Migration in Europe
MigrEU Jean Monnet Module

Co-funded by the Erasmus+ Programme of the European Union

Introduction to Linear Regression 1

Economics of Migration in Europe

Salvatore Carrozzo (salvatore.carrozzo@unito.it)

University of Turin

UNIVERSITÀ DEGLI STUDI DI TORINO

Co-funded by the Erasmus+ Programme of the European Union

Outline

- Introduction
- Linear Regression Fundamentals
    1. Univariate Regression
    2. Multivariate Regression
- Application

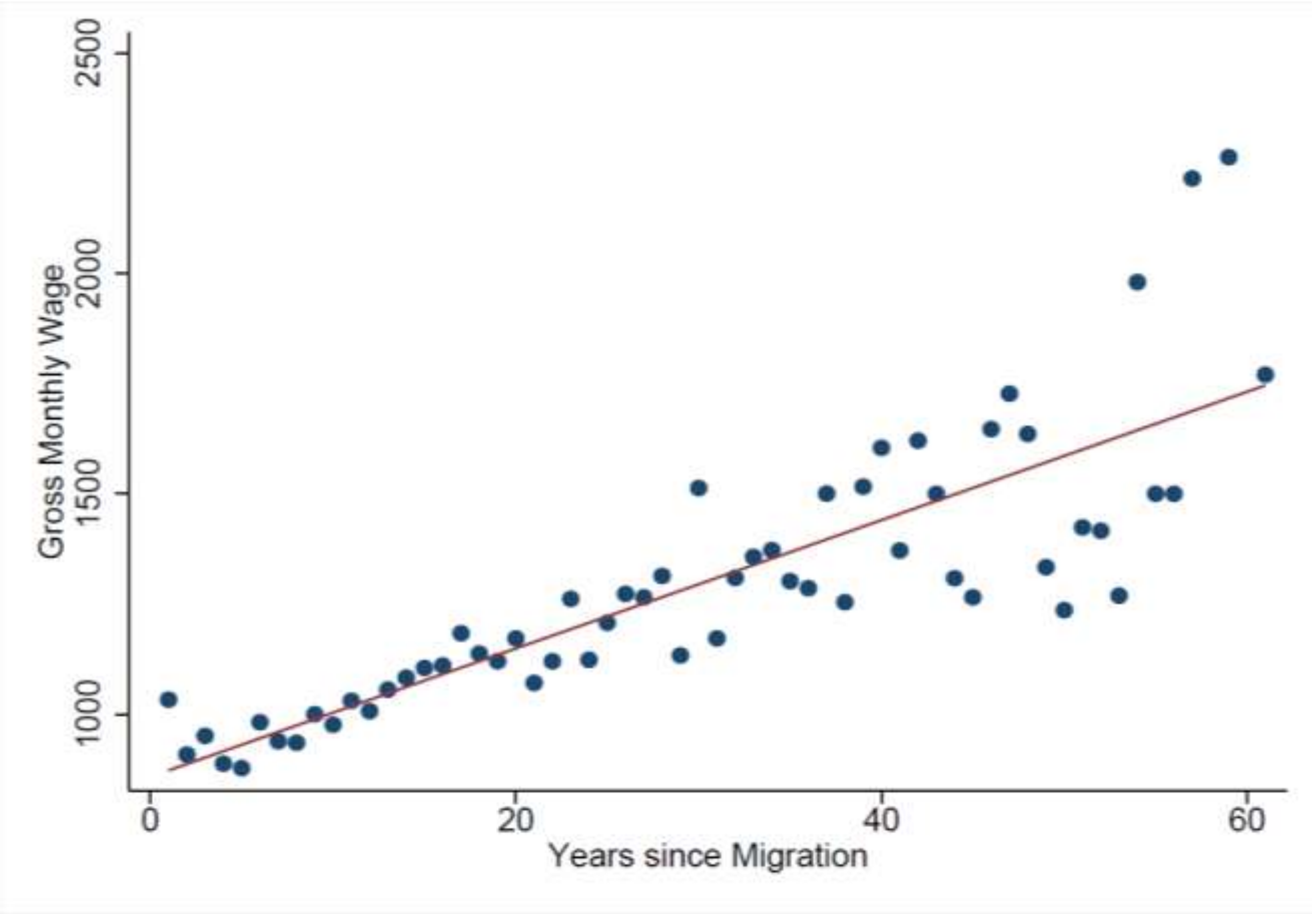
Introduction 1/2

What is a **linear regression**?

Linear regression aims to study the **effect**, if any, of a change in one or more independent variables (explanatory) on a dependent variable (outcome). E.g. the effect of an increase in years since migration on the individual wages.

Co-funded by the Erasmus+ Programme of the European Union

Relationship between Italian Gross Monthly Average Wage of Foreign Workers and Years Since Migration in December 2013



Co-funded by the Erasmus+ Programme of the European Union

Introduction 2/2

- What is the **linear regression design**?

- Linear regression fits data in a linear model to show the relationship between independent variables and dependent variable. The independent variables are multiplied by a coefficient, which shows the average effect of each independent variable on the dependent variable.

Co-funded by the Erasmus+ Programme of the European Union

Types of Regression analysis

Univariate Regression

Univariate regression studies the relationship between an independent variable and a dependent variable.

Multivariate Regression

Multivariate regression studies the relationship between more than one independent variable and a dependent variable.

Univariate Regression

Formula

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (1)$$

β_0: is the intercept of the line

β_1: is the slope of the line

i: indicates a person

ε_i: is an error term due to the fitting

y_i: is the dependent variable

x_i: is the explanatory or independent variable

Co-funded by the Erasmus+ Programme of the European Union

Interpretation of the coefficients

Coefficient formulas

$$\beta_1 = \frac{Cov(y_i, x_i)}{Var(x_i)} \tag{2}$$

$$\beta_0 = E[y_i] - \beta_1 E[x_i] \tag{3}$$

β_1: provides the effect of one unit increase in all x_is on all y_is.

β_0: provides the value of y_is when the independent variable is equal to zero.

Co-funded by the Erasmus+ Programme of the European Union

Table Example (1/2)

Table: The Effect Of An Increase In Years Since Migration On The Foreigners' Gross Monthly Wage

	Wage
Years since migration	13.67***
	(0.787)
N	3331
R^2	0.118

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Co-funded by the Erasmus+ Programme of the European Union

Table Example (2/2)

Comments:

One year more spent in the host country increases the gross monthly wage by 13.67 euro (β_1) on average. If immigrants spend two years more, the increase in the gross monthly wage is 2*13.67 euro=27.34 euro.

Co-funded by the Erasmus+ Programme of the European Union

Multivariate Regression

Formula

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \tag{4}$$

β_0: is the intercept

β_1, β_2: are the coefficients

i: indicates a person

ε_i: is an error term due to the fitting

y_i: is the dependent variable

x_{i1}, x_{i2}: are the explanatory or independent variables

Interpretation of the coefficients

β_2: provides the effect of one unit increase in all x_{i2} on all y_i from averaging out the effect of x_i2 on x_i1.
β_1: provides the effect of one unit increase in all x_{i1} on all y_i from averaging out the effect of x_i1 on x_i2.
β_0: provides the value of y_i when both independent variables are equal to zero.

Co-funded by the Erasmus+ Programme of the European Union

Table Example (1/2)

Table: The Effect Of Both An Increase In Years Since Migration And Total Hours Worked On The Foreigners' Gross Monthly Wage

	Wage
Years since	13.93***
migration	(0.769)
Total hours	9.87***
worked	(0.618)
N	3326
R^2	0.209

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Co-funded by the
Erasmus+ Programme
of the European Union

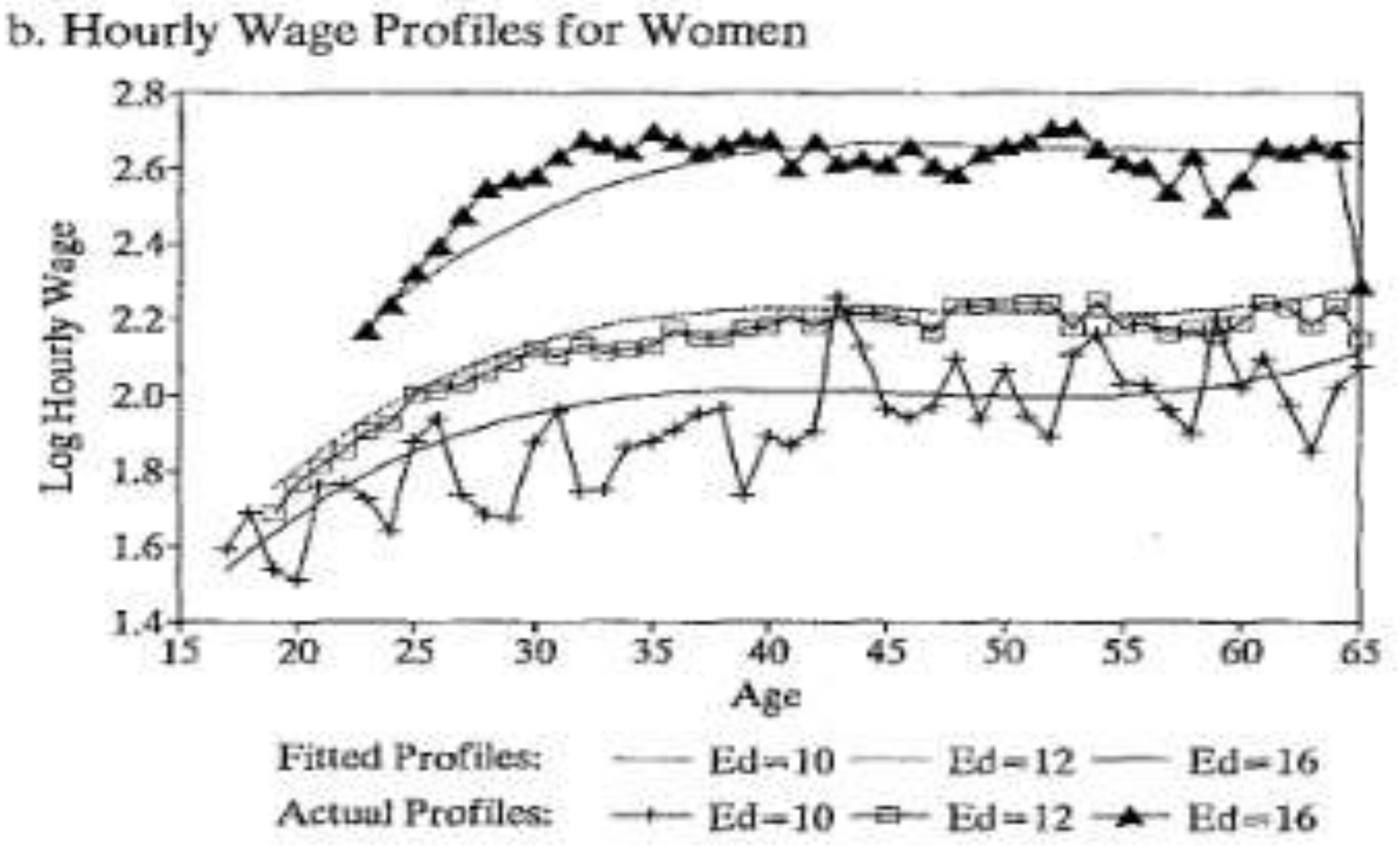
Table Example (2/2)

Interpretation:

One year more spent in host country increases the gross monthly wage by 13.93 euro (β_1) on average. Further, an hour spent more at work worked increases the gross monthly wage by 9.87 euro (β_2) on average.

Co-funded by the Erasmus+ Programme of the European Union

Age profiles of hourly wages for women with different level of education (Card, 1999)



Co-funded by the Erasmus+ Programme of the European Union

Comments

Graph shows the relationship between age and the earning profiles for different level of education. Age proxies the experience and, for this reason, has a positive effect on hourly wages. Further, higher education level leads to higher earnings per se. The solid lines among dots are the regression lines.

Co-funded by the Erasmus+ Programme of the European Union

Table: the relationship between education and wage

Estimated education coefficients from standard human capital earnings function fit to hourly wages, annual earnings, and various measures of hours for men and women in March 1994–1996 Current Population Survey[a]

	Dependent variable				
	Log hourly earnings (1)	Log hours per week (2)	Log weeks per year (3)	Log annual hours (4)	Log annual earnings (5)
A. Men					
Education coefficient	0.100	0.018	0.025	0.042	0.142
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
R-squared	0.328	0.182	0.136	0.222	0.403
B. Women					
Education coefficient	0.109	0.022	0.034	0.056	0.165
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
R-squared	0.247	0.071	0.074	0.105	0.247

[a] Notes: Table reports estimated coefficient of linear education term in model that also includes cubic in potential experience and an indicator for non-white race. Samples include men and women age 16–66 who report positive wage and salary earnings in the previous year. Hourly wage is constructed by dividing wage and salary earnings by the product of weeks worked and usual hours per week. Data for individuals whose wage is under $2.00 or over $150.00 (in 1995 dollars) are dropped. Sample sizes are: 102,639 men and 95,309 women.

Co-funded by the Erasmus+ Programme of the European Union

Comments

This regression analysis's table shows the effect of an increase in the education level on labour outcomes. The red circled shows the increase of one more year in the education level on wage, which is 0.1 log points (around 10% increase).

Migration in Europe
MigrEU Jean Monnet Module

Co-funded by the Erasmus+ Programme of the European Union

Introduction to Linear Regression 2

Economics of Migration in Europe

Salvatore Carrozzo (salvatore.carrozzo@unito.it)

University of Turin

UNIVERSITÀ DEGLI STUDI DI TORINO

Outline

- Significance
- Fitting
- Application

Co-funded by the Erasmus+ Programme of the European Union

Significance

- What does it mean significance?

  A linear regression coefficient is significant when it is different from zero.

- How can we test it?

  We check whether the ratio of the coefficent to the standard error (the variability of the coefficient) is larger than 2.

Co-funded by the Erasmus+ Programme of the European Union

Table Example (1/2)

Table: The Effect Of An Increase In Years Since Migration On The Foreigners' Gross Monthly Wage

	Wage
Years since migration	13.67***
	(0.787)
N	3331
R^2	0.118

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Co-funded by the Erasmus+ Programme of the European Union

Table Example (2/2)

Table: The Effect Of Both An Increase In Years Since Migration And Total Hours Worked On The Foreigners' Gross Monthly Wage

	Wage
Years since migration	13.93***
	(0.769)
Total hours worked	9.87***
	(0.618)
N	3326
R^2	0.209

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Co-funded by the Erasmus+ Programme of the European Union

Fitting

- **Do independent variables explain the dependent variable?**

  In order to answer to this question, we compute an index to show whether the model has a good fitting.

- **Why do we care about the fitting?**

  Linear Regression has a predictive use. It shows the ability of one or more independent variables to foresee the values of the dependent variable. (E.g. years since migration predict the current wage.)

Co-funded by the Erasmus+ Programme of the European Union

Predictive use

Linear regression may:

- Foresee the future values of the dependent variable;
- Predict the dependent variable values given different values of the independent variables;
- Give information on the key variables to figure out the dynamics of a dependent variable

Increasing the predictive power

- **How do we increase the predictive ability of a model?**

We increase the number of independent variables until we get a good level of the R^2.

- **Testing the model with R^2**

The R^2 shows the predictive ability of the model. The index range is between 0% and 100%.

Co-funded by the Erasmus+ Programme of the European Union

Table Example

Table: The Effect Of An Increase in Total Hours Worked on Total Wages Within Each Year Since Migration Cell

	Wage
Hours Worked	25.59***
	(0.467)
Observations	69
R^2	0.992

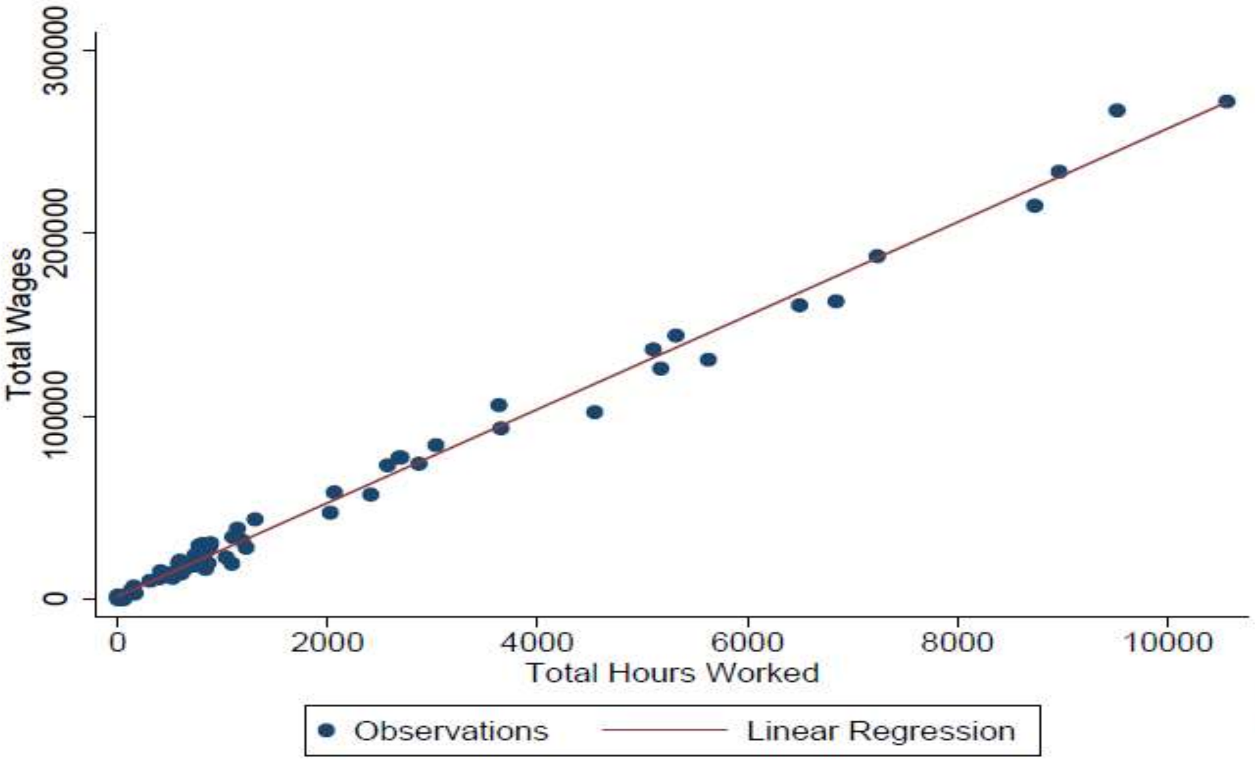
Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Co-funded by the Erasmus+ Programme of the European Union

Graph Example

Figure: Pooled Scatter Plot of Total Hours Worked and Total Wages Within Each Years Since Migration Cell



Co-funded by the Erasmus+ Programme of the European Union

Application 1: Card, 1999

Estimated education coefficients from standard human capital earnings function fit to hourly wages, annual earnings, and various measures of hours for men and women in March 1994–1996 Current Population Survey[a]

	Dependent variable				
	Log hourly earnings (1)	Log hours per week (2)	Log weeks per year (3)	Log annual hours (4)	Log annual earnings (5)
A. Men					
Education coefficient	0.100 (0.001)	0.018 (0.001)	0.025 (0.001)	0.042 (0.001)	0.142 (0.001)
R-squared	0.328	0.182	0.136	0.222	0.403
B. Women					
Education coefficient	0.109 (0.001)	0.022 (0.001)	0.034 (0.001)	0.056 (0.001)	0.165 (0.001)
R-squared	0.247	0.071	0.074	0.105	0.247

[a] Notes: Table reports estimated coefficient of linear education term in model that also includes cubic in potential experience and an indicator for non-white race. Samples include men and women age 16–66 who report positive wage and salary earnings in the previous year. Hourly wage is constructed by dividing wage and salary earnings by the product of weeks worked and usual hours per week. Data for individuals whose wage is under $2.00 or over $150.00 (in 1995 dollars) are dropped. Sample sizes are: 102,639 men and 95,309 women.

Co-funded by the Erasmus+ Programme of the European Union

Comments

- The coefficient (red circled number) is different from zero since the ratio of the coeffienct to the standard error (the blue circled number) is larger than 2.
- Education is a good predictor for the hourly earnings since the R^2 is around 30%.
- Insight: the education affects the earnings for sure. However the magnitude of the coefficient might be larger since education might be also a good proxy for the innate ability.

Application 2: Borjas et al., 1996

TABLE 1—CROSS-SECTIONAL IMPACT OF IMMIGRATION ON NATIVE WAGE [DEPENDENT VARIABLE = ln(WEEKLY WAGE)]

Independent variable	Regression coefficients			
	Male natives		Female natives	
	1980	1990	1980	1990
Relative number of immigrants in metropolitan area $j\ (I_j/N_j)$	−0.0173 (0.0813)	0.2869 (0.0721)	0.4525 (0.0941)	0.5588 (0.1059)
Relative number of immigrants in metropolitan area j and education group $k\ (I_{jk}/N_{jk})$	−0.0119 (0.0410)	0.1346 (0.0293)	0.2876 (0.0621)	0.2865 (0.0622)
Sample size	312,446	299,202	268,649	288,620

Notes: Standard errors are reported in parentheses. The cross-sectional regressions also include a vector of dummy variat dicating the worker's age (18–24, 25–34, 35–54, or 55–6 old) and educational attainment (high-school dropout, high-school graduate, some college, or college graduate). The sample is restricted to native workers who reside in one of the 236 metropolitan areas that can be matched in the 1980 and 1990 Censuses.

Comments

- All the coefficients are different from zero but the ones in the first column.
- R^2 is not reported. Maybe, scholars are only interested to study the effect of an increase in the migrant share on the native wage.
- Insight: An increase in immigrant share might be a proxy for the better wages in the host country. Hence, we are not sure in the case whether an increase in immigrant share predicts a change in native wage.