1 Sample selection

I am restricting the sample to people:

- Aged 20-60.
- Not in the armed forces.
- With non-missing occupation.
- Non-missing educational level.
- At least one year of work after leaving full-time education.

2 Regarding weighting

- The dataset comes with sampling weights.
- According to the documentation the weights apply to people aged 20-60 years old. Even within this sample some observations in 2006 and 2012 have missing weights. The documentation does not say why this could happen.
- Without weighting, the dataset has too few mean and young people.
- For now, all results are weighted.

Table 1: Share with missing weights

	Year				
	1997	2001	2006	2012	Total
Share missing	0.00	0.00	0.14	0.00	0.06

Table 2: Representativeness of the SES sample, 2006-2012

Demographic	(1) Unweighted	(2) Weighted	(3) LFS
2006			
Share female	0.49	0.47	0.46
Breakdown by age bracket			
20-29	0.16	0.22	0.22
30-39	0.24	0.25	0.27
40-49	0.28	0.27	0.28
50-60	0.26	0.23	0.23
2012			
Share female	0.53	0.46	0.47
Breakdown by age bracket			
20-29	0.16	0.23	0.23
30-39	0.24	0.25	0.24
40-49	0.31	0.28	0.29
50-60	0.28	0.24	0.24

Note: weights correspond to `weightall` variable. The third column comes from table A1 in http://doc.ukdataservice.ac.uk/doc/7645/mrdoc/pdf/7645_ses_technical_briefing_may_2014.pdf

Table 3: Summary statistics people with missing weights

Variable	Mean	Std. Dev.	Min.	Max.	N
AGE OF RESPONDENT	41.786	10.877	20	60	1037
SEX OF RESPONDENT	0.518	0.5	0	1	1037
EDUCATION LEVEL HELD	2.555	1.429	0	4	1036
DATASET	2006	0	2006	2006	1037

3 Occupations to keep

- Not all occupations appear in all years. It's unclear to me if this is a result of new jobs appearing, or whether this is a result of people in this occupations not being sampled.
- For now I restrict everything to occupations appearing in all waves.

Table 4: Occupation counts

	1997	2001	2006	2012
Number of occupations	180	219	232	207
Occupations appearing in all waves				
Number of occupations	154	154	154	154
Employment share	0.94	0.92	0.89	0.89

4 Selection of skill

- I am using the variable `edlev` as a measure of education. This is the one coded most simply + it is consistent across all years.
- It shows the level of NVQ qualification of the worker.
- Raw classification is given by:
  - No qualification.
  - Level 1
  - Level 2
  - Level 3
  - Level 4-5
- I group the categories based on how similar their occupational employment distribution is.
- This results in the grouping:
  - **Low-skill:** no qualification-Level 1.
  - **Mid-skill:** Level 2-3.
  - **High-skill:** Level 4-5.

Table 5: Welch index of the occupational employment by education level

	0	**1**	**2**	**3**	**4**
0	1.00				
1	0.60	1.00			
2	0.16	0.27	1.00		
3	-0.13	0.05	0.27	1.00	
4	-0.64	-0.67	-0.72	-0.45	1.00

Table 6: Correlation in occupation employment distribution by education level

	0	**1**	**2**	**3**	**4**
0	1.000				
1	0.765	1.000			
2	0.586	0.725	1.000		
3	0.445	0.630	0.797	1.000	
4	-0.031	0.098	0.250	0.329	1.000

5 Boundary jobs

- I am defining the boundary jobs by year.
- I define a job as follows:
  - For each occupation, I first take the two skills with the highest employment share and compute the total employment shares accounted by these two skills: $sh_1 + sh_2$.
  - The job is in the border if the top employment share accounts for less than a given threshold R, i.e.

$$\frac{sh_1}{sh_1 + sh_2} < R$$

- I use 55% and 60% as possible thresholds.
- 60% seems like a reasonable choice.
- Border jobs are changing from year to year. Only 4 jobs are in the boundary over the whole period under the 60% definition.
  - Almost all jobs (91) are in the boundary at some point in the period.
  - These jobs are in the three types of borders.

Table 7: Share of boundary jobs

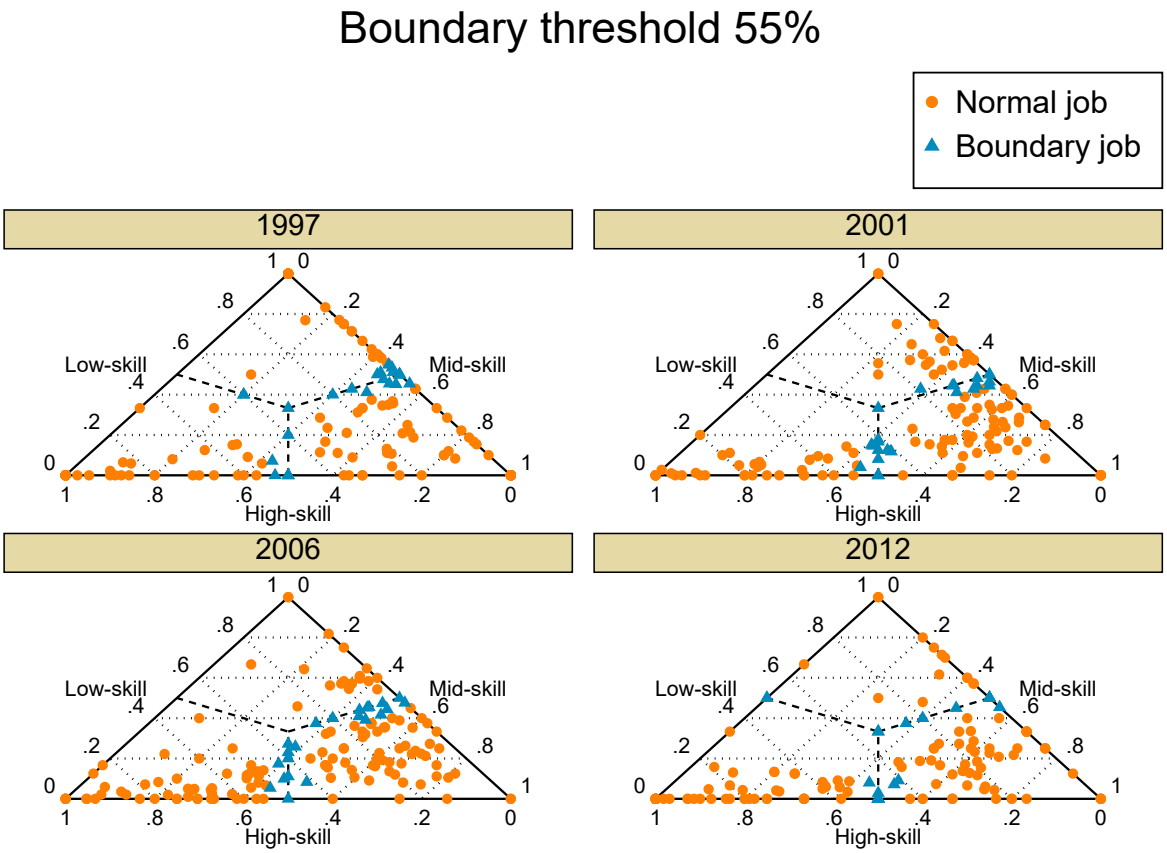
	(1)	(2)	(3)	(4)
Boundary threshold	**1997**	**2001**	**2006**	**2012**
55%	0.17	0.16	0.14	0.16
60%	0.24	0.29	0.27	0.24
65%	0.39	0.44	0.50	0.34
70%	0.44	0.58	0.58	0.48
Observations	154	154	154	151

Table 8: Count of border jobs by year

	Year			
Border type	**1997**	**2001**	**2006**	**2012**
55% threshold				
Low-High	3	1	0	4
Low-Mid	17	11	12	10
Mid-High	6	12	10	10
60% threshold				
Low-High	3	1	1	4
Low-Mid	25	25	21	15
Mid-High	9	18	19	17

Figure 1: Job classification under different boundary definitions

(a)



(b)