8	-1.2	3.8	5.0	-0.2	0.0
Poland	-9.1	-1.2	-6.7	-8.1	-0.9	-0.2	0.3	7.3	1.6	-1.2
Portugal	1.6	0.1	-4.8	2.1	0.2	-0.2	-0.1	6.0	-1.5	-0.3
Romania	-12.7	0.0	-6.7	-8.2	-1.9	0.0	0.4	9.7	-5.3	-0.8
Slovenia	-8.5	-0.7	-6.1	-7.2	-1.7	-0.6	0.2	9.6	-0.4	-1.6
Slovakia	1.7	-0.6	-1.0	-2.2	-0.5	0.0	0.8	4.3	0.2	0.8
Finland	8.0	-0.4	-0.6	2.5	0.1	0.3	-0.2	5.9	-0.9	1.4
Sweden	6.7	-0.2	-1.3	-0.8	-0.2		0.3	8.0	-1.2	2.1
United Kingdom	7.6	0.0	-0.1	4.4	0.3	-0.2	0.8	4.4	-0.8	-1.1
Iceland	5.9	-1.1	-2.3	-5.7	-0.6	-0.2	1.7	6.7	0.8	6.5
Norway	6.2	0.0	-1.2	-2.2	0.2	0.0	1.3	8.4	-0.8	0.4
Switzerland	5.9	0.7	1.9	0.8	0.6	-0.1	-0.9	3.0	-0.2	0.1