Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Federal Department of Home Affairs FDHA
**Federal Office for Gender Equality FOGE**

# Standard analysis model for monitoring compliance with wage equality between women and men in federal procurement (methodology)

Instructions for standardised monitoring procedure

Bern, October 2017

## List of contents



## 1 General situation

Tendering companies are bound by the Swiss law on public procurement to comply with the principle of wage equality between women and men, thus conforming to Swiss legislation and international agreements. The aim is also to prevent distortions of competition which disadvantage employers who implement a fair wage policy. The Federal Office for Gender Equality (FOGE) is mandated by the Federal Procurement Conference (FPC) to carry out controls.

The standard analysis model describes the Confederation’s procedure for controlling compliance with wage equality regulations in companies with a minimum of 50 employees. It consists of two components:

a) a statistical methodology: semi-logarithmic OLS regression analysis;
b) a model with the following explanatory characteristics: years of training, potential years of employment, years of service, required level of professional skills, professional position and sex.

A tool for monitoring compliance with this provision was developed and tested during a pilot phase (2001-2003) by the Office for Research on Labour and Social Policies (BASS) on behalf of the FOGE and the Federal Procurement Conference (FPC).[1]

The Logib tool enables regression analysis to be performed in Excel, with the specifications of the standard analysis model, making it easy for companies to access and carry out their own tests free of charge.

[1] The report on the pilot phase presents the legal background, the statistical methodology and the initial experiences gained with the tool when it was used to conduct controls between 2001 and 2003 in five pilot companies. It is available on the FOGE website in both German and French : *Contrôle du respect de l’égalité de salaire entre femmes et hommes dans les marchés publics de la Confédération. Rapport sur la phase-pilote portant sur la mise en œuvre de l’art. 8, al. 1, let. c de la loi sur les marchés publics, juin 2004.* (Monitoring compliance with equal pay between women and men in public procurement. Report on the pilot phase bearing on the implementation of Art. 8, para.1, letter c of the Public Procurement Act, June 2004). The tool has undergone various modifications since the report’s publication.

The present document describes the current situation regarding the methodological approach and serves as a guide for conducting the *standardised analysis* for monitoring compliance with wage equality.[2]

[2] The first version of this document was written in 2005 by Silvia Strub from the Office for Research on Labour and Social Policies (BASS).

# 2 Overview of the measurement methodology

## 2.1 A step by step measurement procedure using regression analysis

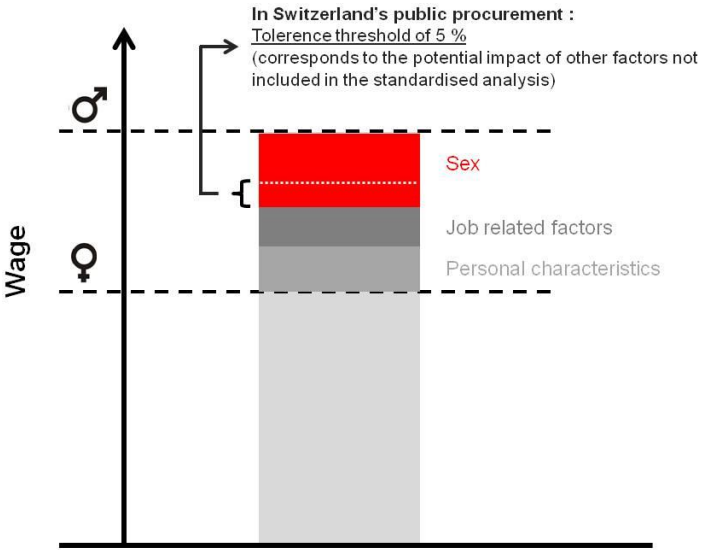
The monitoring tool uses regression analysis methodology.

*Regression analysis* is usually used to evaluate discrimination in the economy in general. It is also possible to apply this method to the companies themselves with regard to certain aspects such as the size of the company and the male-female ratio. Regression analysis enables to measure the individual impact of various factors on wages.

The *standardised* analysis used for monitoring compliance with wage equality in public procurement consists in explaining the wages of employees by their *personal characteristics* (training, years of service and potential professional experience), *job related factors* (required level of professional skills and professional position) as well as their *sex* (see Graph 1 below). All other things being equal, this method allows isolation of the share of the wage variance that is due solely to an employee's sex. In other words, it makes it possible to determine the wage gap between men and women with similar personal and professional characteristics within a company.

Given that standard regression used in the monitoring of public procurement only takes into account the five factors listed above, a tolerance threshold of 5% has been introduced to cover any share of the gap that could be explained by other objective explanatory factors that could be specific to the company in question.

The tolerance threshold of 5% is used only for the monitoring of public procurement. In other words, if after taking into account the five factors (training, years of service, potential professional experience, required level of professional skills and professional position), the gender-specific wage gap, i.e. the difference explained by gender, is statistically significantly greater than the tolerance threshold of 5%, the public procurement requirements to obtain a mandate from the Swiss Confederation are considered as not being fulfilled.




Graph 1: Monitoring of wage equality by standardised regression analysis

When gender-based wage inequality is significantly above the tolerance threshold of 5%, the monitored company has to undertake measures to correct this discrimination. Then, the company must prove, by using the method described here, that it henceforth complies with the conditions for public procurement regarding wage equality.

Given that only systematic gender-based wage discriminations are taken into consideration in Federal procurement controls and that a tolerance threshold of 5% is used, the results of the monitored company cannot be used as proof that there is no wage discrimination in general under the terms of the Gender Equality Act EqA.

## 2.2 Database

The *standardised control* requires, for all employees[3] of the company, information in anonymous form concerning (at least) the following characteristics:

- sex
- age
- training
- years of service
- required level of professional skills
- professional position
- individual working hours
- gross wage[4] (basic wage, 13th salary, allowances, bonuses and premiums)
- function.

# 3 Standardised control

As part of the standardised analyses, *two equations of regression* are calculated in order to determine the size of the wage difference between women and men:

a) The wage gap between women and men when taking into account their *personal characteristics* (*regression based on characteristics related to personal qualifications only (regression pq)*, with human capital factors such as training, years of service and potential years of employment);

b) The wage gap between men and women when taking into account both their personal characteristics and job related factors (*standard regression,* which takes into account in addition to personal characteristics, required level of professional skills and professional position).

The two-stage approach – the calculation of the personal qualifications (PQ) regression followed by the standard regression which includes job related factors – enables identification of the type and extent of any wage discrimination and other gender-based inequalities. The decision regarding compliance with conditions regarding wage equality for obtaining a public tender is based on the result of the standard regression.

[3] There are a number of exceptions to this rule which are indicated in chapter 3.3.

[4] All wage components (basic wage, 13th salary, allowances, bonuses, premiums, etc.) must be entered.

In company A, women earn on average 15% less than men. Taking into account that the women have less experience and lower qualifications, the gender-specific wage gap is 10% (result of PQ regression). If we also take into account the fact that fewer women hold executive posts or posts with a high level of responsibility, the gender-specific wage gap is 3% (standard regression). In the same professional position, with the same required level of professional skills and with the same qualifications, women earn 3% less than men in company A. Given that this result is below the tolerance threshold of 5% applied to public procurement, company A fulfils the conditions with regard to wage equality for obtaining a public tender.

## 3.1 Methodology

In principle there are two possible methodological approaches for determining the extent of discrimination between two groups (in this case, women and men) by means of regression analysis: the dummy variable method and the Oaxaca method. In the controls on compliance with wage equality in public procurement, the dummy variable is used. The following paragraphs present the methodological details enabling the standard regression analysis to be carried out on the basis of this approach.

## 3.2 Regression equation specification

The *general form* of the regression equation is

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_n X_n + \varepsilon_i$$

$Y_i$ = dependent variable (explained variable) for a given observation (e.g. wage of the person *i*);

$X_1$ à $X_n$ = independent variables (explanatory variable) for a given observation (e. g. human capital characteristics of the person *i*);

$\beta_0$ = point of intersection of the regression line with ordinate axis (intercept), i.e. the constant.

$\beta_1$ à $\beta_n$ = coefficients characterising the regression line. The value of each coefficient indicates the number of additional units of $Y_i$ associated with an increase of one unit of the independent variable under consideration.

$\varepsilon_i$ = error term for a given observation. The error term indicates the discrepancy between a predicted value and its real value (e.g. difference between the actual wage and the estimated wage of the person *i*).

The regression line closest to the data is estimated in such a way that the total of all error terms is minimised (OLS regression, ordinary least squares methodology).

The actual specifications for the regression based on characteristics related to personal qualifications only (PQ regression) and the standard regression are presented below.

**Regression equations for standardised analyses**

**a) Regression based on characteristics related to personal qualifications only**

$$ln\,(Y_i) = \beta_0 + \beta_1 * EDUC_i + \beta_2 * EXP_i + \beta_3 * EXP_i\,\hat{}\,2 + \beta_4 * ANC_i + \beta_5 * SEXE_i + \varepsilon_i$$

where $ln\,(Y_i)$ = logarithm of the person *i*'s gross wage standardised to a full-time job.

*The dependant variable "wage" is converted into a logarithm due to a right-skewed wage distribution and in order to obtain a more direct interpretation of the results (=> impact of EDUC in percentage, etc.).*[5]

[5] Conversion of standardised gross wage: the various wage components are converted for all employees into a monthly basis for a full-time job at 40 hours per week (even if the company's usual workweek is more or less than 40 hours).

$EDUC_i$ = characteristic of level of education (number of years of training of the person *i*).
*The highest qualification is converted into years of training.*[6]

$EXP_i$ = characteristic of level of professional experience (number of potential years of employment[7] of the person *i*).
$\beta_2 * EXP_i + \beta_3 * EXP_i \text{ ^2}$: *the evolution of the impact of professional experience is usually non-linear.*

$ANC_i$ = characteristic of level of duration of employment in the present company (number of years of service of the person *i*).

$SEXE_i$ = dummy variable for the sex
*is = 1, if the person i is a woman;*
*is = 0, if the person i is a man.*

$\beta_0$ = constant (corresponds to "basic wage" if the characteristics of level are equal to 0).

$\beta_1, ..., \beta_4$ = coefficients (indicate the impact of corresponding factors on wages).

$\beta_5$ = coefficient of the variable *SEX*; in other words this is a « discrimination coefficient».
*E.g.:* ***$\beta_5$ = - 0.1 means: with equal qualifications (taking into account years of service, training and potential years of employment), women earn ~10% less than men.***[8]

**b) Standard regression**

In the standard regression, in addition to personal characteristics, job related factors[9] are taken into account. These are ordinal variables with the following categories:

- For the variable "required level of professional skills", four categories of tasks have been defined:
  1 Complex problem solving and decision making
  2 Relatively complex technical or practical skills
  3 Tasks requiring special professional knowledge
  4 Simple and repetitive tasks
- For the variable "Professional position": five categories:
  1 Senior and higher management
  2 Middle management
  3 Lower management
  4 Lowest management
  5 No management function.

In order to include these two variables in the standard regression, a dummy variable is created (variable 0/1) for each category.

[6] Conversion of completed training into years: university = 17 years, university of applied sciences (HES) = 15 years, higher vocational training and education = 14 years, teaching certificate = 15 years, Matura = 13 years, vocational training and education = 12 years, in-house vocational training = 11 years, compulsory schooling = 7 years. The category "Other training" is given a value of 11.

[7] Conversion: potential experience = age minus number of years of training minus six years prior to schooling.

[8] In the semi-logarithmic equations, the coefficients of a dummy variable can be interpreted approximately as an impact expressed in percentage of the explanatory factor on the dependent variable. However, this interpretation becomes less precise as the coefficient increases. The exact formula for the calculation is as follows: $100*(e^{\beta} - 1)$.

[9] These are factors which constitute the level of qualifications required for the job and the professional position as conceived and used originally in the Swiss Earnings Structure Survey (ESS)).

The standard regression model is, therefore, based on the QP regression, but contains in addition dummy variables related to the professional position and the required level of professional skills. The discrimination coefficient is calculated in a similar manner to the QP regression.

## 3.3 Preparation of data and data plausibility check

The data supplied by the companies are prepared according to the regression equation specifications. All employees who were employed by the company during the reference month and received wages must be included in the analysis, with the exception of the following: apprentices and interns, who due to their training status received lower wages, as well as employees with permanent residence abroad (expats).

Before proceeding with the regression analysis, a number of tests must be carried out in the data plausibility check.

## 3.4 Hypothesis test: is the "discrimination coefficient" statistically significantly higher than the tolerance threshold?

Any estimation of a parameter contains an element of unpredictability which may arise from various sources (for example data errors). The aim is, therefore, to determine if the discrimination coefficient, indicating the part of the gender wage gap that is unexplained (as an absolute value), is statistically significantly higher than the tolerance threshold S.

A unilateral hypothesis test with a confidence level of 95% is carried out. The hypothesis to be tested (null hypothesis $H_0$) is:

- $H_0$: the discrimination coefficient is equal to the tolerance threshold S.

If the hypothesis $H_0$ is not confirmed, it is rejected in favour of the alternative hypothesis ($H_1$):

- $H_1$: the discrimination coefficient is greater than the tolerance threshold S.

In order to be able to test the hypothesis $H_0$, a student test (or *t*-test) is carried out on the basis of the results from the regression equation and is compared with a "critical value t-$t_{krit}$ " .[10] If the statistic *t*-test is greater than the critical value t-$t_{krit}$ (as an absolute value), the data refute the null hypothesis. In other words, this means that the wage discrimination against one sex is statistically significantly greater than 5 per cent (in absolute terms). In this case, the conditions with regard to wage equality are considered unfulfilled for obtaining a public tender from the Confederation.

In terms of formula, the student test statistic (*t*-test) is the quotient of

- the difference between the discrimination coefficient D ($\beta_5$) and the tolerance threshold S

with

- the standard error (shortened to S.E.)

The test statistic is thus expressed as **T = (D – S)/S.E.** If D < -0.05, the tolerance threshold S = -0.05 *(negative value*) should be included in the calculation. If D > 0.05, S = 0.05 *(positive value) should be included).*[11] [12]

[10] In the estimated model, the quantity to be tested, the "critical value t-$t_{krit}$", depends on the number of observations N (e. g. number of persons in the company) minus the number of coefficients. Tables with *t* values can be found in any statistics manual.

[11] If the discrimination coefficient is negative (D<0), this means that the gender-specific wage inequality disadvantages women (in other words, when all other elements are equal, women earn less than men), whereas if D is greater than 0 (D>0), this means that men are disadvantaged by the gender-specific inequality.

[12] The exact number to use, due to the conversion calculation explained at note 8 is 0.05127 if D is greater than 0 and -0.04877 if D is less than 0.

The value of the *t*-test is then compared with $t_{krit}$ (equal to 95% quantile – as it is a unilateral test – of the student test distribution). Hypothesis $H_0$ is rejected when $t > t_{krit}$ when D is greater than 0.05, or if $t < t_{krit}$ when D is less than -0.05.[13]

With the dummy variable method, it is simple to assess whether discrimination is significant or not, because the discrimination coefficient D of $\beta_5$ and the standard error (S.E.) are the direct result of the regression equation.

**Example:**

Test to determine if the discrimination effect is statistically significantly greater than the tolerance threshold (e.g. Company X with 105 employees)

**Standard regression:**

| | |
|---|---|
| Discrimination coefficient D ($\beta_5$) | - 0.138 |
| Standard error of discrimination effect (S.E. de $\beta_5$) | 0.043 |
| Tolerance threshold S | - 0.05 |
| *t*-test statistic *t* = (D–S) / S.E. | - 2.049 |
| Critical value $t_{krit.}$ ($\alpha$ = 0.05) | - 1.66 |

*Decision:*

If t < $t_{krit.}$ It follows: discrimination effect D < tolerance threshold S [*or as absolute values: "D greater than S"*]

*This means:*

Significant = yes, because - 2.049 < -1.66 [*or 2.049 or greater than 1.66*]

Explanations:

The women in company X earn, with equal qualifications (training, experience, years of service, professional position, required level of professional skills) approximately 13% less than the men (discrimination coefficient D = -0.138). The *t*-test obtained from the discrimination effect, from the tolerance threshold and the standard error of D equals -2.049. This value is less than the "critical value *t* "($t_{krit}$) of -1.66 (or greater in absolute values). Therefore, the null hypothesis must be rejected, according to which the discrimination effect is not greater than the tolerance threshold of 5%. In other words: the data have clearly refuted the null hypothesis. With equal human capital levels and taking into account job related factors, the gender-specific wage gap is significantly higher than the tolerance threshold of 5% from a statistical point of view.

[13] If the hypothesis $H_0$ is rejected, the risk of making an error is equal to 5% (α is set here at 5%, = 100 – the confidence level of 95%). This risk is the probability of rejecting hypothesis $H_0$ when it is actually true (false positive).