Alphametrics (AM)

Alphametrics Ltd

Forecasting skill supply and demand in Europe: *Developing a consistent dataset*

Paper presented at *Skillsnet* technical workshop

9-10 June 2009, Thessaloniki, Greece

Robert Stehrer (wiiw)

Terry Ward (Alphametrics)

Aim: **to construct a set of consistent data**

broken down **on the labour demand side** by:

- employment in different occupations and sectors

- with details of education attainment levels

- plus sex and age of person concerned

on the labour supply side by**:**

- sex and age in more detail

- educational attainment level

- employment status

- occupation of those employed

*A*lphametrics

Data source

Data in this detail can only come from the **Labour Force Survey**

– provides comparable data for all EU countries plus others

– in sufficient detail for purposes of study

– but not in published microdata set – too aggregated

– Instead use of special extractions by Eurostat from full microdata

– these relate not to individuals as such but all individuals who share a set of common characteristics

Alphametrics

Starting-point for demand side

Data for EU27 countries+CH for 1995-2007 (if not for all) for:

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

- 27 occupations (ISCO 2-digit level) (potentially 115 or so at ICSO 3-digit level)

- 3 educational categories (high, medium, low) (potentially 12 or so at ISCED 2-digit level)

- 4 age groups (15-54, 55-64, 65-69, 70+)

- men and women

Gives potentially 38,810 data cells for each country each year

Alphametrics

Future possible data extension

Data on demand side could be extended in future by:

- shifting from ISCO 2-digit to 3-digit (115 or so occupations)

- shifting from broad education levels to ISCED 2-digit (12 or so levels)

- adding field of study (15 fields)

- so increasing data cells ….

- to a mere 9,936,000 for each country each year

Alphametrics

Data on supply side

Data for EU27 countries plus CH for 1995-2007 for:

- 56 age groups (55 individual years of age 15-69 plus 70+)

- 3 education categories

- 3 employment states (employed, unemployed, inactive)

- 27 occupations for those employed

- men and women

Means only around 15,000 data cells

Alphametrics

Problem: LFS sample survey with small sample size in some countries – raises issues of data reliability

Sample size, 2007

|	Av. Nunber per quarter	% population 15-74		Av. Nunber per quarter	% population 15-74
BE	22,100	0.3	LU	4,000	1.1
BG	26,000	0.4	HU	60,100	0.8
CZ	49,300	0.6	MT	5,600	1.8
DK	20,700	0.5	NL	83,400	0.7
DE	131,300	0.2	AT	37,900	0.6
EE	4,800	0.5	PL	40,000	0.1
IE	59,100	1.8	PT	33,100	0.4
EL	56,100	0.7	RO	49,900	0.3
ES	126,300	0.4	SI	13,900	0.9
FR	63,900	0.1	SK	21,900	0.5
IT	127,200	0.3	FI	34,400	0.9
CY	7,300	1.3	SE	51,300	0.8
LV	7,500	0.4	UK	86,900	0.2
LT	12,400	0.5	CH	43,400	0.8

Further data problems

- Data improvements over time

- Missing sectors and occupations in some countries in some years – employment zero in one year, positive in next, then back to zero

- Documented breaks in series (15 cases over period 1995-2007)

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

- No answers or blanks

Alphametrics

Breaks in series – illustration

An example is Austria between 2003 and 2004:

- ISCO 34 (Other associate professionals),

- ISCO 52 (Sales persons and demonstrators)

- increase in ISCO 34 from 187,200 to 474,100

- reduction in ISCO 52 (Sales persons and demonstrators) from 297,400 to 159,800 plus smaller reductions in other occupations

- Substantial shift in occupational pattern…

Alphametrics

Division of employment by occupation, Austria 2007 and 2003
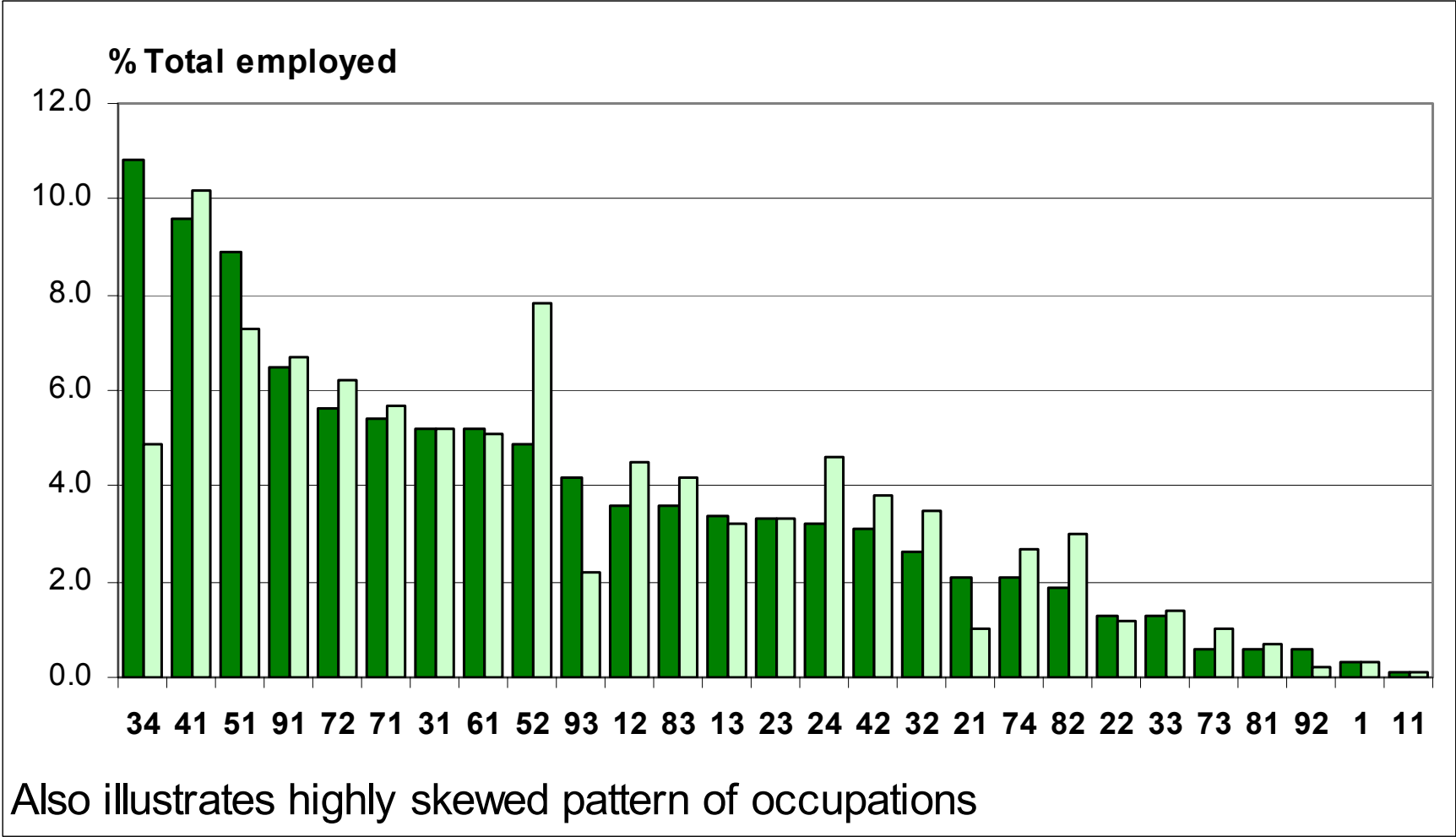
Strategy for data adjustment

- Missing cells set to zero

- 'No answers' allocated pro rata (but results depend on sequencing – i.e. order in which dimensions adjusted)

- Adjustment for breaks more problematic

- Even if breaks known, no clear solution – no data to provide link

- In a few cases, offsetting shifts in a limited number of sectors or occupations

- But in many more cases, shifts spread across sectors or occupations

- And in some cases, both sectors and occupations affected

Alphametrics

Data adjustment strategy (cont.)

Different strategies tried:

– Adjustment for pair-wise shifts can work in some cases

But cannot be a general strategy

– Adjustment of levels (numbers employed) tends to lead to volatile results

– **Preferred strategy:**

to adjust shares in sequence:

- NACE-ISCO-ISCED-AGE-SEX

- i.e. start by adjusting NACE shares of total employment

- then ISCO shares within NACE, ISCED shares within NACE-ISCO and so on

Alphametrics

Method of adjustment

Following NACE-SICO-ISCED-Age-Sex sequence:

- Start with 2007 and work back

- Determine threshold for annual change – can be sector specific

- Calculate average (or median) change in shares

- Apply actual change in shares or average (median) if above/below threshold

Alphametrics

Problems of applying preferred strategy

- Problems with cells being zero in 2007 but positive in some earlier years

- Problems with series being positive in 2007 but zero in some earlier years

- Trends in shares not always clear – e.g. U-shaped

- Sequencing of adjustment matters (as with missing values)

- Results after adjustment might differ from overall LFS data

  – e.g. share of women in total employment might differ from original - Should we then make further adjustments to correct? But difficult to be fully compatible with original LFS series AND adjust at micro level

- Austrian case again provides illustration…

*A*lphametrics

Austria: men and women employed before and after ISCO adjustment
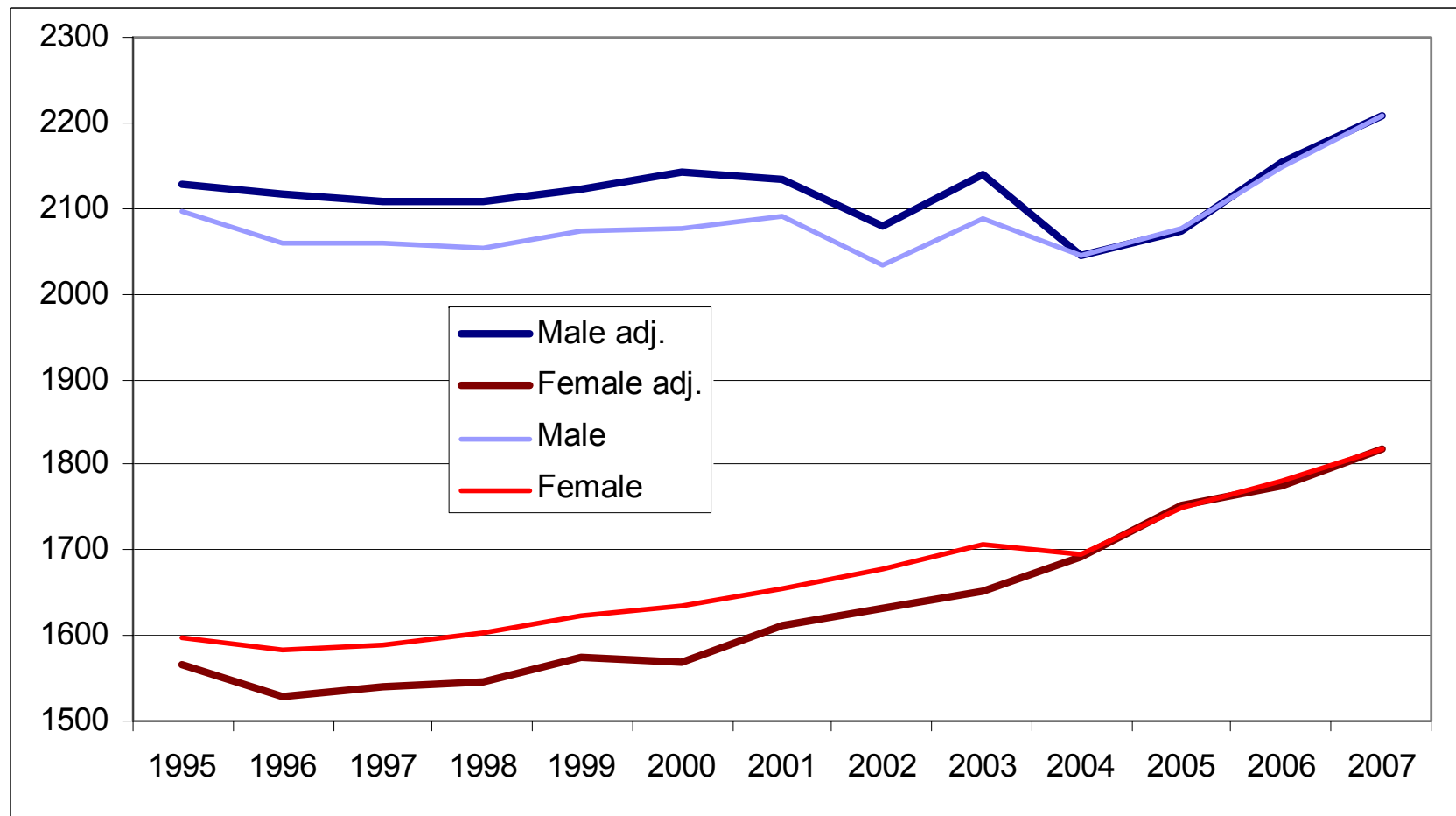
Issues related to modelling strategy

- – At what level and how closely should attempt be made to match original data for totals?

- – How should very small cells be dealt with?

- – How should cells that are positive only in some years be dealt with?

- – Does smoothing of series for adjustment purposes matter for modelling?

- – If accept differences in aggregates from LFS – e.g. men and women – does labour supply data need to be adjusted too to match?

Alphametrics