Wage determination in the US

M.Watremez, J.Chadoutaud

ENAC

May 11, 2012

Introduction

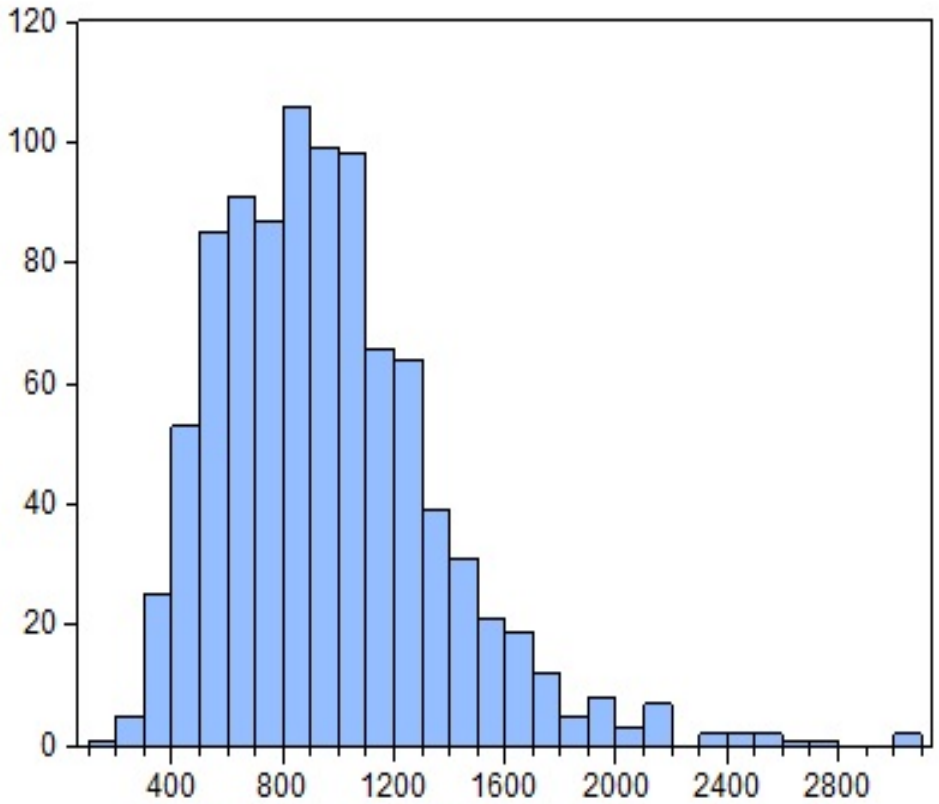
How can you get rich ?

Table of contents

First model: presentation of the variables

- **Quantitative variables :**
  - Average weekly hours
  - Years of education
  - Years of experience
  - Years with current employer
  - Age
  - Number of siblings
- **Dummy variables :**
  - familial status : $=1$ if married
  - skin color : $=1$ if black
  - geographic region : $=1$ if live on south
  - living area : $=1$ if live in urban area

Dependant variable : Wage



Series: WAGE	
Sample 1 935	
Observations 935	
Mean	957.9455
Median	905.0000
Maximum	3078.000
Minimum	115.0000
Std. Dev.	404.3608
Skewness	1.199259
Kurtosis	5.696661
Jarque-Bera	507.4271
Probability	0.000000

Figure : Wage histogram

Some remarks about the explanatory variables :

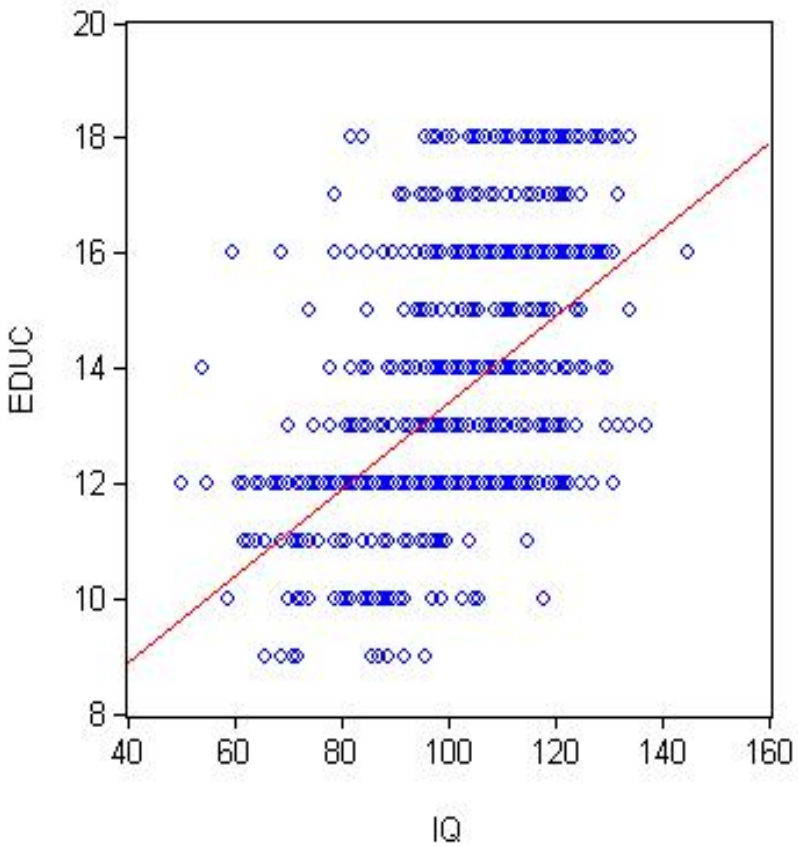


Figure : Education vs IQ

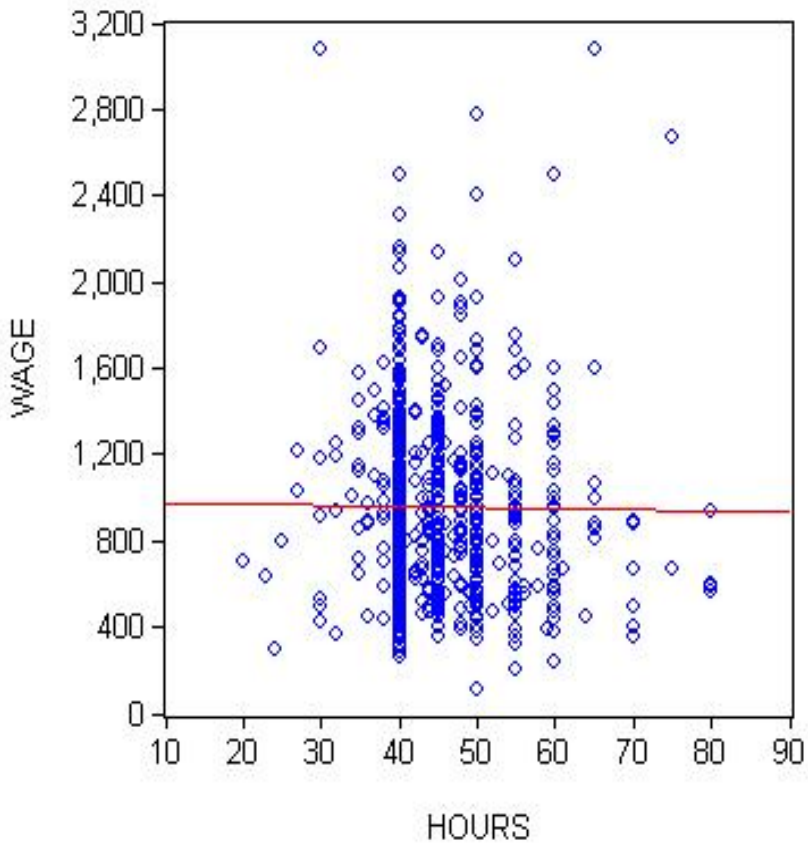


Figure : Wage vs Hours

Table of contents

First equation

Dependent Variable: WAGE
Method: Least Squares
Date: 05/03/12 Time: 13:31
Sample: 1 935
Included observations: 935

	Coefficient	Std. Error	t-Statistic	Prob.
C	-678.7060	180.4970	-3.760206	0.0002
URBAN	168.2340	26.12980	6.438396	0.0000
TENURE	5.465555	2.412825	2.265210	0.0237
SOUTH	-50.95740	25.59870	-1.990624	0.0468
SIBS	-3.023807	5.423062	-0.557583	0.5773
MARRIED	177.8918	37.89673	4.694120	0.0000
IQ	3.488324	0.975847	3.574664	0.0004
HOURS	-3.215964	1.622862	-1.981662	0.0478
EXPER	9.568107	3.599924	2.657864	0.0080
EDUC	52.00638	7.061792	7.364474	0.0000
BLACK	-120.9769	39.67422	-3.049258	0.0024
AGE	10.16323	4.552763	2.232321	0.0258

R-squared	0.243508	Mean dependent var	957.9455
Adjusted R-squared	0.234493	S.D. dependent var	404.3608
S.E. of regression	353.7886	Akaike info criterion	14.58803
Sum squared resid	1.16E+08	Schwarz criterion	14.65015
Log likelihood	-6807.903	Hannan-Quinn criter.	14.61172
F-statistic	27.00961	Durbin-Watson stat	1.852249
Prob(F-statistic)	0.000000		

Figure : least squares

least squares

- Coefficients and their T-statistic
- Adjusted R-squared

Wage equation

$$WAGE_i = \beta_0 + \beta_1 URBAN_i + \beta_2 TENURE_i + \beta_3 SOUTH_i + \beta_4 SIBS_i + \beta_5 MARRIED_i + \beta_6 IQ_i + \beta_7 HOURS_i + \beta_8 EXPER_i + \beta_9 EDUC_i + \beta_{10} BLACK_i + \beta_{11} AGE_i \quad (1$$

Equation optimization

Table: CORRELATION Workfile: PROJECT WAGE2::Wage2\

View | Proc | Object | Print | Name | Edit+/- | CellFmt | Grid+/- | Title | Comments+/-

Correlation

	AGE	BLACK	EDUC	EXPER	HOURS	IQ	MARRIED	SIBS	TENURE	SOUTH	URBAN	WAGE
AGE	1.000000	-0.035646	-0.012254	0.495330	0.024812	-0.043741	0.106980	-0.040719	0.270602	-0.029478	-0.006749	0.156702
BLACK	-0.035646	1.000000	-0.179457	0.055849	-0.107850	-0.387869	-0.053448	0.305277	-0.078236	0.236458	0.070200	-0.210878
EDUC	-0.012254	-0.179457	1.000000	-0.455573	0.091009	0.515697	-0.058566	-0.239288	-0.036167	-0.097033	0.072151	0.327109
EXPER	0.495330	0.055849	-0.455573	1.000000	-0.062126	-0.224913	0.106349	0.064310	0.243654	0.021257	-0.047386	0.002190
HOURS	0.024812	-0.107850	0.091009	-0.062126	1.000000	0.073839	0.032563	-0.049603	-0.055528	-0.029519	0.016573	-0.009504
IQ	-0.043741	-0.387869	0.515697	-0.224913	0.073839	1.000000	-0.014668	-0.284773	0.042159	-0.209785	0.038936	0.309088
MARRIED	0.106980	-0.053448	-0.058566	0.106349	0.032563	-0.014668	1.000000	-0.004327	0.072605	0.022757	-0.040248	0.136583
SIBS	-0.040719	0.305277	-0.239288	0.064310	-0.049603	-0.284773	-0.004327	1.000000	-0.039161	0.066320	-0.031469	-0.159204
TENURE	0.270602	-0.078236	-0.036167	0.243654	-0.055528	0.042159	0.072605	-0.039161	1.000000	-0.061691	-0.038486	0.128266
SOUTH	-0.029478	0.236458	-0.097033	0.021257	-0.029519	-0.209785	0.022757	0.066320	-0.061691	1.000000	-0.109898	-0.159387
URBAN	-0.006749	0.070200	0.072151	-0.047386	0.016573	0.038936	-0.040248	-0.031469	-0.038486	-0.109898	1.000000	0.198406
WAGE	0.156702	-0.210878	0.327109	0.002190	-0.009504	0.309088	0.136583	-0.159204	0.128266	-0.159387	0.198406	1.000000

Figure : Correlation matrix

- Are there any correlations between variables? Apparently, *NO*

Equation optimization

Wald Test:
Equation: EQ_TOT

Test Statistic	Value	df	Probability
F-statistic	0.310899	(1, 923)	0.5773
Chi-square	0.310899	1	0.5771

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(5)	-3.023807	5.423062

Restrictions are linear in coefficients.

Figure : Wald test

Wald test

- $$H_0 : \beta_4 = c_5 = 0 \qquad (2)$$
- F-statistic gives : We don't reject the null hypothesis
- Conclusion:

$$\beta_4 = 0 \qquad (3)$$

Equation optimization

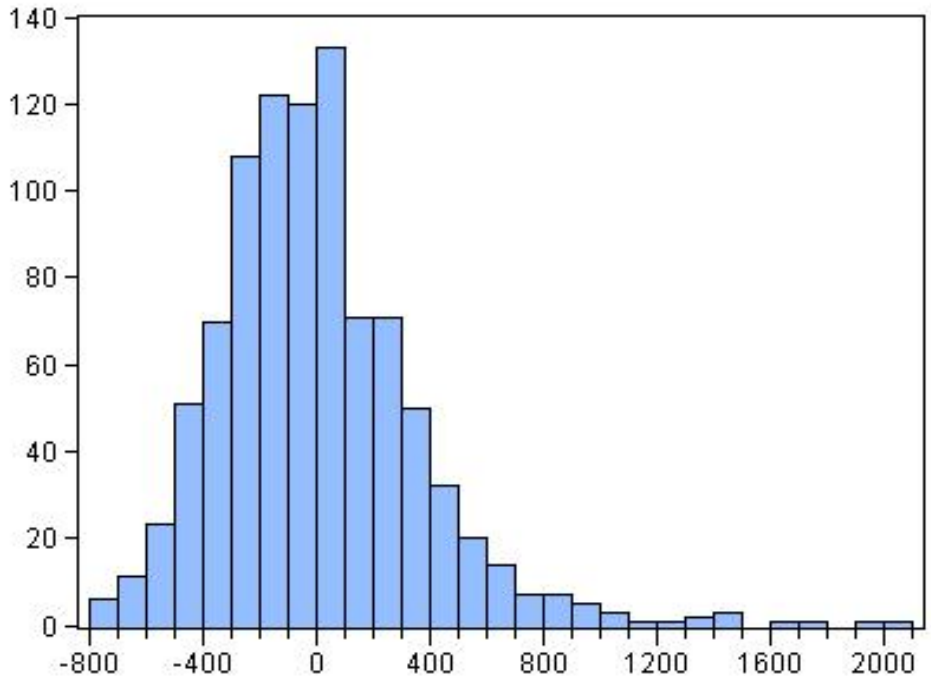
Heteroskedasticity Test: White			
F-statistic	1.023234	Prob. F(50,884)	0.4312
Obs*R-squared	51.15283	Prob. Chi-Square(50)	0.4282
Scaled explained SS	150.1225	Prob. Chi-Square(50)	0.0000

Figure : heteroskedasticity

heteroskedasticity

- $$H_0 : Homoskedasticity, H_1 : Heteroskedasticity \quad (4)$$
- $$Prob.Chi - Square(50) = 0.42 > 0.05 \quad (5)$$
- Conslusion: We don't reject the null Hypothesis, but we removed HOURS

Equation optimization



Series: RESID	
Sample 1 935	
Observations 935	
Mean	-7.86e-14
Median	-39.67455
Maximum	2089.579
Minimum	-799.4021
Std. Dev.	352.5040
Skewness	1.274840
Kurtosis	6.997161
Jarque-Bera	875.7119
Probability	0.000000

Exogeneity and Normality

- Mean=0 : Exogeneity
- According to J-B test, we reject Normality...

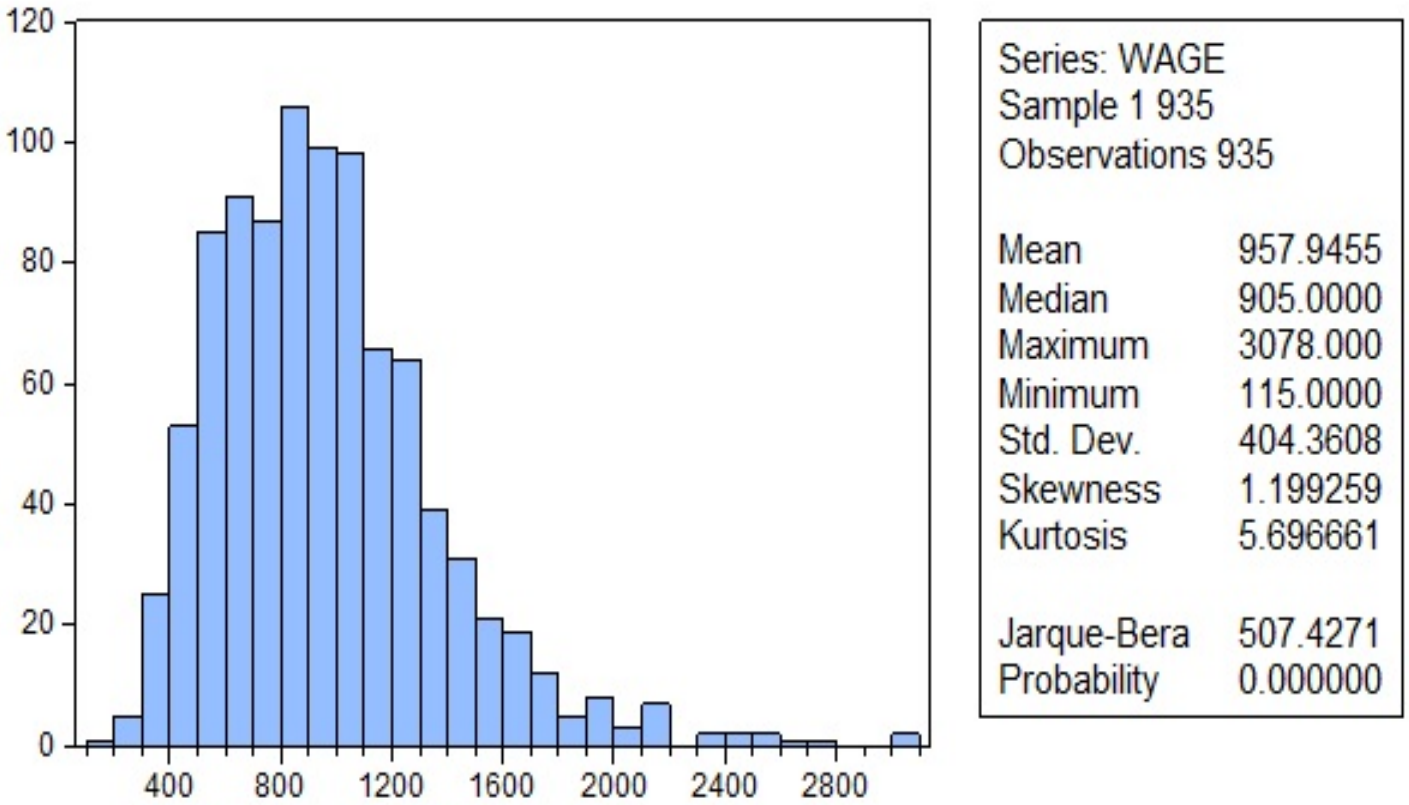


Figure : Wage histogramm

Optimized equation

Dependent Variable: WAGE
Method: Least Squares
Date: 05/08/12 Time: 13:51
Sample: 1 935
Included observations: 935

	Coefficient	Std. Error	t-Statistic	Prob.
C	-823.8317	164.5668	-5.006063	0.0000
AGE	9.746359	4.551200	2.141492	0.0325
BLACK	-119.2974	38.54503	-3.095013	0.0020
EDUC	51.97934	7.018052	7.406520	0.0000
EXPER	9.893982	3.600402	2.748021	0.0061
IQ	3.539171	0.970911	3.645209	0.0003
MARRIED	175.2001	37.91924	4.620350	0.0000
SOUTH	-50.35592	25.61090	-1.966191	0.0496
TENURE	5.790615	2.410475	2.402271	0.0165
URBAN	167.9889	26.13561	6.427588	0.0000

R-squared	0.240042	Mean dependent var	957.9455
Adjusted R-squared	0.232647	S.D. dependent var	404.3608
S.E. of regression	354.2147	Akaike info criterion	14.58832
Sum squared resid	1.16E+08	Schwarz criterion	14.64009
Log likelihood	-6810.040	Hannan-Quinn criter.	14.60806
F-statistic	32.46353	Durbin-Watson stat	1.861535
Prob(F-statistic)	0.000000		

Figure : least squares

least squares

- Coefficients and their T-statistic
- Adjusted R-squared

Final equation

After optimization...

$$\begin{aligned} WAGE_i = & -823 + 168\,URBAN_i + 5,79\,TENURE_i - 50,4\,SOUTH_i \\ & + 175\,MARRIED_i + 3,54\,IQ_i + 9,89\,EXPER_i + 52,0\,EDUC_i \\ & + -119\,BLACK_i + 9,75\,AGE_i \end{aligned} \tag{6}$$

Table of contents

Second model : presentation of the variables

- **Quantitative variables :**
  - Years of education
  - Years of experience
  - Years with current employer
  - Number of dependents
- **Dummy variables :**
  - familial status : $=1$ if married
  - skin color : $=1$ if nonwhite
  - gender : $=1$ if female
  - different sectors : construction, manufacture, trade, services, professional occupation, professional services, clerical occupation, services occupation

Table of contents

First equation

Dependent Variable: WAGE
Method: Least Squares
Date: 05/08/12 Time: 16:56
Sample: 1 526
Included observations: 526

	Coefficient	Std. Error	t-Statistic	Prob.
C	1.184722	0.843518	1.404501	0.1608
CLEROCC	0.430651	0.429210	1.003358	0.3162
CONSTRUC	-0.711915	0.655607	-1.085886	0.2780
EDUC	0.389140	0.058361	6.667775	0.0000
EXPER	0.011688	0.011776	0.992505	0.3214
FEMALE	-1.551940	0.281797	-5.507288	0.0000
MARRIED	0.326335	0.283036	1.152982	0.2495
NDURMAN	-1.215293	0.479157	-2.536315	0.0115
NONWHITE	-0.034517	0.409068	-0.084381	0.9328
NUMDEP	0.135873	0.103441	1.313531	0.1896
PROFOCC	2.028540	0.366205	5.539354	0.0000
PROFSERV	-1.159432	0.446729	-2.595382	0.0097
SERVICES	-1.829079	0.519348	-3.521875	0.0005
SERVOCC	-0.206211	0.432625	-0.476650	0.6338
TENURE	0.120351	0.020310	5.925596	0.0000
TRADE	-2.295843	0.407130	-5.639084	0.0000
TRCOMMPU	-1.301348	0.676791	-1.922822	0.0551

R-squared	0.450689	Mean dependent var	5.896103
Adjusted R-squared	0.433422	S.D. dependent var	3.693086
S.E. of regression	2.779838	Akaike info criterion	4.914448
Sum squared resid	3933.296	Schwarz criterion	5.052300
Log likelihood	-1275.500	Hannan-Quinn crite.	4.968423
F-statistic	26.10094	Durbin-Watson stat	1.823987
Prob(F-statistic)	0.000000		

Figure : least squares

least squares

- Adjusted R-squared
- T-statistic

Equation optimization

- Walt test
- Heteroskedasicity
- Correlations
- Exogeneity and Normality

Final equation

	Coefficient	Std. Error	t-Statistic	Prob.
C	1.457280	0.692599	2.104074	0.0359
EDUC	0.353597	0.051704	6.838828	0.0000
TENURE	0.128837	0.018063	7.132479	0.0000
FEMALE	-1.453555	0.264901	-5.487159	0.0000
MARRIED	0.512785	0.261336	1.962169	0.0503
NDURMAN	-0.871014	0.439028	-1.983962	0.0478
TRADE	-1.925829	0.346178	-5.563117	0.0000
SERVICES	-1.548458	0.464432	-3.334087	0.0009
PROFSERV	-0.758692	0.368993	-2.056115	0.0403
PROFOCC	1.983632	0.301895	6.570601	0.0000

R-squared	0.441487	Mean dependent var	5.896103
Adjusted R-squared	0.431745	S.D. dependent var	3.693086
S.E. of regression	2.783947	Akaike info criterion	4.904445
Sum squared resid	3999.185	Schwarz criterion	4.985534
Log likelihood	-1279.869	Hannan-Quinn criter.	4.936195
F-statistic	45.32018	Durbin-Watson stat	1.845418
Prob(F-statistic)	0.000000		

Figure : Optimized equation

After optimization...

$$\begin{aligned} WAGE_i = {} & 1,46 + 0,35\,EDUC_i + 0,13\,TENURE_i - 1,45\,FEMALE_i \\ & + 0,51\,MARRIED_i - 0,87\,NDURMAN_i - 1,93\,TRADE_i \\ & - 1,55\,SERVICES_i - 0,76\,PROFSERV_i + 1,98\,PROFOCC_i \end{aligned} \quad (7)$$

Conclusion

- Similar results with the first equation
- Differences between the working sectors
- Better model than the first one

Any questions ?