

Forecasting of the inflow of graduates on the labour market at all education levels

A report for CEDEFOP Medium-term forecast of skills supply in Europe

Arnaud Dupuy



Introduction

- 🔗 The report focuses on:
 - 🔗 A description of the methodology used to derive projections of the number of people entering the labour market by educational levels,
 - 🔗 A description of the data used to this aim, i.e. aggregate data from the UNESCO-OECD-Eurostat (UOE) data collection on education statistics on population by age groups and graduates by level and field of education, and
 - 🔗 A documentation of the limitations of the data with respect to the aim of the project.



Methodology I

- ⌚ We proceed in two steps:
 - ⌚ Step 1: We make projections of the number of students and graduates by educational levels (3 levels, 0-2, 3-4 and 5-6), age category (15-19, 20-24, 25-29 and gender
 - ⌚ Step 2: We make projections of the number of graduates entering the labor market by educational level



Methodology: Step I

- Step 1: We follow Ekamper 2007 and derive projections of the number of students and graduates by educational levels (0-2, 3-4 and 5-6) using the well-known participation ratio method.

- Formally, the procedure is the following:

- 1. Participation ratios for 1998-2005 are calculated according to:

$$\text{Pr}^c_{dt} = S^c_{dt} / P_{dt}$$

- 2. We formulate predictions of the participation ratios, Pr^c_{dt+i} , using the series P^c_{dt}

- 3. Using independent projected population, say P^c_{dt+i} , we calculated the Projected number of students with certain characteristics:

$$S^c_{dt+i} = \text{Pr}^c_{dt+i} \times P_{dt+i}$$

Methodology: step I

☞ The same procedure can be used to derive projections of the number of graduates.

☞ 1. Graduation ratios for 1998-2005 are calculated according to:

$$Gr^c_{dt} = G^c_{dt} / S^c_{dt}$$

☞ 2. We formulate predictions of the graduation ratios, Gr^c_{dt+i} , using the series G^c_{dt}

☞ 3. Using the projected number of students, say S^c_{dt+i} , we calculated the Projected number of graduates with certain characteristics: $G^c_{dt+i} = Gr^c_{dt+i} \times S^c_{dt+i}$

Methodology: step II

- Step 2: Projections of the number of graduates entering the labor market
- To derive projections of the number of individuals that enter the labor market for each gender, by level and year (for each country), we simply subtract to the number of graduates at each level the number of graduates that go on with their study.
- This requires information about the share of graduates at each level that go on with their study, α^c_{dt}
- The number of graduates at level c entering the labor market at t is then given by $(1 - \alpha^c_{dt+i})G^c_{dt+i} = E^c_{dt+i}$

Data description I

- Source: from the joint UIS (UNESCO institute of statistics)/OECD/Eurostat questionnaires on education statistics which constitute the core of the database on education. Each of the 27 countries considered provided data coming from administrative records on the basis of commonly agreed definitions.
- ISCED 0-6 is the basis for international education statistics. We distinguish three levels: 0-2, 3-4 and 5-6. Unfortunately, while specific information for educational level 3C (short) is available for the number of graduates in some countries, no distinction from the aggregate level 3 is possible for the number of students by educational level.

Data description II

- Variables of interest:
 - The number of students by ISCED level, age category (15-19, 20-24, 25-29 and 30-34) and gender for the period 1998-2005 and,
 - The number of graduates by ISCED level, age category and gender for the period 1998-2005.
- The data is furthermore divided in age groups: 15-19, 20-24- 25-29 and 30-34.

Data description III

- Source 2: Eurostat population and Eurostat population projections database from which information on population by gender, age and EU member state and future population projections (under a number of different assumptions) by gender, age (single year age groups) and EU member state for the period 2006-2025 (base year 2004).
- Data on past population trends is available for the EU-27 countries plus Norway and Switzerland.
- However, data on population projections is only available for the EU-27 countries and not for Switzerland and Norway.



Data limitations: Missing values I

- Number of students 29 countries, 2 genders and 8 isced levels => 464 time series

Number of missing values/ age bracket	15-19	20-24	25-29	30-34
0	301 (64.8%)	298 (64.2%)	284 (61.2%)	269 (57.9%)
1	5	5	3	3
2	26	34	38	38
3	25	20	21	22
4	35	31	34	26
5	10	13	13	14
6	27	26	26	27
7	17	18	20	25
8	18	19	25	40
The last 3 values completed (for 2003, 2004, 2005)	370 (79.7%)	373 (80.3%)	362 (78%)	350 (75.4%)

Data limitations: missing values II

- Number of graduates: 29 countries, 2 genders and 4 isced levels, which results in 232 time series

Number of missing values/ age bracket	15-19	20-24	25-29	30-34
0	106 (45.6%)	116 (40.5%)	108 (46.5%)	101 (43.5%)
1	0	2	4	0
2	16	14	20	23
3	20	22	18	21
4	17	14	18	17
5	3	0	2	4
6	16	13	10	12
7	20	20	19	13
8	34	31	33	41
The last 3 values completed (for 2003, 2004, 2005)	139 (59.9%)	142 (61.2%)	150 (64.6%)	140 (60.3%)

Solutions I

Missing values

- For series with 1 missing:
 - Step 1: estimate regression on 7 observation and impute the prediction for the missing year.
 - Step 2: use extrapolation method (several variant are thinkable) to derive projections of the participation and graduation ratios
- For series with more missing values this method is not very precise. We could use the mean of the last three observations as a projection of the ratios
- Note that for complete time series (0 or 1 missing after imputation), both methods can be used => some sensitivity analysis can be done.
- Alternative scenario are also possible. We could use projections from countries similar (or mean EU).

Solutions II

- Projections of the share of graduates entering the labour market by educational level.
- We would simply subtract to the number of graduates at each level the number of graduates that go on with their study. However, this requires information about the share of graduates at each level that go on with their study, α^c_{dt}
- Unfortunately, no data about α^c_{dt}
- Assumption I: educational system is strictly hierarchical = enter educational level c only graduates from educational level c-1,
- With Assumption I, we can use our data on the number of students and number of graduates to derive an estimation of the share of graduates entering the labor market.

$$S^{c+1}_{dt+1} - S^{c+1}_{dt} = \alpha^c_{dt} G^c_{dt} - G^{c+1}_{dt+1}$$

