

# Modelling historical growth in adult educational attainment

FE and Skills Analysis, Department for Business, Innovation and Skills

## 1 Summary

- 1.1 This paper sets out the work BIS have carried out to model components of historical growth in attainment at Levels 2+, 3+ and 4+.
- 1.2 As a result of this work, we now better understand the quality of the modelled historical data that were originally used as a comparator in the July 2009 discussion paper on [“Measuring adult educational attainment using the Labour Force Survey”](#). We have also extended the analysis to consider growth in the Level 4 and above series. The modelled growth estimates for L2+ and L3+ have been improved and an assessment made of the sensitivity of the modelled estimates to changes in the underlying data.
- 1.3 The measure of uncertainty produced around the Level 2 and above series is helpful in interpreting the data, which seem to be a good fit with the unadjusted LFS data until around 2006. The Level 3 and above and Level 4 and above series are considered to be more approximate, due to more assumptions being required in the modelling to ensure a good fit (see sections 6 and 7 for more details), but we have included the results from these in the interests of transparency.
- 1.4 This further work has confirmed that the use of the historical modelling work should be restricted to providing some understanding of the trend of attainment in the population. It cannot provide a measure of the level without reference to some point in history on the LFS. Nonetheless, where the modelling suggests a change in the rate of growth, we believe this can provide an insight to how one might expect the trend in the LFS should be changing.
- 1.5 Taking the limitations of the different models into account, the revised modelled growth data suggest a broadly similar level of growth between 2006 and 2008 as that arising from the adjusted LFS estimates yielded from research that has been recently conducted.
- 1.6 A summary of the different growth series is set out below:

**Table 1: Percentage point growth in modelled historical series and adjusted LFS**

|                   | Estimated growth 2001 to 2006 | Estimated growth 2006 to 2008 | Growth in adjusted LFS 2006 to 2008 |
|-------------------|-------------------------------|-------------------------------|-------------------------------------|
| Level 2 and above | 3.9pp                         | 2.3pp                         | 2.2pp                               |
| Level 3 and above | 4.0pp                         | 1.9pp                         | 2.2pp                               |
| Level 4 and above | n/a                           | 1.3pp                         | 1.4pp                               |

## 2 Context for the work to model historical growth in attainment

2.1 Since 2001, the Labour Force Survey (LFS) has been used to calculate, and set targets for, the highest level of qualification held by adults of working age in England. LFS estimates of educational attainment are adjusted by BIS prior to their publication in a Statistical First Release (SFR). These estimates are referred to throughout this paper as SFR estimates, to distinguish this use of the LFS data from the adjusted method newly proposed by RM Data Solutions.

**Table 2: March 2009 SFR estimates of Level of highest qualification held by working age adults<sup>11</sup> in England 2001-8**

|      | <b>Total<br/>(thousands)</b> | <b>% Level 4<br/>and above</b> | <b>% Level 3<br/>and above</b> | <b>% Level 2<br/>and above</b> | <b>No<br/>qualifications</b> |
|------|------------------------------|--------------------------------|--------------------------------|--------------------------------|------------------------------|
| 2001 | 28,466                       | 25.2                           | 44.7                           | 65.0                           | 15.1                         |
| 2002 | 28,603                       | 26.2                           | 46.0                           | 66.4                           | 14.1                         |
| 2003 | 28,763                       | 27.2                           | 47.0                           | 67.2                           | 13.7                         |
| 2004 | 28,982                       | 27.9                           | 47.5                           | 67.6                           | 13.4                         |
| 2005 | 29,234                       | 28.5                           | 48.4                           | 68.9                           | 12.6                         |
| 2006 | 29,425                       | 30.0                           | 49.4                           | 69.9                           | 12.2                         |
| 2007 | 29,601                       | 30.9                           | 50.6                           | 70.7                           | 11.4                         |
| 2008 | 29,803                       | 31.2                           | 50.8                           | 71.2                           | 10.9                         |

2.2 The level 2/3 trajectory model is designed to help determine the number of Level 2/3 achievements needed to meet the 2011 PSA targets (of 79% of the population to hold L2 qualifications) and the 2020 PSA target of 90%). It also allows us to see the impact of different delivery scenarios on the proportion of the population holding level 2 qualifications for each year going forwards. The trajectory model works by considering all the elements which contribute to the growth in each year and making assumptions about how they will change in the future. This is discussed in detail in sections 3 below. Although the modelling approach was not originally designed for this purpose, it does provide a means to estimate past changes in the proportion of the population holding level 2 qualifications.

2.3 As set out in the July 2009 discussion paper, while BIS has over the past two years delivered very close to the number of planned achievements, the proportion of L2+ qualifications held in the population (as reported in the SFR) has not reflected the growth we would expect based on our modelling. The subsequent sections of this paper set out how BIS has modelled historical growth and how the resulting series compare with the SFR estimates.

---

<sup>1</sup> Available at: <http://www.thedataservice.org.uk/statistics/sfrmar09/summary.htm>

### 3 Modelled historical growth data for the level 2 and above attainment series

3.1 As a result of the work carried out over the last 6 months, the modelled growth estimates have been revised compared with those quoted in both the July discussion paper and Annex F of the interim research report published in October. The most recent estimates are set out in Table 3 below.

**Table 3 SFR and BIS Modelled Annual Growth in Level 2 attainment 2001 to 2008**

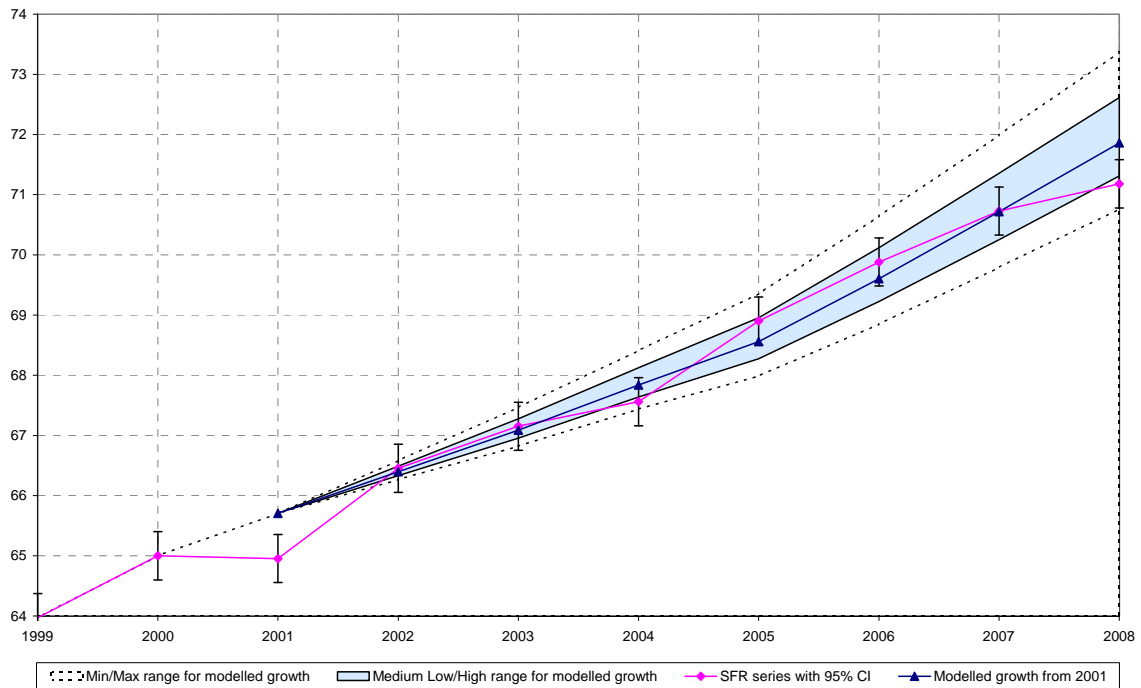
| Year | SFR percentage at Level 2+ | SFR Growth | Smoothed SFR percentage at Level 2+ | Growth in smoothed SFR series | Modelled Growth |
|------|----------------------------|------------|-------------------------------------|-------------------------------|-----------------|
| 2001 | 65.0%                      |            | 65.7%                               |                               |                 |
| 2002 | 66.4%                      | 1.4        | 66.6%                               | 0.9%                          | 0.7%            |
| 2003 | 67.2%                      | 0.8        | 67.4%                               | 0.8%                          | 0.7%            |
| 2004 | 67.6%                      | 0.4        | 68.1%                               | 0.7%                          | 0.7%            |
| 2005 | 68.9%                      | 1.3        | 68.9%                               | 0.8%                          | 0.7%            |
| 2006 | 69.9%                      | 1.0        | 69.8%                               | 0.9%                          | 1.1%            |
| 2007 | 70.7%                      | 0.8        | 70.6%                               | 0.9%                          | 1.1%            |
| 2008 | 71.2%                      | 0.5        | 71.4%                               | 0.8%                          | 1.2%            |

3.2 Between 2007 and 2008, SFR estimates suggest that the working age population with qualifications equivalent to level 2 or above increased by 0.46 percentage points. However, BIS models suggest a rate of growth of 1.2 percentage points.

3.3 Historically, there has been a degree of volatility in SFR growth which is sensitive to various types of error. Consequently, the SFR data series has been smoothed, using Holt-Winters exponential smoothing applied to a quarterly series starting in 1997, and growth in this series has been compared with the modelled growth. In recent years, the modelled growth appears to have been higher than growth in both the smoothed SFR series and the basic SFR series. Perhaps coincidentally, this has happened in the same period as initiatives, such as Train to Gain (TtG), designed to raise level 2 attainment.

3.4 The modelled estimates require a number of assumptions to be made and we have performed sensitivity analysis based on varying these assumptions. Figure 1 shows a set of modelled trajectories with additional lines corresponding to high and low values of parameters. The 'minimum' and 'maximum' series are formed by assuming all parameters are at their most extreme plausible values. They give an envelope of plausible predictions. Since it is unlikely that all parameters would simultaneously be at their extreme values, we have also plotted lines corresponding to parameters midway between the best estimate and the extreme.

**Figure 1: Comparison of modelled and SFR series for L2+**



3.5 The data underlying Figure 1 are set out below.

**Table 4: Sensitivity range for modelling of Level 2 and above attainment**

|                  | Estimates of growth from 2001 to 2006 | Estimates of growth from 2006 to 2008 |
|------------------|---------------------------------------|---------------------------------------|
| Minimum          | 3.1pp                                 | 1.9pp                                 |
| Medium low       | 3.5pp                                 | 2.1pp                                 |
| Central estimate | 3.9pp                                 | 2.3pp                                 |
| Medium high      | 4.4pp                                 | 2.5pp                                 |
| Maximum          | 4.9pp                                 | 2.7pp                                 |

3.6 While the chart above presents growth in terms of changes in the level of attainment within the population, this can only be achieved by using the SFR estimates as a starting estimate of level, and modelling growth from that point forwards. Historically, as for any sample survey, there has been a degree of variability in the SFR estimates. Consequently, when attempting to show a modelled estimate of level, the SFR data series has been smoothed using double exponential smoothing. The modelled series above has been constructed using a value of 65.7% from the smoothed SFR series as the initial condition.

3.7 Because estimates of level are sensitive to the choice of starting point along the SFR time series, and the choice of smoothing method (and indeed whether to smooth), it is recommended that the modelled data are only used to estimate growth in the attainment series, rather than the absolute level at any point in time.

## 4 Components of Modelled Growth in the Level 2 and above historical model

4.1 The modelled growth can be attributed to the effects of several elements:

- The effect of age cohorts entering and leaving the working age population each year, i.e. 59 year old women and 64 year old men retiring each year and being replaced by 19 year olds, reaching working age.
- The effect of migration, made up of immigration and emigration.
- Deaths within the working age population.
- Level 2 achievements being gained in the population specifically broken down into:
  - Publicly funded First Full Level 2 achievements.
  - Privately funded First Full Level 2 achievements.
  - Additional part level 2 achievements combining to make a first full level 2 and learners achieving level 4 qualifications without first being qualified to level 2.

4.2 Each of the components is modelled or estimated with varying degrees of confidence. Due to the fact that the age ranges of the population are different each effect is modelled on males and females separately and then aggregated into a combined total.

4.3 A fundamental assumption is that definitions of full qualifications are the same between administrative data sources and the LFS. Potential problems may be around combinations of partial qualifications to make up a full qualification.

4.4 The annual growth can be broken down into components as follows:

**Table 5: Modelled components of growth in percentage point terms of proportion of population at level 2 or above, Based on actual Labour Force Survey values.**

| Year | Modelled growth | Growth over previous year from each component |           |           |                    |                     |                   |
|------|-----------------|---|-----------|-----------|--------------------|---------------------|-------------------|
|      |                 | Population Replacement                        | Mortality | Migration | Public Up-skilling | Private Up-skilling | Other Up-skilling |
| 2001 |                 |   |           |           |                    |                     |                   |
| 2002 | 0.71            | 0.44  | -0.12     | 0.15      | 0.09               | 0.11                | 0.05              |
| 2003 | 0.71            | 0.44  | -0.13     | 0.14      | 0.09               | 0.11                | 0.05              |
| 2004 | 0.73            | 0.39  | -0.10     | 0.22      | 0.10               | 0.09                | 0.04              |
| 2005 | 0.74            | 0.40  | -0.09     | 0.16      | 0.11               | 0.12                | 0.04              |
| 2006 | 1.1             | 0.55  | -0.12     | 0.17      | 0.32               | 0.17                | 0.05              |
| 2007 | 1.1             | 0.41  | -0.13     | 0.23      | 0.36               | 0.15                | 0.05              |
| 2008 | 1.2             | 0.39  | -0.12     | 0.26      | 0.48               | 0.11                | 0.05              |

4.5 The individual components of growth are described in detail in the next few sections, which examine the sensitivity to changes in the underlying data and assumptions. Each section provides a description of the data used, any limitations of the data, and assessments of the sensitivity of the model to possible uncertainties in the data.

## 5 Sensitivity analysis of components of Modelled Growth

5.1 Some of the components of growth have a degree of uncertainty. BIS have tried to use the most reasonable values in the modelling, and have produced sensitivity tables to examine the impact on the total modelled growth of changing each component to its maximum and minimum realistic values. This assumes that components are independent of each other. Note that no sensitivity analysis has been carried out for the level 3 model. This is because the gap between what is predicted and what is observed is so large that sensitivity analysis would be meaningless.

5.2 Table 2 (last two columns) above shows that there is quite a good fit between the SFR smoothed growth and the modelled growth based on the smoothed SFR) for 2001-2005 by slight alterations to single parameters. However, from 2005 to 2008, the modelling shows higher growth than the SFR.

**Table 6: Modelled components of growth in percentage point terms of proportion of population at level 2 or above – for the periods 2001-2005 and 2005-2008.**

| Period    | Smoothed SFR growth | Modelled growth | Growth over previous year from each component |           |           |                     |                      |                    |
|-----------|---------------------|-----------------|---|-----------|-----------|---------------------|----------------------|--------------------|
|           |                     |                 | Popula - tion Replace- ment                   | Mortality | Migration | Public Up- skilling | Private Up- skilling | Other Up- skilling |
| 2001-2006 | 4.1                 | 4.0             | 2.1   | -0.6      | 0.8       | 0.7                 | 0.6                  | 0.2                |
| 2006-2008 | 1.6                 | 2.3             | 0.8   | -0.3      | 0.5       | 0.8                 | 0.3                  | 0.1                |

5.3 We have separated these two time periods in the sensitivity tables. An example is shown below.

**Table 7: Example of a sensitivity analysis table**

|                            | Numbers of Other upskilling | SFR smoothed growth minus modelled growth 2001-2006 | SFR smoothed growth minus modelled growth 2006-2008 |
|----------------------------|-----------------------------|---|---|
| Maximum realistic scenario | 40,000                      | <i>-0.24ppts</i>                                    | <i>- 0.72ppts</i>                                   |
| Current estimate           | 20,000                      | + 0.10ppts  | <i>- 0.59ppts</i>                                   |
| Minimum realistic Scenario | 12,500                      | + 0.22ppts  | <i>- 0.54ppts</i>                                   |

5.4 These tables show the difference between the Smoothed SFR growth and the modelled growth (based on the smoothed LFS) for different scenarios for the two time periods. Positive figures indicate LFS smoothed growth is higher. Negative figures (*in red italics*) indicate modelled growth is higher. Differences of around +/-0.1 ppts show a very good match. The aim of these sensitivity is to see the impact of different realistic scenarios, and where possible to match the SFR smoothed growth with the modelled growth for 2001-2005.

**Population Replacement (effect of age cohorts entering and leaving the working age population)**

- 5.5 Each year, 59 year old women and 64 year old men reach retirement age and are effectively replaced within the working age population by 19 year olds reaching adult age.
- 5.6 As the two groups have quite different average levels of attainment, there will be an impact on the level with the population as a whole. There are four main elements to the calculation of age cohort estimates.
  - Number of 19 year olds entering the working age population
    - This is from the LFS data for the previous year.
  - Number of 59/64 year olds leaving the working age population
    - This is from the LFS data for the preceding year, the number of 64 year old men, and 59 year old women.
  - Qualifications of 19 year olds entering the working age population
    - Based on SFR estimates from the previous year.
  - Qualifications of 59/64 year olds leaving the working age population
    - Based on SFR estimates from the previous year.

**Limitations of data sources**

- 5.7 These numbers are taken directly from the SFR estimates, and are subject to sample error as the single-age cohorts are small. While the data could be smoothed, this is likely to introduce bias.

**Table 8: Level 2+ Qualified Recruitment and Retirement from the LFS**

| Year | Level 2+    |            |         |
|------|-------------|------------|---------|
|      | Recruitment | Retirement | Net     |
| 2002 | 422,000     | 255,000    | 166,000 |
| 2003 | 434,000     | 266,000    | 169,000 |
| 2004 | 452,000     | 262,000    | 190,000 |
| 2005 | 471,000     | 257,000    | 213,000 |
| 2006 | 500,000     | 278,000    | 223,000 |
| 2007 | 490,000     | 330,000    | 159,000 |
| 2008 | 502,000     | 345,000    | 157,000 |

**Sensitivity**

- 5.8 Given the volatility in the SFR estimates, the sensitivity analysis performed was limited to considering the impact of using alternative DCSF estimates of qualifications of 19 year olds entering the working age population.
- 5.9 DCSF use matched data from a variety of administrative sources to produce attainment rates for younger people. Their published statistics suggest a rapid growth in attainment for those reaching age 19, compared with a relatively flat trend underlying the rather volatile series from the SFR. Earlier years in the DCSF series will suffer from some of the administrative records being incomplete, leading to attainment rates being underestimated. We have therefore used a constant rate prior to 2006 in demonstrating the impact of using these data.

5.10 The impact of making this change is more substantial in the 2001 to 2005 period, with the modelled growth now almost 0.5 percentage points below that shown by the SFR. However, the usefulness of the DCSF data in this manner prior to 2006 is questionable and this variation is not carried forward to other analysis in this paper.

**Table 9: impact of using DCSF attainment rates for 19 year olds**

|   | <b>SFR smoothed growth minus modelled growth 2001-2006</b> | <b>SFR smoothed growth minus modelled growth 2006-2008</b> |
|---|--|--|
| <b>Current estimate</b>   | <b>0.33</b>  | <b>-0.98</b>   |
| <b>Estimate using DCSF figures for attainment of 19 year olds</b> | <b>0.48</b>  | <b>-0.92</b>   |

### **Mortality (effect of deaths in the working age population)**

5.11 Each year a number of deaths occur in the working age population, this component estimates the number of deaths within the population qualified to at least level 2 or level 3.

5.12 Mortality rates are calculated for the working age population using the number of deaths registered per year as a proportion of the working age population (calculated separately for males and females).

#### **Limitations of data sources**

5.13 The number of deaths recorded should be highly accurate. However, qualification levels are not recorded on death certificates, and we currently assume that mortality of L2 holders is the same as the total working age population.

5.14 Single age band results from the SFR shows that people at the younger end of the working age range are higher qualified on average than at the older end of the working age range. We also know that more deaths occur at the older end of the age range so this could indicate that the qualification level of the people dying may be lower than assumed in the model. While this suggests overall growth shown in the model may be a slight under-estimate, the differences will be marginal in terms of the overall scale of the growth being modelled.

#### **Sensitivity**

5.15 Mortality rates for qualified and unqualified people across the working age population will be different, due to:

1. Differences between age distributions – unqualified population skewed towards older people, particularly those in 50+ age bracket
2. Effect of economic status & wealth
3. Area of residence and type of occupancy (linked to above)
4. Differences in general health & well-being

**Table 10: Level 2 percentage mortality sensitivity**

|                            | Percentage mortality of L2+s compared to rest of population     | SFR smoothed growth minus modelled growth 2001-2006 | SFR smoothed growth minus modelled growth 2006-2008 |
|----------------------------|---|---|---|
| Maximum realistic scenario | 4% below that of those below L2s                                | + 0.06ppts  | -0.60ppts   |
| Current estimate           | Same as rest of population                                      | + 0.10ppts  | - 0.59ppts  |
| Minimum realistic Scenario | Unlikely that mortality of L2+s will be greater than below L2-s |   |   |

**Migration (effects of immigration and emigration)**

5.16 Migration figures are calculated using TIM (Total International Migration) results for England and Wales broken down into the country of citizenship.

5.17 Migration movements broken down according to the citizenship of the migrant is known in total but not split by age band. However, using the TIM age breakdown reports we can calculate the numbers immigrating and emigrating from England. In order to use this ‘citizenship migration’ data we therefore have to assume that the proportions of British and non-British migrants are the same for the working age population as for the total population.

5.18 It is assumed that anybody with a citizenship of British, European Union 15 or Old Commonwealth (immigrant or emigrant) is qualified to the same level as the working age population as reported in the SFR. This assumption is based on evidence from OECD comparisons that Britain and the EU15 are similar in qualification levels. Immigrants or emigrants who are citizens of other countries, than those listed above, are assumed to be qualified to the same level as that reported by the SFR estimates for new arrivals in the last year.

5.19 The TIM figures are derived from the IPS (International Passenger Survey) that interviews over a quarter of a million people entering or leaving the UK through the main airports, seaports and the channel tunnel.

**Limitations of data sources**

5.20 No data is available on qualifications held by emigrants, so we make the assumption that individuals of a given nationality have the same level of qualifications on entering and leaving. It is difficult to identify qualification rates for individual countries because of the small sample size and the LFS having to capture a diverse range of qualifications.

5.21 The migration statistics are also an area that has been recognised as challenging, with a quality review under way, and it is worth noting that the official immigration figures are different to LFS numbers of people

who say they have arrived in the past year. Immigrants may be hard to contact, through being single and at work, might not be resident at addresses used in the LFS sample frame, or may be unwilling to be contacted by bodies associated with Government.

5.22 We have investigated whether better sources of information on the educational attainment of immigrants and emigrants might be available. This work has failed to find a source, although we have identified research that has been completed on how better information might be collected on surveys<sup>2</sup>. While there are sources that may provide information in the future, it is unlikely that developments in these areas will be pursued in light of issues such as response burden and cost. Nonetheless areas under consideration include discussions between CLG and other government departments about the feasibility of a migration surveys; if undertaken this survey will have some educational content on immigrants.

**Table 11: LFS and TIM/GAD immigration comparison**

|                                     | Q4 01   | Q4 02   | Q4 03   | Q4 04   | Q4 05   | Q4 06   | Q4 07   | Q4 08   |
|-------------------------------------|---------|---------|---------|---------|---------|---------|---------|---------|
| LFS<br>Arrived <<br>1year ago       | 158,000 | 153,000 | 163,000 | 219,000 | 251,000 | 225,000 | 236,000 | 196,000 |
| LFS<br>Proportion<br>at level 2+    | 51.9    | 52.8    | 53.4    | 53.7    | 49.7    | 51.9    | 53.7    | 60.3    |
| TIMS/GAD<br>Number of<br>immigrants | 370,000 | 400,000 | 392,000 | 459,000 | 445,000 | 409,000 | 491,000 | 493,000 |

### Sensitivity

5.23 The sensitivity testing considered variations in the levels of qualifications for migrants, but the assumption that these were the same for those entering and exiting England was not tested. Percentage qualifications levels of migrants that are not British, EU15 or Old Commonwealth citizens are assumed to be the same as the SFR proportion at Level 2+ of new arrivals (table above). However, as the number of immigrants in the modelling is above the LFS figure and the qualification levels may be understated, this may not be the case although it is unlikely that the actual figure is more than 10% away from the SFR derived value.

**Table 12: Level 2 Percentage qualification level of migrants that are not British, EU15 and Old Commonwealth (rest of the world)**

|                            | Percentage of non-British migrants holding L2 qualifications | SFR smoothed growth minus modelled growth 2001-2006 | SFR smoothed growth minus modelled growth 2006-2008 |
|----------------------------|--|---|---|
| Maximum realistic scenario | Derived from SFR + 10%                                       | - 0.21ppts  | - 0.73ppts  |
| Current estimate           | Derived from SFR   | + 0.10ppts  | - 0.59ppts  |

<sup>2</sup> Using the Omnibus Survey to test questions on emigration. Helen Evans, Roma Chappell and Emma Wright, Office for National Statistics. Population Trends 127 Spring 2007

|                            |                                  |            |            |
|----------------------------|----------------------------------|------------|------------|
|                            | arrived in last year respondents |            |            |
| Minimum realistic Scenario | Derived from SFR - 10%           | + 0.41ppts | - 0.45ppts |

5.24 The percentage qualifications levels of migrants that are British, EU15 or Old Commonwealth citizens are assumed to be the same as the proportion at Level 2+ of the working age population. Higher levels of qualifications from OECD countries could increase this but it is unlikely to differ from the English percentage by more than 10%.

**Table 13: Percentage of British, EU15 and old commonwealth migrants holding L2 qualifications**

|                            | Percentage of British, EU15 and Old Commonwealth migrants holding L2 qualifications | SFR smoothed growth minus modelled growth 2001-2006 | SFR smoothed growth minus modelled growth 2006-2008 |
|----------------------------|---|---|---|
| Maximum realistic scenario | Same as England population -10%   | + 0.05ppts  | - 0.61ppts  |
| Current estimate           | Same as England population  | + 0.10ppts  | - 0.59ppts  |
| Minimum realistic Scenario | Same as England of population + 10%   | + 0.15ppts  | - 0.57ppts  |

5.25 The tables above show that changing the qualification levels of “rest of the world” migrants has a larger effect on the modelled numbers than changing the qualifications levels of the British, EU15 and Old Commonwealth migrants.

5.26 We have assumed that the proportions of British and non-British migrants are the same for the working age population as for the total population. This indicates that the net migration is as follows:

**Table 14: Net Migration**

|                                 | 2002    | 2003    | 2004    | 2005    | 2006    | 2007    | 2008    |
|---------------------------------|---------|---------|---------|---------|---------|---------|---------|
| British, EU15, Old Commonwealth | -26,000 | -42,000 | 2,000   | -33,000 | -50,000 | -30,000 | -37,000 |
| Rest of the World               | 147,000 | 158,000 | 199,000 | 216,000 | 198,000 | 210,000 | 218,000 |

5.27 If the assumption is incorrect and the split is not equal some of the net migration would be moved between the two groups. Table 13 shows the sensitivity if 20,000 net migration is moved between groups per year.

**Table 15: Rest of world split incorrectly from TIM figures**

|  | Assumption of number of “rest of the world” split incorrect from TIM figures (and converse for British/EU15/Old Commonwealth) | SFR smoothed growth minus modelled growth 2001-2006 | SFR smoothed growth minus modelled growth 2006-2008 |
|--|---|---|---|
|  |   |   |   |

|                            |  |            |            |
|----------------------------|--|------------|------------|
| Maximum realistic scenario | Assumed rest of the world TIM net migration -20,000  | + 0.06ppts | - 0.62ppts |
| Current estimate           | Assumed rest of the world TIM net migration          | + 0.10ppts | - 0.59ppts |
| Minimum realistic Scenario | Assumed rest of the world TIM net migration + 20,000 | + 0.14ppts | - 0.56ppts |

5.28 The above table shows that if 20,000 migrants are moved between groups per year the modelled growth does not change by more than 0.1ppt per year.

### **Public Up-skilling (achievements gained which were funded by the Learning and Skills Council)**

5.29 Aside from the effects of changes to the working age population, individuals can become more skilled as a result of training activity funded by either the public or private sector.

5.30 We have comprehensive data about publicly funded upskilling, for which there are three main elements

- Full Level 2 achievements by programme type (Further Education, Train to Gain and Apprenticeships), from ILR data
- For Full Level 2 achievements by programme type, estimates of the proportion of these achievements which are 'first' at that level by the individual. These estimates are based on the LSC's Prior Qualifications Survey. This looks at the qualifications held by a sample of LSC-funded learners using LFS-style questions.
- Jumpers rates for First Full Level 3 achievements, based on the most appropriate source by programme. Jumpers are Level 3 achievers who have not previously achieved a Level 2 qualification.

### **Limitations of data sources**

5.31 The ILR data on qualifications achieved is put through a series of validation checks when it is received by the Data Service from the provider. Validation Rules that are run against submitted data, and guidance available to data providers can be found here:

<http://www.theia.org.uk/downloads/ilrdocuments/ilrdetail.htm>

5.32 Within the Data Service and BIS, processed data and resulting outputs are quality assured and signed off, in accordance with National Statistics guidelines. This will shortly be made available on the Data Service website.

5.33 Data quality and standards are in place for quality of the data returned by providers:

[http://www.theia.org.uk/NR/rdonlyres/A69791E5-B14F-40DA-B059-4036AFB64CE9/0/natilrspecification2009\\_10Appendix\\_Psp26Jan2009v1.pdf](http://www.theia.org.uk/NR/rdonlyres/A69791E5-B14F-40DA-B059-4036AFB64CE9/0/natilrspecification2009_10Appendix_Psp26Jan2009v1.pdf)

5.34 To identify the proportion of achievements which are first achievements at that level by the individual, we mainly use information from Prior Qualifications Surveys. The analysis behind this is discussed at: [http://www.thedataservice.org.uk/NR/rdonlyres/5EF148B3-BC49-4593-9E88-696B687D22F5/0/Firstness\\_Rates\\_For\\_FullLevel2\\_FullLevel3.pdf](http://www.thedataservice.org.uk/NR/rdonlyres/5EF148B3-BC49-4593-9E88-696B687D22F5/0/Firstness_Rates_For_FullLevel2_FullLevel3.pdf)

5.35 The Prior qualifications Surveys involve trained interviewers asking detailed questions which are similar to those used in the LFS education section. So these surveys will have similar problems as those faced by the LFS, but as they are smaller (with samples of a few hundred for some learning routes), there will be more sample error.

5.36 Data on the number of first full L2 achievements is shown below.

**Table 16: Full L2 and first full L2 achievements (000's)**

|        | First full L2 achievements |
|--------|----------------------------|
| 2002/3 | 35,700                     |
| 2003/4 | 47,100                     |
| 2004/5 | 58,300                     |
| 2005/6 | 129,600                    |
| 2006/7 | 139,000                    |
| 2007/8 | 192,600                    |

5.37 For years prior to 2002/03, we have assumed a constant level of achievements.

5.38 Note that first full L2 achievements include Level 3 jumpers.

### Sensitivity

5.39 We have considered the sensitivity of the number of first achievements below.

**Table 17: Publicly funded achievements**

|                            | Number of first publicly-funded L2 qualifications | SFR smoothed growth minus modelled growth 2001-2005 | SFR smoothed growth minus modelled growth 2005-2008 |
|----------------------------|---|---|---|
| Maximum realistic scenario | 5% more first achievements than current estimate  | + 0.00ppts  | - 0.68ppts  |
| Current estimate           | Based on ILR and prior quals survey               | + 0.10ppts  | - 0.59ppts  |
| Minimum realistic Scenario | 5% less first achievements than current estimate  | + 0.19ppts  | - 0.50ppts  |

5.40 The table above shows that a 10% change in the number of publicly funded first achievements has very little impact over 2001-2005 years as there were relatively few public achievements in that time period. A much larger effect is seen from 2005 to 2008 as that is the time period in which a large growth of achievements occurred.

## Private Vocational Up-skilling (i.e. achievements gained without LSC allocated funding)

5.41 Vocational qualifications are recorded on the NISVQ dataset. By subtracting the vocational qualifications recorded on the ILR from the NISVQ we should have a measure of private vocational upskilling. As the series only extends back to 2003/04, a constant level has been assumed in earlier years.

**Table 18: Privately funded achievements**

|               | 2001/02 | 2002/03 | 2003/04 | 2004/05 | 2005/06 | 2006/07 | 2007/08 |
|---------------|---------|---------|---------|---------|---------|---------|---------|
| Full L2       | 75,600  | 75,600  | 75,600  | 112,900 | 124,600 | 100,300 | 66,600  |
| First Full L2 | 43,000  | 43,000  | 43,000  | 61,600  | 67,700  | 59,100  | 44,800  |

### Limitations of data sources

5.42 NISVQ does not include data from all providers and is scaled up to compensate for this. So subtracting ILR vocational qualifications from it may lead to small errors. Ideally we should match ILR with NISVQ qualifications to confirm that all ILR qualifications are recorded on the NISVQ, and that the remainder are privately-funded.

5.43 To estimate the proportion of these achievements which are first achievements at that level for the individual concerned, we have analysed the LFS to look at the qualifications held by those currently studying for a full L2. As LFS respondents cannot reliably say whether their courses are publicly or privately-funded, we have made some adjustments for the existing estimates for number publicly-funded first achievements. This analysis assumes a constant rate throughout the life of the model (set at 40%); while a moving figure may produce better results, the evidence for producing these is not available.

5.44 The level 3 jumpers assumption were created in a similar way by looking at respondents studying for a level 3 qualification who stated that their highest level of prior qualification was below level 2.

### Sensitivity

5.45 The table below shows that varying the proportion of private vocational qualifications which are assumed to be first by 10% has an impact of well over 0.1 ppts in both periods..

**Table 19: Privately Funded achievement firstness**

|                            | Percentage of private vocational qualifications which are first | SFR smoothed growth minus modelled growth 2001-2006 | SFR smoothed growth minus modelled growth 2006-2008 |
|----------------------------|---|---|---|
| Maximum realistic scenario | 50%   | - 0.11ppts  | - 0.67ppts  |
| Current estimate           | 40%   | + 0.10ppts  | - 0.59ppts  |
| Minimum realistic Scenario | 30%   | + 0.30ppts  | - 0.51ppts  |

### Other Up-skilling (achievements gained that cannot be described by either of the above)

- 5.46 There are other forms of upskilling which we do not have detailed records of but are able to make some assumptions about.
- 5.47 These include the following forms of level 2 qualification:
- Level 4 jumpers. People gaining a level 4 qualification without already having achieved a full L2 or higher qualification. The LFS shows around 140,000 people each year are currently studying for an L4 qualification without a prior L2 qualification. These would include Open University students and people with foreign qualifications. Longitudinal LFS work suggests around 12,000 achieve each year, although some may be foreign students who leave the country shortly afterwards.
  - LSC fund part L3 qualifications e.g. single A levels which count as full L2 qualifications. ILR data shows there are around 10,000 of these per year. The prior qualifications survey estimates that around 25% of these are first qualifications.
  - LSC also fund part L2 qualifications which may take people to a full L2 if it is their 5th GCSE, for example.
  - The last two types of academic qualification may be funded privately as well as by the LSC, and we have no records of them.
- 5.48 Although the numbers set out above would seem to suggest a number of sources of other first qualifications being achieved each year, there are currently many unknowns. For example, a good number of the foreign students will not remain in the country and some of their qualifications will have been under-reported on the LFS. Consequently we have made a midpoint estimate of about 20,000 of these types of level 2 qualification each year. There are insufficient data in the LFS to generate a useful time series for this estimate, and we have not been able to identify other administrative data (e.g. from awarding bodies) within the timescale of this work.

### Sensitivity

- 5.49 The table below shows that variations in this series do have a noticeable impact. Further work is therefore required to increase the reliability of the assumptions being made.

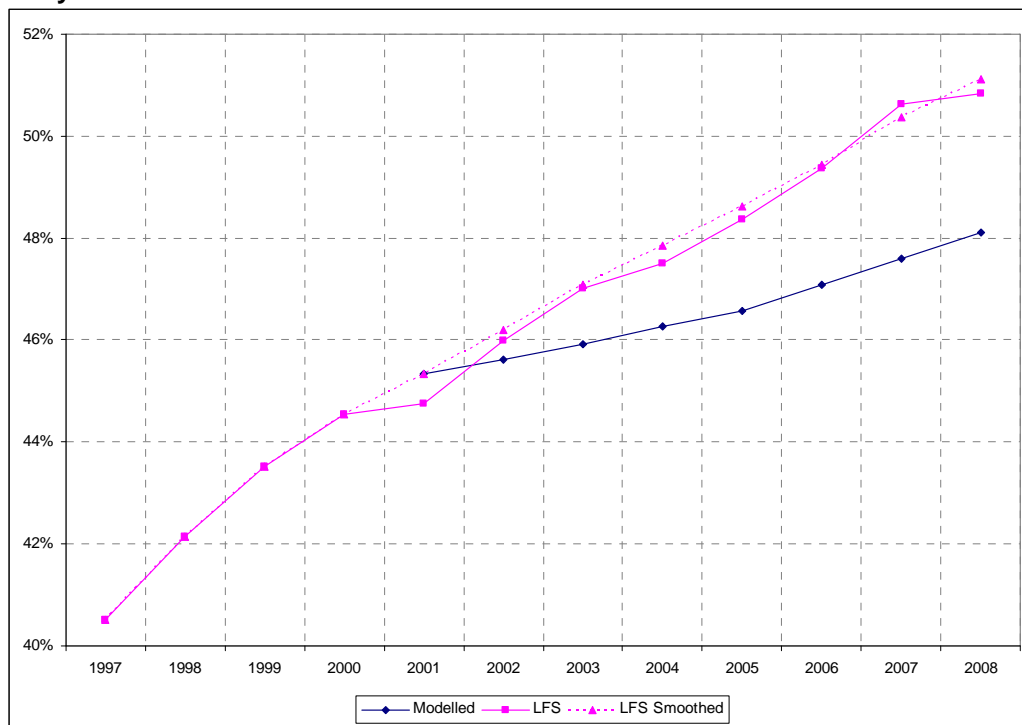
**Table 20: Other upskilling qualifications**

|                            | <b>Numbers of Other upskilling</b> | <b>SFR smoothed growth minus modelled growth 2001-2006</b> | <b>SFR smoothed growth minus modelled growth 2006-2008</b> |
|----------------------------|------------------------------------|--|--|
| Maximum realistic scenario | 40,000                             | <i>-0.24ppts</i>   | <i>- 0.72ppts</i>  |
| Current estimate           | 20,000                             | + 0.10ppts   | <i>- 0.59ppts</i>  |
| Minimum realistic Scenario | 12,500                             | + 0.22ppts   | <i>- 0.54ppts</i>  |

## **6 Modelled historical growth data for the level 3 and above attainment series**

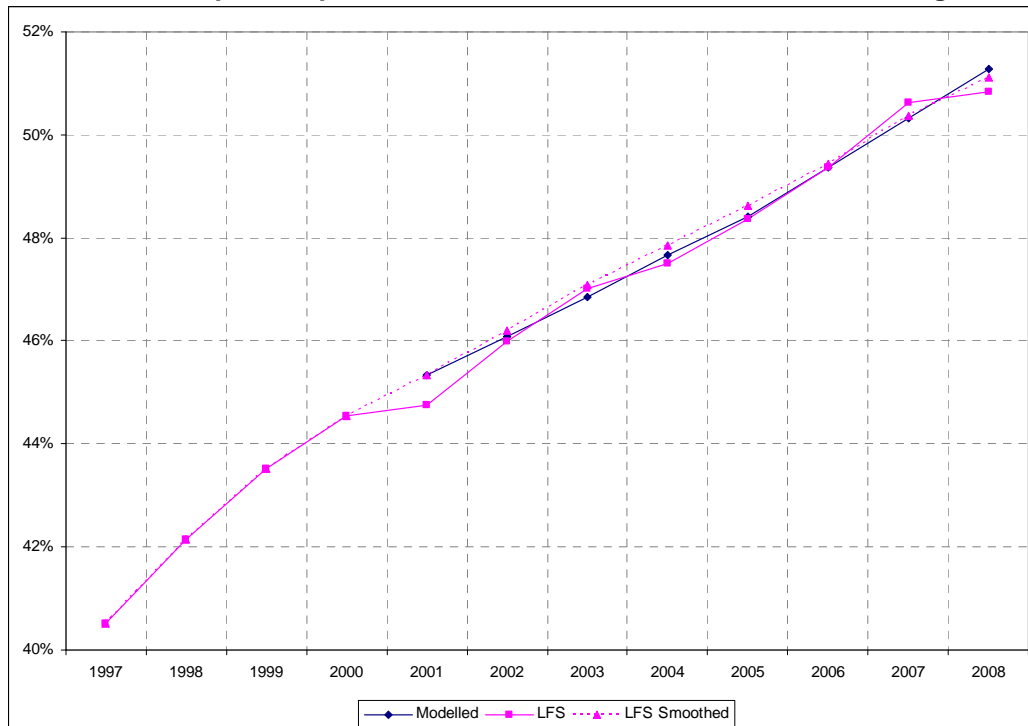
- 6.1 The modelled historical growth for the Level 3+ series is constructed in a broadly similar way to that for L2+. This section of the paper therefore presents similar summary information as for L2+. Differences in results are entirely due to the difference in the numerical parameters, such as numbers of achievements, that drive the modelled series.
- 6.2 In the case of L2+ attainment, a small proportion of the growth in the SFR series in the first part of the decade could not be explained by achievements measured by the main administrative data sets (once demographic effects have been accounted for as far as possible). The remainder could however, be explained as arising from a number of areas, for which we had evidence but not actual time series. We were therefore able to assume an 'other upskilling' series based on this evidence.
- 6.3 The situation is somewhat different for level 3, where the key challenge in increasing the robustness of this model has been to identify sources of administrative data that report on all of the routes through which people can reach L3.
- 6.4 Figure 2 shows the observed and modelled growth in level 3 attainment based on demographic change and achievements from both the ILR and private vocational qualifications. Note that SFR growth is low between 2001 & 2002 and 2007 & 2008. The SFR series has been smoothed (using Holt-Winters exponential smoothing) to provide a more consistent comparator for the (inherently smoother) modelled series.

**Figure 2: Level 3 comparison plot using achievements identified in time series for ILR and privately funded vocational achievements**



6.5 It is possible to obtain adequate agreement the SFR and admin data-based series by assuming that there are an additional 135,000 level 3 achievements per annum; this is illustrated in Figure 3. We discuss the evidence for this assumption below in paragraph 6.12.

**Figure 3: Level 3 comparison plot with additional achievements to match SFR growth.**



6.6 While BIS has been able to identify some of the upskilling activity that contributes to this number (see paragraph 6.12 below), we do not have the consistent and comprehensive data necessary to provide satisfactory input into the model. While comparison with the smoothed SFR data from 2001 to 2006 suggest this number has been stable over time, we do not know whether it is reasonable to assume this will remain stable in light of changes to the scale and nature of public sector upskilling that BIS have been delivering. This necessarily limits the value of the modelled growth data for the L3+ series.

6.7 The comparison of this modelled growth with the SFR series does suggest a similar pattern to that observed for the L2+ series. That is, there is a reasonably good fit until 2006, after which point the modelled growth series would suggest a higher growth rate than is demonstrated in the SFR. However, what is not known is the extent to which this might be due to the weak assumption we have made regarding the level of other upskilling.

6.8 Table 21 shows modelled and SFR growth at level 3 and above.

**Table 21: Modelled & SFR results assuming extra achievements**

| Year | SFR percentage at Level 3+ | SFR Growth | Modelled Growth |
|------|----------------------------|------------|-----------------|
| 2001 | 44.8%                      | 0.0%       | 0.0%            |
| 2002 | 46.0%                      | 1.2%       | 0.7%            |
| 2003 | 47.0%                      | 1.0%       | 0.8%            |
| 2004 | 47.5%                      | 0.5%       | 0.8%            |
| 2005 | 48.4%                      | 0.9%       | 0.7%            |
| 2006 | 49.4%                      | 1.0%       | 1.0%            |
| 2007 | 50.6%                      | 1.3%       | 0.9%            |
| 2008 | 50.8%                      | 0.2%       | 0.9%            |

6.9 The components of growth for level 3 are set out below. The additional 135,000 annual achievements are represented in the other upskilling column.

**Table 22: Modelled components of growth in percentage point terms of proportion of population at level 3 or above**

| Year | Modelled growth based on actual SFR | Growth over previous year from each component |           |           |                    |                     |                   |
|------|-------------------------------------|---|-----------|-----------|--------------------|---------------------|-------------------|
|      |                                     | Population Replacement                        | Mortality | Migration | Public Up-skilling | Private Up-skilling | Other Up-skilling |
| 2001 |                                     |   |           |           |                    |                     |                   |
| 2002 | 0.77                                | 0.31  | -0.09     | 0.06      | 0.07               | 0.04                | 0.37              |
| 2003 | 0.78                                | 0.30  | -0.10     | 0.07      | 0.07               | 0.05                | 0.39              |
| 2004 | 0.82                                | 0.31  | -0.08     | 0.14      | 0.08               | 0.04                | 0.33              |
| 2005 | 0.75                                | 0.29  | -0.07     | 0.09      | 0.09               | 0.05                | 0.30              |
| 2006 | 0.97                                | 0.43  | -0.09     | 0.08      | 0.15               | 0.06                | 0.35              |
| 2007 | 0.95                                | 0.31  | -0.10     | 0.14      | 0.17               | 0.07                | 0.36              |
| 2008 | 0.95                                | 0.25  | -0.09     | 0.18      | 0.20               | 0.06                | 0.35              |

## Public & Private Up-skilling

6.10 Table 25 shows the number of first full achievements used in the model.

**Table 23: Level 3 publicly-funded achievements**

|        | First full L3 achievements |
|--------|----------------------------|
| 2002/3 | 25,300                     |
| 2003/4 | 33,000                     |
| 2004/5 | 40,800                     |
| 2005/6 | 57,400                     |
| 2006/7 | 61,200                     |
| 2007/8 | 76,700                     |

6.11 Table 26 shows the number of privately funded achievements used. These have been derived using the same data sources and techniques as for level 2.

**Table 24: Level 3 privately-funded achievements**

|               | 2001/02 | 2002/03 | 2003/04 | 2004/05 | 2005/06 | 2006/07 | 2007/08 |
|---------------|---------|---------|---------|---------|---------|---------|---------|
| Full L3       | 40,400  | 40,400  | 40,400  | 52,400  | 57,000  | 61,000  | 62,700  |
| First Full L3 | 16,200  | 16,200  | 16,200  | 21,000  | 22,800  | 24,400  | 25,100  |

## Other Up-skilling

6.12 As discussed in paragraph 6.5 above, producing a good fit in the model requires an assumption that there are 135,000 other individuals upskilling to level 3 each year. As for level 2, there are sources of level 3 achievements which would not be identifiable from the administrative time series we are already using. Those which we have been able to identify to date are:

- Level 4 jumpers. People gaining a level 4 qualification without already having achieved a full L3 or higher qualification. The LFS shows around 300,000 people each year are currently studying for an L4 qualification without a prior L3 qualification. These would include Open University students and people with foreign qualifications. These are discussed in more detail below.
- LSC also fund part L3 qualifications which may take people to a full L3 if it is their 2<sup>nd</sup> GCSE, for example.

6.13 Work on identifying level 4 jumpers has used the LFS to conduct a number of analyses. Comparing the qualifications people are currently studying for against what they currently hold provides the following figures.

**Table 25: Qualifications currently studying for on LFS.**

|                            | <b>Q4 2004</b> | <b>Q4 2006</b> | <b>Q4 2008</b> |
|----------------------------|----------------|----------------|----------------|
| <b>Studying for L3</b>     | 401,000        | 365,000        | 406,000        |
| <b>L3 Jumpers</b>          | 82,000         | 69,000         | 75,000         |
| <b>L3 Firstness</b>        | 59.8%          | 72.3%          | 61.4%          |
| <b>Studying for L4+</b>    | 1,816,000      | 1,827,000      | 1,965,000      |
| <b>Jumpers from &lt;L3</b> | 327,000        | 316,000        | 307,000        |
| <b>Jumpers from &lt;L2</b> | 140,000        | 128,000        | 132,000        |

Base: 19-59/64, England - Source: Labour Force Survey

- 6.14 However, because courses take a number of years to complete, these figures need converting to annual achievements. We can do this by multiply the learners (stock) figure by:
- an in-year completion rate to give the number of completers each year. This is approximately the reciprocal<sup>3</sup> of the average course length.
  - a success rate, to give the number of achievers per annum.
- 6.15 Further work is needed to identify these figures but we would expect average course lengths for level 4 to be 2-4 years, with most first degree courses being 3 years. Applying a conservative figure of 25% of learners completing per year would give 75,000pa. This analysis would suggest this group is a major potential contributor to the other upskilling we have assumed.
- 6.16 An alternative data source to the LFS is the Higher Education Statistics Agency (HESA) data, which has details of entrants into HE and qualifiers for each academic year, including their prior qualifications. This will not be comparable to the LFS data above, which gives estimates for the numbers studying for a level 4 or above at a given point in time, but it can give some context as to the reliability of the LFS estimate of the proportion of level 4 participants who are not qualified to level 3. Preliminary investigations have taken place in this area and will be continued alongside the further work to be undertaken using HESA data in the main research project.

---

<sup>3</sup> By Liittle's law. This applies to a stock close to long-term average equilibrium, and is independent of the distribution of residence times of individuals.

## 7 Modelled historical growth data for the level 4 and above attainment series

- 7.1 BIS models the level of the population holding L4+ qualifications to project the likely impact planned delivery of L4+ qualifications will have on the proportion of the working age population qualified to this level.
- 7.2 Important factors in the L4+ modelling are the number in the population who are already qualified, the number newly qualified, arrivals of immigrants to England with qualifications, and changes to the underlying working population. These inputs are used to make a best estimate of the proportion qualified to that level.
- 7.3 An additional complexity to the L4+ modelling, compared with the L2+ and L3+ models, is the modelling of the impact of students domiciled overseas who remain in England after the completion of their studies.
- 7.4 The L4+ modelling work is at a relatively experimental stage – development of the model started in autumn 2008, and the model was first used in spring 2009. It has not previously been used to explore the historical data in the same way as has been done for L2+ and L3+. Nonetheless, some work has been conducted to make comparisons along similar lines.
- 7.5 The development of the L4+ model involved a process of “rolling back” to consider how well the model projected from a 2004 base against out-turn figures from later years. This process showed the model to have a good fit to year-on-year growth in proportion of the population qualified to Level 4 or above (as measured by the LFS) - see table 26. The model was not rolled back further (as was done for the L2+ and L3+ models) as this was not deemed necessary for the development of the model.
- 7.6 The estimated growth figure from the existing L4+ model suggests growth of around 1.3 percentage points between 2006 and 2008. This is within the sampling error range of the unadjusted SFR growth estimate, as are the growth estimates for each of 2007 and 2008.

**Table 26: SFR and modelled growth in proportion of the working age population qualified to Level 4 or above**

| Year | SFR     |        | Modelled growth |
|------|---------|--------|-----------------|
|      | Outturn | Growth |                 |
| 2004 | 27.9%   |        |                 |
| 2005 | 28.5%   | +0.6%  | +0.7%           |
| 2006 | 30.0%   | +1.5%  | +0.7%           |
| 2007 | 30.9%   | +0.9%  | +0.7%           |
| 2008 | 31.2%   | +0.3%  | +0.7%           |

7.7 Growth in the number in the working age population qualified to L4+ (the numerator) is estimated in the model based on the net change arising from:

- Numbers of new mainstream qualifiers each future year available from the HESA administrative data.
- Qualified people moving forward or out of the model
- Overseas students gaining level 4 qualifications in England and remaining in the country (currently about 15,000 each year), based on analysis of latest outturn HESA data, and an assumption that around 30% of these qualified students remain in England after qualifying.
- Students gaining level 4 qualifications in other (non-England) UK higher education institutions who move / return to England (currently about 10,000 each year), based on analysis of latest outturn HESA data and the latest DLHE survey.
- Immigrants that arrive in England with a level 4 qualification (currently about 90,000 each year), based on latest Quarter 4 LFS data.

7.8 The L4+ model uses ONS/GAD population estimates as the denominator, while the Level 2+ and Level 3+ historical modelling uses LFS based population estimates (which have a slightly more limited coverage than the whole population as measured by ONS). The impact of using such data in L4+ projections has not been tested.

7.9 Similarly to the L3+ model, the L4+ model also includes an assumed level of net additional up-skilling arising from other sources, including outward migration and death which are not modelled separately in the current L4+ model. This is currently assumed to be around 30,000 additional achievements per year which was the approximate shortfall of our modelled series below the LFS outturn series.

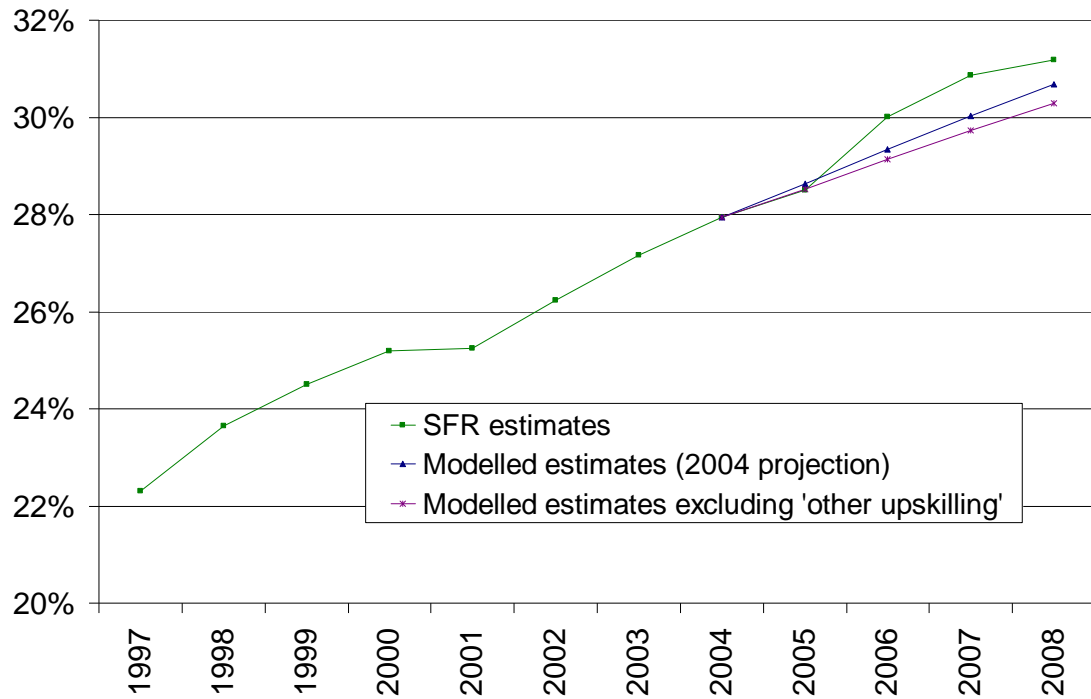
**Table 27: Modelled components of growth in percentage point terms of proportion of population at Level 4 or above**

| Year | Modelled growth | Fluctuation in working age population | Population replacement | Upskilling in English HEIs & FECs | Other modelled factors | Other upskilling |
|------|-----------------|---------------------------------------|------------------------|-----------------------------------|------------------------|------------------|
| 2005 | 0.7%            | -0.1%                                 | -0.3%                  | 0.9%                              | 0.2%                   | 0.1%             |
| 2006 | 0.7%            | -0.2%                                 | -0.3%                  | 0.9%                              | 0.2%                   | 0.1%             |
| 2007 | 0.7%            | -0.1%                                 | -0.4%                  | 0.9%                              | 0.2%                   | 0.1%             |
| 2008 | 0.7%            | -0.1%                                 | -0.4%                  | 0.9%                              | 0.2%                   | 0.1%             |

7.10 The 'Other modelled factors' component covers the effect of qualified immigrants, overseas domiciled students & students at other UK institutions who remain in / return to the UK after completing their studies

7.11 The 'Other upskilling' column includes net additional up-skilling arising from other sources, including outward migration and death. This is an estimated value, as discussed in paragraph 7.9.

**Figure 4: Comparison of SFR series with 2004-based projection including and excluding 'other up-skilling' component**



7.12 The overall conclusion from comparison of the L4+ model with the SFR data is that there is no suggestion that growth in recent years has been lower in the SFR than one would expect. Further work to consider the robustness of the L4+ model and the underpinning assumptions will be conducted as new data become available.