Alphametrics (AM)

Alphametrics Ltd

CEDEFOP

European Centre for the Development
of Vocational Training

cambridge econometrics

connecting you to the future

Adjusting and extending the LFS data used for the projections

Skillsnet technical workshop

14-15 June 2010, Thessaloniki

Robert Stehrer and Terry Ward

**wiiw, Vienna
Alphametrics, UK**

Researchcentrum voor Onderwijs
en Arbeidsmarkt

**Research Centre for Education
and the Labour Market**

ROA

WARWICK INSTITUTE *for*
EMPLOYMENT RESEARCH

IER

Data source - LFS

Only the European **Labour Force Survey** contains details
required for all countries

- Employment and labour force by sex, age, sector,
  occupation, education level

- Data comparable across countries and reasonably
  consistent over time

- Data especially extracted by Eurostat from microdata –
  not published dataset - too aggregated for our purposes

- For employment:
  - Men+women

  - 4 age groups

  - 60 industries (NACE Rev. 1.1, 2-digit level)

  - 27 occupations (ISCO 2-digit level)

  - 3 educational categories (high, medium, low)

  - 41,664 per country for each year

Data problems

- Problems stem from: data improvements over time

- Documented breaks in series (15 cases over period 1995-2008)

- Undocumented breaks in series – these can arise for range of reasons – related to sample size and survey methods but also changing nature of jobs

- Missing sectors, occupations or education levels in some countries in some years due to sample size

- Need for adjustment to ensure consistency

New objectives

- To break down education levels in more detail – i.e. more than three broad categories of low, medium and high

- To break down occupations in more detail – i.e. more than 27 ISCO categories

Education attainment level by ISCED – LFS data
% Total employed

|	Low - basic educ		Medium - upper secondary			High - tertiary	
	1	2	3c - Voc <3 yrs	3a+b+c Other	4a+b+c	5	6
EU	5.9	16.7	1.9	46.3	3.1	25.4	0.8
DE	1.8	13.1	0.0	51.8	7.5	24.4	1.4
FR	7.4	17.4	0.0	45.1	0.1	29.4	0.6
IT	7.2	31.0	0.6	43.8	1.3	15.8	0.2
UK	0.2	9.5	12.6	44.9	0.1	31.7	1.1
ES	13.6	28.2	0.2	24.3	0.1	32.9	0.8
CZ	0.0	5.9	0.0	77.3	1.7	14.4	0.6
SK	0.1	4.4	0.0	79.1	0.0	16.1	0.2
PL	0.4	9.1	0.0	64.0	3.8	22.2	0.5
HU	0.3	12.3	0.0	63.5	2.2	21.2	0.4
AT	0.6	16.9	1.2	52.7	10.6	16.4	1.6
LV	0.6	13.1	0.0	55.7	6.5	23.8	0.3
RO	6.6	18.2	0.0	56.9	4.5	13.7	0.1
SI	1.5	14.1	0.0	60.9	0.0	21.6	1.9

Alternative education breakdown

- Conclusion - LFS does not provide sufficient breakdown by ISCED

- Alternative approach – to use data by field of study

- Data broken by 15 fields and can further sub-divide by 3 broad education levels ….
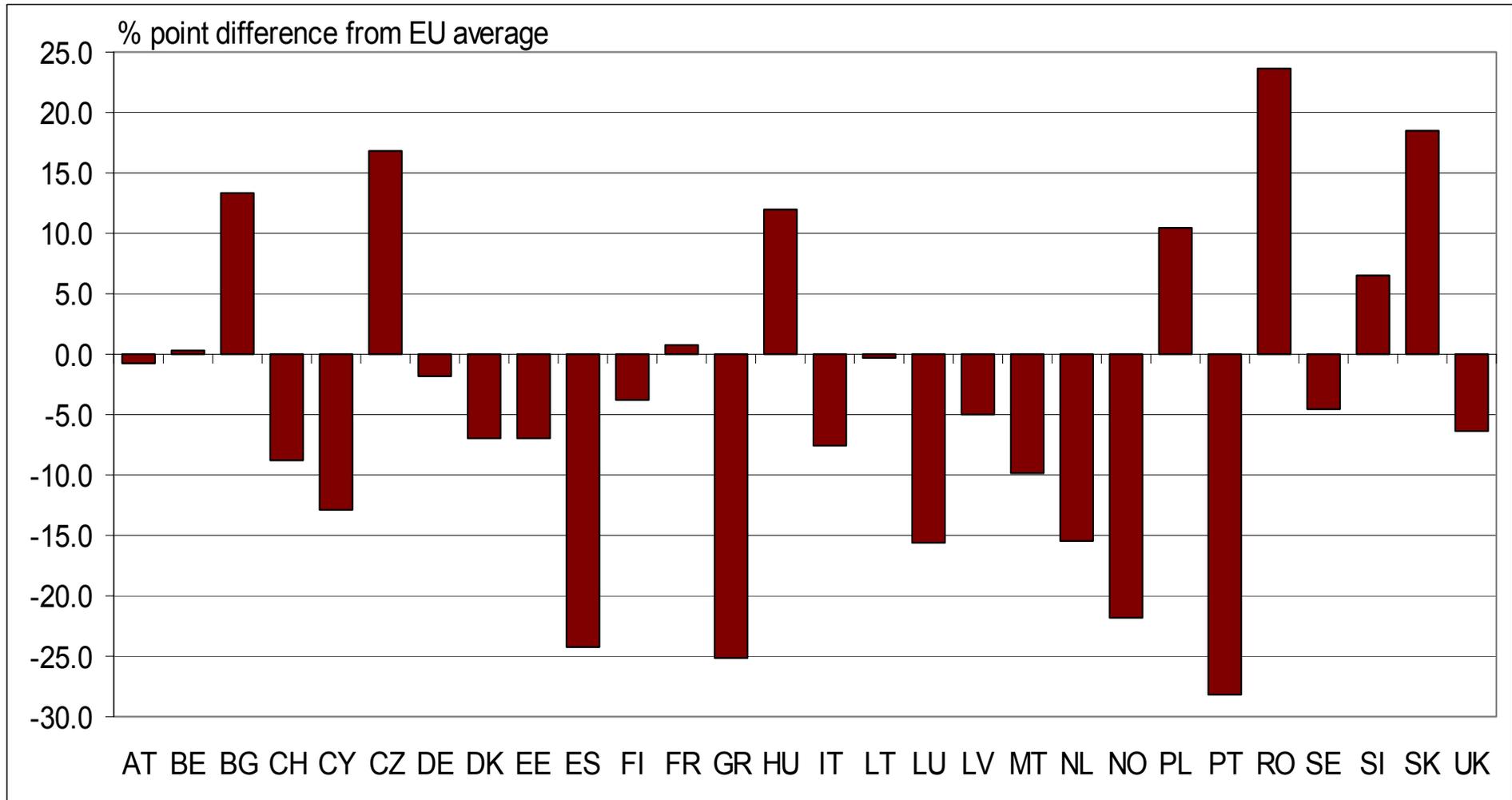
Upper secondary education

01 Agriculture and veterinary
02 Computer science
03 Engineering, manufactg,, construction
04 Foreign languages
05 General programmes
06 Health and welfare
07 Humanities, arts
08 Life sciences

09 Mathematics, statistics,
10 Physical sciences
11 Services
12 Social sciences, business, law
13 Teacher training, education
14 Computer use
15 Computing

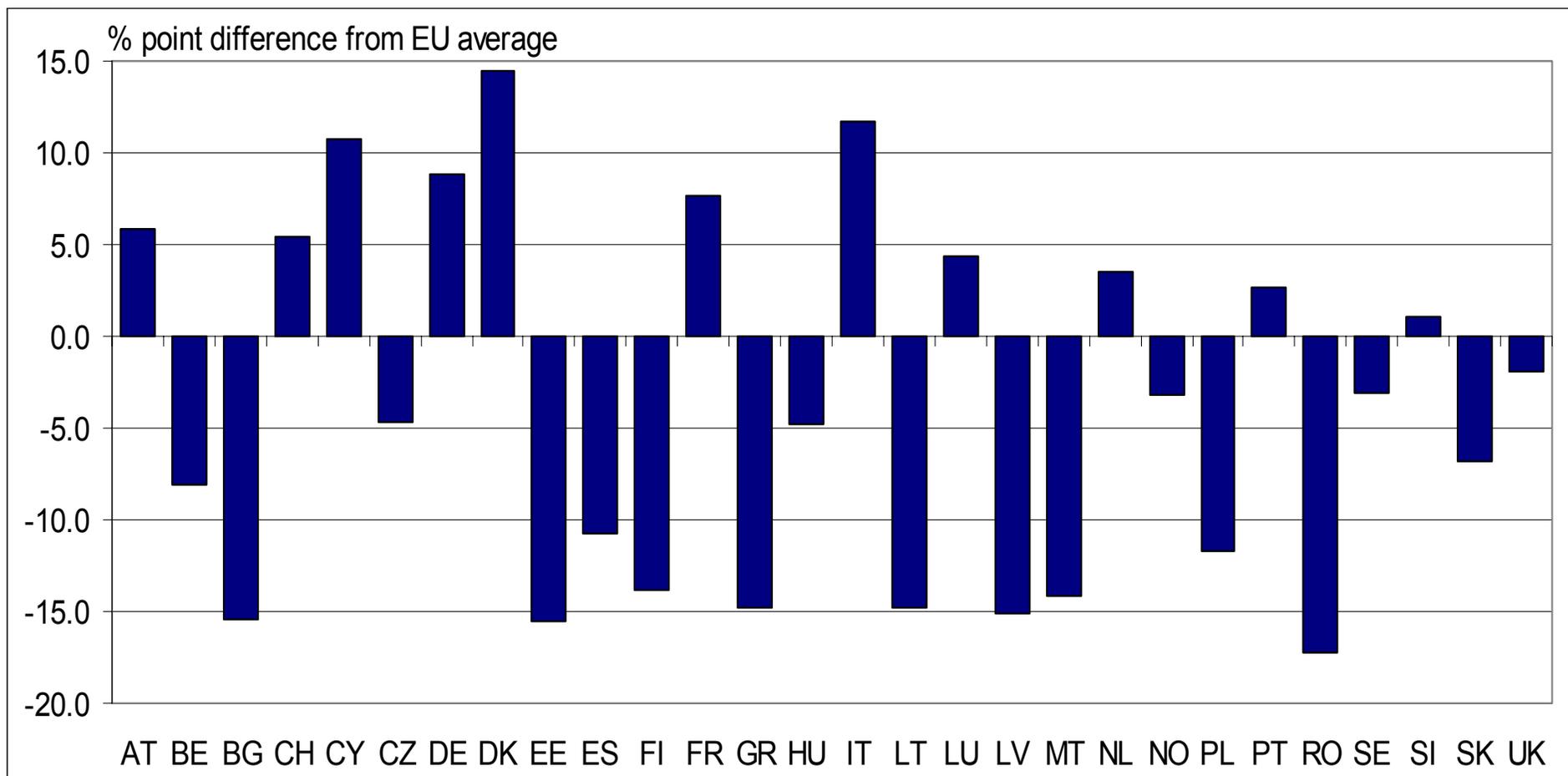# Upper secondary education
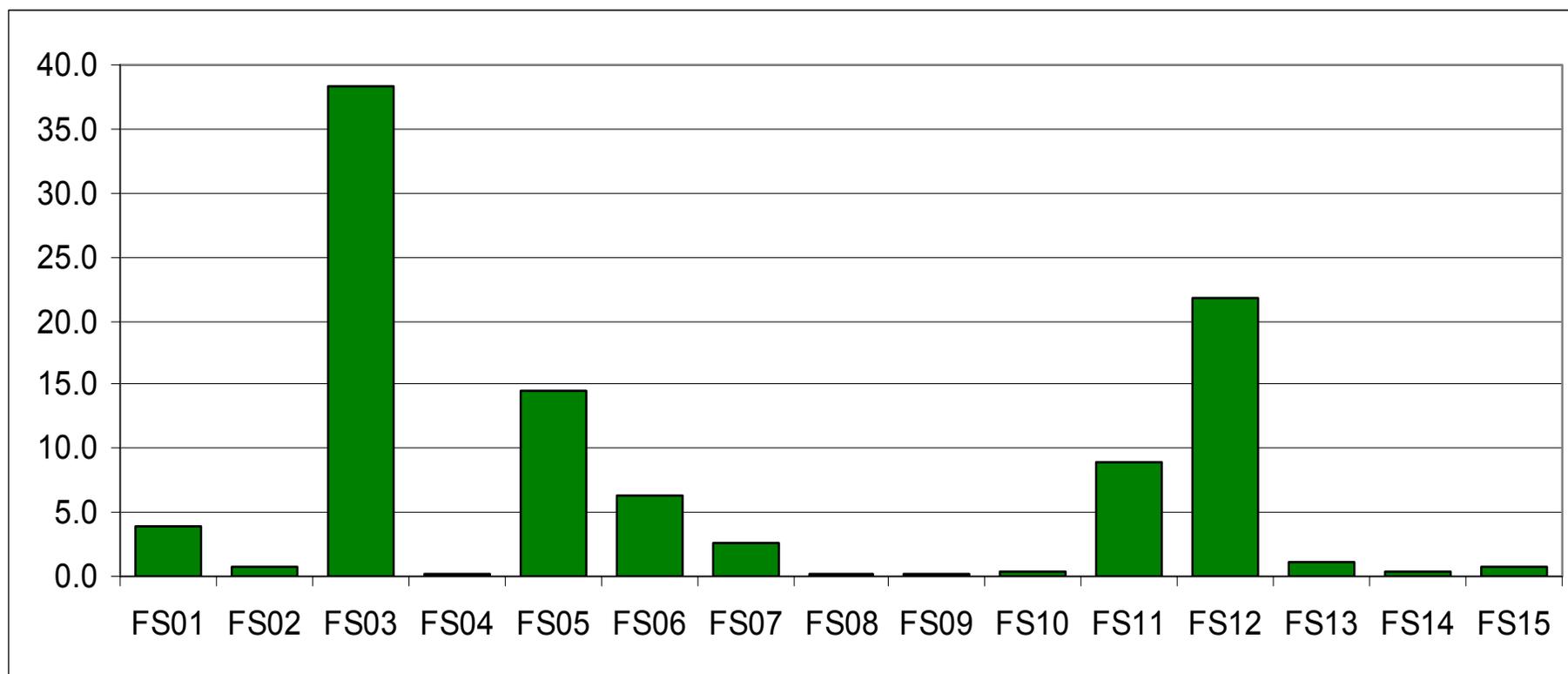
Difference in share of engineering, manufacturing, construction from EU average

% point difference from EU average

Upper secondary education

01 Agriculture and veterinary

03 Engineering, manufacturing, construction

05 General programmes

06 Health and welfare

07 Humanities, arts

11 Services

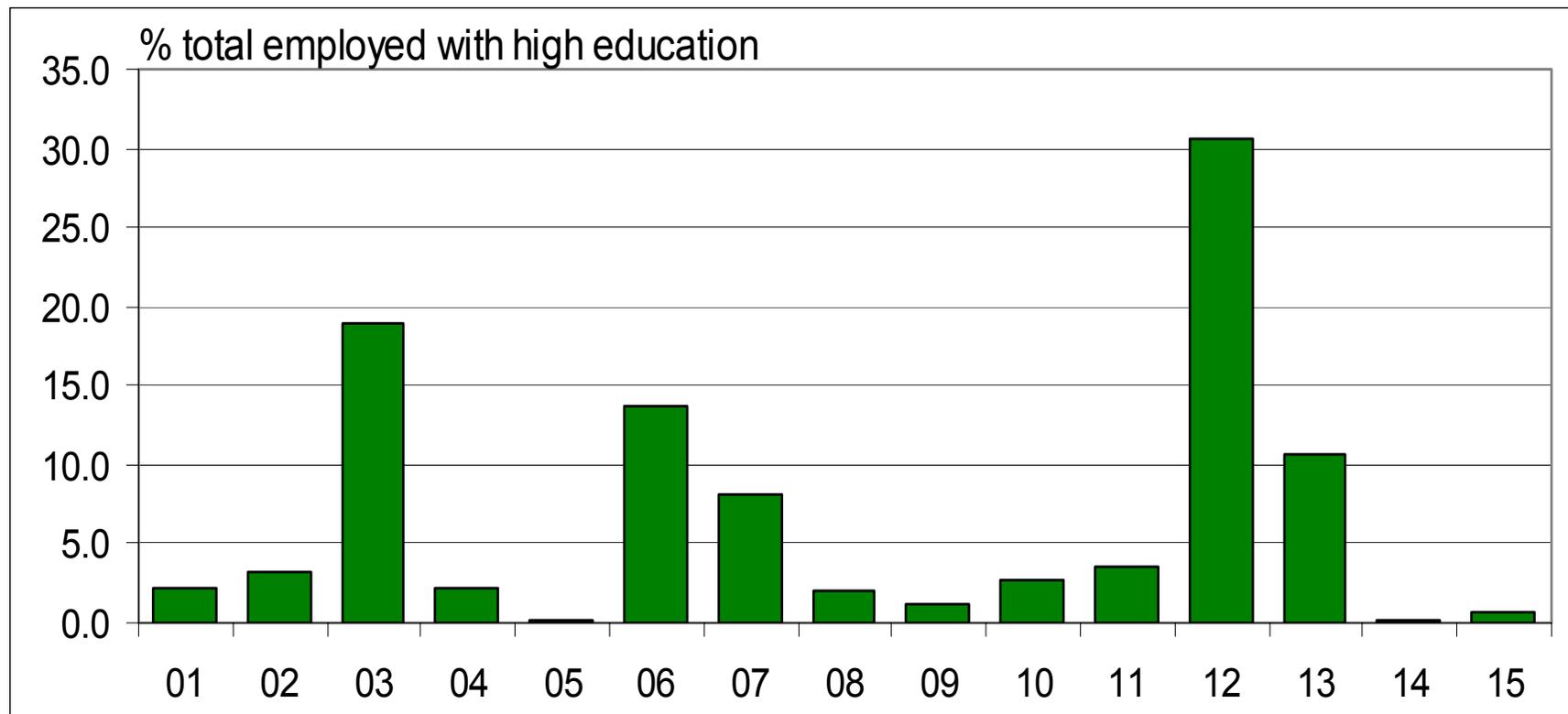12 Social sciences, business, law

13 Teacher training, education

% Division

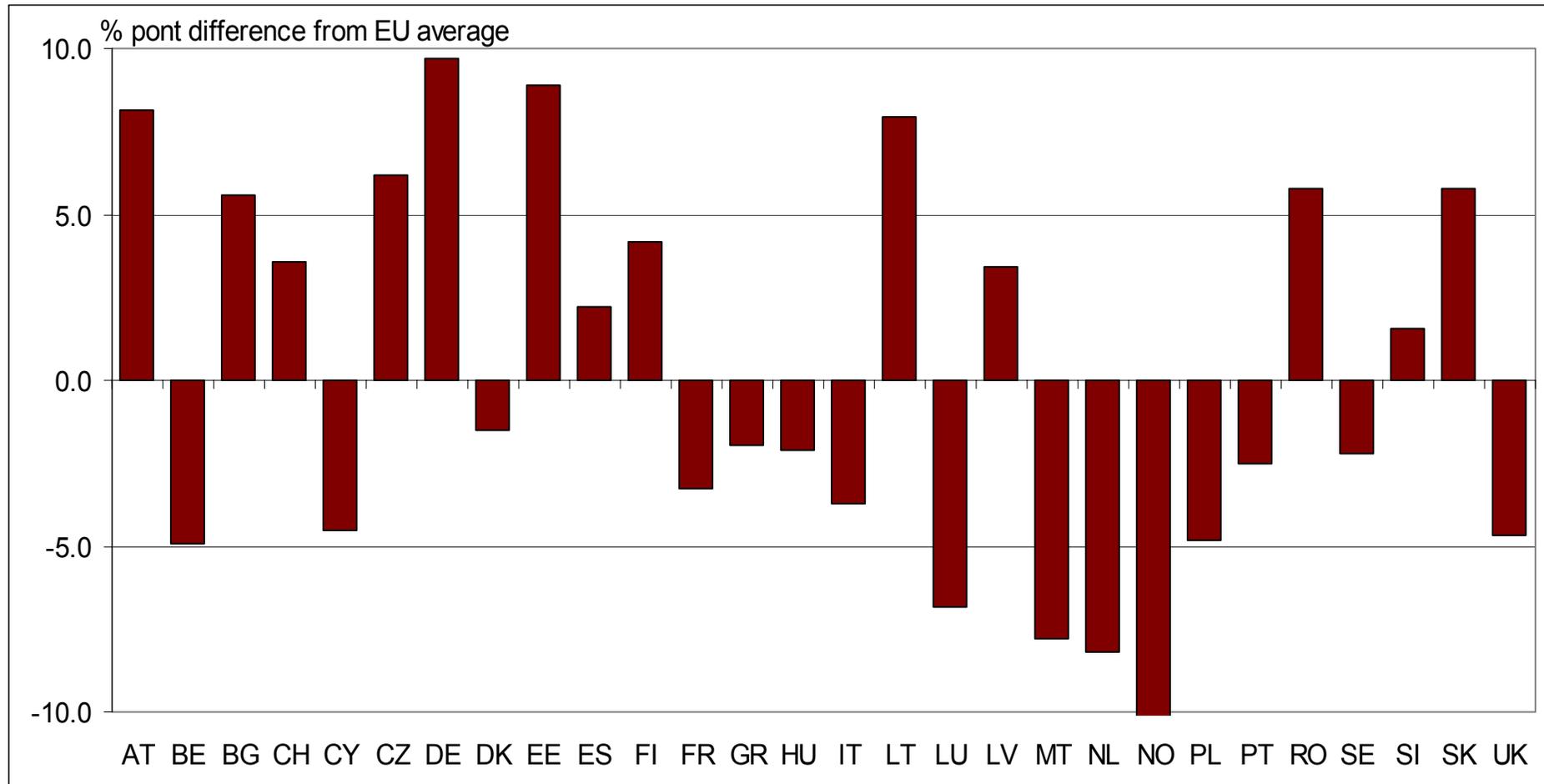|	FS01	FS03	FS05	FS06	FS07	FS11	FS12	FS13
EU	4.0	38.3	14.6	6.2	2.6	8.9	21.7	1.1
DE	2.6	36.5	6.7	8.7	2.3	9.7	30.5	1.6
FR	5.0	39.0	2.0	6.2	6.5	6.2	29.4	0.0
IT	2.8	30.7	20.8	1.0	1.5	6.3	33.5	0.8
UK	2.0	32.0	2.0	17.1	3.8	16.4	19.9	2.9
ES	0.7	14.1	63.4	5.0	0.7	3.6	11.0	0.4
NL	4.0	22.8	10.6	16.2	2.3	12.4	25.3	3.9
BE	2.4	38.6	21.9	7.0	2.4	10.9	13.6	1.0
AT	4.5	37.5	9.0	4.5	1.6	13.4	27.6	1.4
SE	3.0	33.7	17.9	11.2	4.6	8.4	18.7	1.0
GR	0.4	13.2	65.2	3.2	2.9	4.4	6.9	0.2
PT	1.0	10.1	8.5	2.8	22.1	4.7	24.4	2.1
PL	9.3	48.7	12.6	3.5	0.3	12.8	10.0	0.4
CZ	4.7	55.1	5.1	4.4	1.6	9.6	17.0	1.4
HU	3.4	50.2	13.8	4.9	0.9	8.3	16.9	0.2
RO	4.6	61.9	15.3	2.6	1.6	4.1	4.5	1.3
BG	4.0	51.6	33.0	0.2	0.7	3.9	6.4	0.0

Tertiary education

01 Agriculture and veterinary
02 Computer science
03 Engineering, manufactg,, construction
04 Foreign languages
05 General programmes
06 Health and welfare
07 Humanities, arts
08 Life sciences

09 Mathematics, statistics,
10 Physical sciences
11 Services
12 Social sciences, business, law
13 Teacher training, education
14 Computer use
15 Computing

% total employed with high education

Code	Value
01	2.2
02	3.2
03	19.0
04	2.2
05	0.2
06	13.7
07	8.1
08	2.1
09	1.2
10	2.7
11	3.5
12	30.7
13	10.6
14	0.2
15	0.7

Tertiary education
Difference in share of engineering, manufacturing, construction from EU average

% pont difference from EU average

AT BE BG CH CY CZ DE DK EE ES FI FR GR HU IT LT LU LV MT NL NO PL PT RO SE SI SK UK

Tertiary education
Difference in share of social science, business, law from EU average

% point difference from EU average

Chart showing % point difference from EU average by country: AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GR, HU, IT, LT, LU, LV, MT, NL, NO, PL, PT, RO, SE, SI, SK, UK
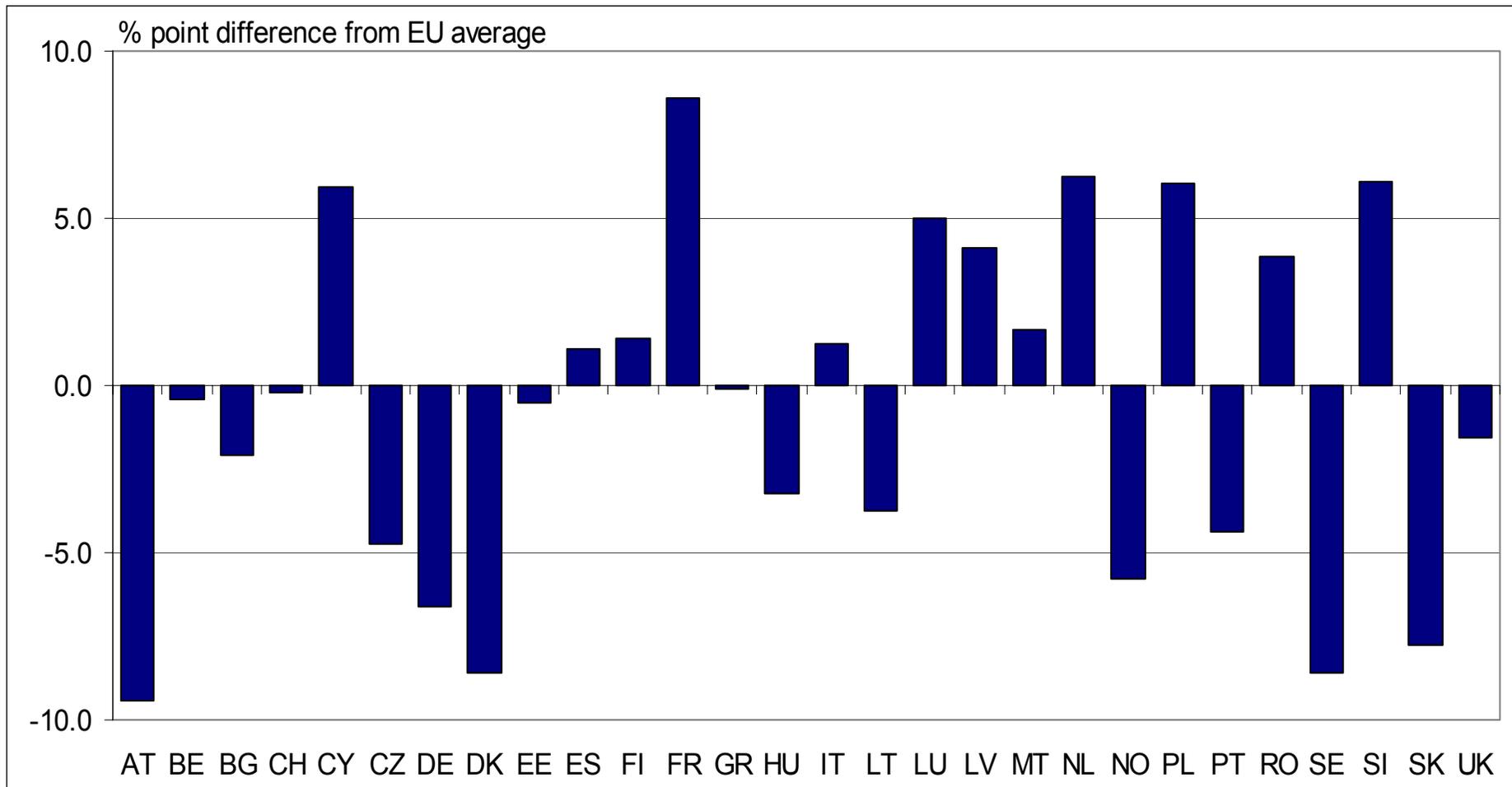
Occupation breakdown

- ISCO 3-digit – 116 separate occupations instead of 27 of ISCO 2-digit

- Increases potential cells for each country for each year to 167,040

- In practice, employment not evenly split across all occupations but concentrated in a few even if ignore sectors …

Occupation breakdown (ISCO 2 to 5)

ISCO2	ISCO1 % of Total	ISCO2 % of ISCO1	1	2	3	4	5	6	7	8	
					3rd digit of ISCO3 code						
21 Engineers	14	25	5	1	31	64					(Computing, engineers)
22 Health professionals		13	11	68	22						(Health profs, nurses)
23 Teachers		29	13	44	30	3		11			(Secondary, primary)
24 Other		33	34	14	3	20	15	2	13		(Business)
31 Engineers	16	23	62	20	7	3		9			(Engineers)
32 Health		16	10	44	47						(Health assoc, nurses)
33 Teachers		7	18	37	13	30					(Pre-primary)
34 Other		53	36	7	31	9	2	8	8		(Finance, admin)
41 Office clerks	11	81	18	15	18		7	(9)	35		(Secretaries,num,trans, oth)
42 Customer serv clerks		19	52	48							(Cashiers,client info)
51 Personal serv	14	62	2	35	38	11		14			(Hotel/rest, care)
52 Sales		38	0	100	0						(Shop assistants)

Occupation breakdown (ISCO 6 to 9)

ISCO2	ISCO1 % of Total	ISCO2 % of ISCO1	3rd digit of ISCO3 code									
			1	2	3	4	5	6	7	8	9	
61 Agricultural	4	96	34	10	49	3		2				(Mart gard,crop/animal)
71 Building	14	44	3	45	38	14						(Building frame, finishers)
72 Metal, machine		35	25	17	37	21						(Welders,tool-makrs,fitters,electric)
73 Precision,craft		4	37	17	7	38						(Precision,craft printers)
74 Other		15	35	27	31	6						(Food, furniture, clothing)
81 Plant operat	8	12	6	21	7	16	22	14	11			(Metal.chemical)
82 M/c oprtr,assembler		36	14	4	8	2	4	15	13	28	12	(Assem)
83 Drivers		51	5	73	21	1						(Motor vehicle driver, mob operate)
91 Sales/service	10	62	4	0	64	10	17	4				(Cleaners)
92 Agric labourers		8	100									
93 Labourers		30	27	40	34							(Construction, manufacturing, transport)

Occupation breakdown by sector

- Can reduce number of potential cells by moving from 60 NACE 2-digit sectors to 35 E3ME sectors – but still 104,832

- In practice, for some 80% of these cells, there is no entry for any year – i.e. occupation does not exist in sector

- Of the remaining cells, around 30-50 % (depending on country) have entry in only one year, between 12-20 % in only two years ...

- Accordingly, only 10-20% have entries for all years – i.e. 2 - 4% of total potential cells

- Issue – how to treat data when cells have entries only for some years?

- Cells could be zero in 2008 but positive in some earlier years, or positive in 2008 but zero in some earlier years

- Could extrapolate to fill in missing cells, but trends not always clear – could be e.g. U-shaped

Adjusting the data

- Adjusting data for missing values gives rise to similar problems as for breaks in the series - and adds to those already encountered with ISCO 2-digit data

- Data missing for any year is imputed using data from year before and after

- Where only 2-digit (or 1-digit) ISCO data available allocated pro rata to 3-digit categories

- Sequencing of adjustment affects results

- Sequencing as before - SEX-AGE-ISCED-ISCO-NACE – i.e. start by adjusting SEX within AGE-ISCED-ISCO-NACE, then AGE within ISCED-ISCO-NACE

- Start with 2008 and work back

- Results after adjustment differ from LFS total – inevitable, not possible to be compatible with LFS total AND adjust at detailed level

Dataset produced

- Dataset internally consistent for each country

- Relatively small adjustments and differences from LFS totals for variables without breaks – i.e. sex and age

- Larger adjustments for other variables with breaks - ISCO and ISCED especially

- Disaggregation to ISCO 3-digit adds to problems, as indicated

- Option to aggregate some of variables – e.g. – for some purposes – depends on use to which data put

- Danger that forecasts produced affected by data adjustment method

- Problems even more acute for field of study data – short time series limits forecasting possibilities. Initial solution probably to use breakdown for latest year (2008)