Department for Business, Innovation & Skills

Guide to the simplified student loan repayment model (June 2015)

This model is being released as an updated version of that published in June 2014. If you have any feedback on the model please send it to helen.woodward@bis.gsi.gov.uk

Contents

How to use the simplified model	**page 2**
Technical guide	**page 3**
StEP model assumptions	**page 11**

How to use the simplified model

This new version of the simplified model is identical to operate as previous versions. On opening the worksheet you may be asked if you wish to enable macros. You must enable macros in this model or the simulation won't run. On the worksheet called 'Global' is a button called 'Run simulation'. Clicking on this runs the simulation, and generates a RAB charge in cell B33, which represents the proportion of loan value which is not expected to be repaid.

Some of the most useful parameters to change and experiment with are the macro economic forecasts on the Global worksheet. Row 15 shows RPI growth, which influences the interest rate charged to borrowers and the discount rate used for the NPV calculation. Row 16 shows RPIX growth which gives the assumed rate of increase in loan amounts. Rows 17 and 18 give earnings growth for two points in the year, March for repayment threshold increases, and September for earnings increases.

The other key parameters to experiment with are repayment policy parameters; the yellow cells in rows 4 to 11. These change the threshold, repayment rate above this threshold, maximum real interest rate, the upper threshold for maximum interest, the maximum repayment period, debt amounts for each year of a three year course, gender split, and the discount which has been applied.

A sample of 10,000 male and 10,000 female earnings paths have been taken from the full model and copied into the 'paths' worksheet. The main change from the previous simplified model is an improvement of the earnings distribution and linking non-employment to earnings, along with updated macro economic forecasts.

The simplified model only calculates a RAB charge for students who have studied a 3 year first degree for an appropriate distribution of entry ages. This iteration of the simplified model introduces some of the more detailed modelling assumptions from the full model: i.e. direct repayments, investment income, death, disability, migration, repayment frictions. The technical guide on the following pages gives more detail.

Technical guide

ICR Student Loan repayment modelling: StEP

The Stochastic Earnings Path (StEP) model is the financial model used by the Department for Business, Innovation & Skills (BIS) to estimate the financial cost of income-contingent student loans to Government – the current iteration is known as StEP3. The main improvements compared to the previous iteration of the StEP model (known as StEP2) concern the linking of earnings to employment and the improvement of the earnings distribution. BIS is grateful for the advice of external analysts in developing and quality assuring the model.

The StEP model uses wage equations to estimate the next year's wage, and when these are used repeatedly, we can generate an earnings path. The wage equations use multiple years of historical earnings when forecasting future earnings; and they also make extensive use of the Student Loans Company (SLC) administrative data. The StEP2 model used a single wage equation for each gender – the effect of this was that for borrowers with "average" prior earnings and characteristics the forecasts were reliable, but for low or high earners the equation ensured that they would stay as low or high earners respectively. While this state of affairs often occurs in reality, it is not always the case. To create more flexibility, multiple wage equations are used in the StEP3 model depending on whether the borrower has been a low, medium, or high earner historically. This allows some opportunity for a person to experience a change of fortune over their lifetime. The resulting lifetime earnings distribution is a little narrower than in the previous iteration of this model, but more realistic.

The StEP2 model also generated a non-employment (defined as unemployment plus economic inactivity) path for each individual by using a model based on history of non-employment, age and course level, but crucially not earnings. The probability of being non-employed was effectively independent of earnings, and this was not very realistic. In the StEP3 model non-employment is now linked to historical earnings (also earnings is partially determined by non-employment history) which takes into account that low earning individuals are more likely to become non-employed than high earners.

Recent graduates have different characteristics from the graduate population found in survey data. They tend to have lower wages and lower employment rates than the average graduate. The StEP model gives weight to the SLC data on recent graduates; and StEP3 uses the course subject and institution information to improve the forecasting of earnings in early repayment years.

The main data sources used in the model are:

- SLC administrative data – to initialise the model with the correct income and employment distributions, also used for earnings in early repayment years. Used for modelling of payments direct to the SLC.

- British Household Panel Survey (BHPS) data – used in the earnings and employment models
- Labour Force Survey (LFS) data – to convert income percentiles to cash amounts, regarded as more reliable than cash values from BHPS due to large sample sizes
- Destinations of Leavers from Higher Education (DLHE) survey – used in the graduate age adjustment
- ONS International Passenger Survey (IPS) – data on migration
- ONS life tables – data on deaths
- UCAS and HESA data – forecasts of student numbers and loan amounts.
- Office for Budget Responsibility (OBR) macroeconomic forecasts – forecasts of earnings growth, Bank of England base rates, and RPI.

The flow diagram on page 6 explains, at a high level, the processes which the StEP model goes through to produce the forecasts, along with how each data source feeds into the full model. Fuller explanations follow from page 7 onwards.

Purpose of the simplified loan repayment model

This is an update of the simplified loan repayment model which was published in June 2014, and incorporates the substantial methodological aspects of the full StEP model. It is a simplified and standalone version of the full loan repayment model used by BIS but the key principles behind it are consistent with the version of the full loan repayment model adopted in spring 2015.

Like the full repayment model, the simplified model operates at the level of individual borrowers (it is a micro-simulation model) and considers a sample of these to arrive at an indication of the likely resource cost of the loans.

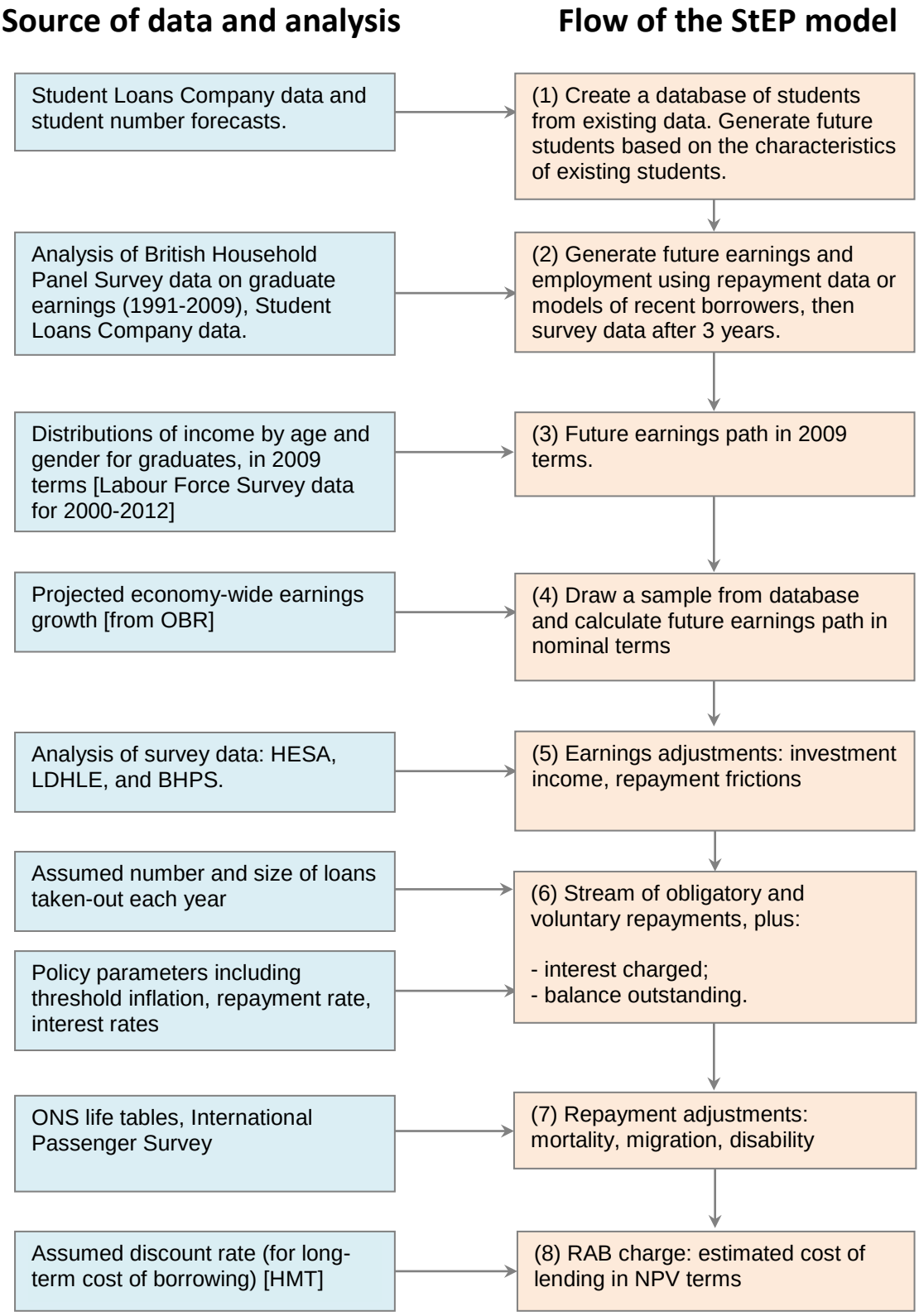
The main simplifications are the way in which this model:

- only considers entrants in 2014 not other years;
- only considers graduates from three-year degree level courses;
- and only considers a uniform amount of borrowing across every borrower.

The results from the full and simplified models are therefore not directly comparable and represent slightly different things. However, the simplified model can be expected to give a reasonable illustration of the general sensitivity of future repayments and the expected cost of loans to:

- key assumptions such as future graduate earnings and the discount rate; and
- policy parameters such as the loan repayment threshold, repayment rate, interest rate and maximum repayment period.

Processes and sources underlying the StEP model



Steps 1, 2 & 3: Forecasting earnings paths in 2009 terms

Note: The simplified model does not include the required regression models to generate earnings paths (including non-employment) for the sampled borrowers. Instead the model contains 20,000 earnings path simulations generated using the regression models. The methodology behind the earnings forecasts is explained below.

Earnings in the first three years after Statutory Repayment Due Date (SRDD)

The earnings in the first three years after SRDD are derived from earnings variables in the SLC administrative data set. Depending on the SRDD, there are a number of ways in which the SLC data is used.

The 'hot start' condition (SRDD 2010 or earlier)

For cohorts who are already more than three years into their repayment term, the actual earnings are extracted from the SLC data to be used as lagged dependent variables in the regressions for future earnings forecasts.

The 'cold start' condition (SRDD 2013 or later; used for the RAB charge and in the simplified model)

For cohorts who are yet to reach their SRDD, and therefore have no earnings history, specially derived models estimate their earnings (which can be zero in the case of non-employment) based on regression analysis of the SLC administrative data set. These regression models are based on gender, age, course type (i.e. whether the borrower completed a first degree or a sub degree), course subject, institution, and the previous earnings and non-employment status generated by the regression model.

Note: The earnings paths contained within the simplified model use the 'cold start' methodology as they represent future borrowers. The latest earnings available from the SLC data set are for the financial year 2012/13.

The 'warm start' condition (SRDD 2011 and 2012)

There will also be cases where we have some actual data on past earnings, but not enough for a hot start. In these cases we use a combination of the available repayments data, and estimates of earnings created in the same way as the cold start.

Benefits of using the SLC administrative data

Using the British Household Panel Survey (BHPS) data in early years leads to significant over-estimation of repayments. For example the survey may capture the earnings of an individual aged 30, but may not give an accurate indication of how long the individual has held their degree. It is likely that the majority of 30 year olds in the survey graduated several years earlier so both their earnings and earnings growth are unlikely to represent that of a new graduate of the same age. The loan borrower population includes some borrowers who drop out of their course before graduating and will exclude a proportion of graduates who could afford not to take out loans during their study. All of these factors indicate that we should expect the borrower population to earn less than the graduate population on average.

We cannot use the SLC data indefinitely. After some time the highest earners begin to fully repay their loans, and start to drop out of the longitudinal dataset. Regressions based on a longer run of SLC data would become biased and so after three years we assume that the trend in earnings can be best represented by the BHPS data.

Earnings for the rest of the repayment period

After the first three years, the historical earnings are fed into regression models based on the BHPS data. The historical earnings are converted into BHPS-equivalent values and the regression model calculates an earnings path using these historical values as lagged dependent variables, non-employment history, age (represented as age bands to reflect different rates of earnings growth depending on the point of an individual's career), course type and gender. The age bands show that the rate of earnings growth very much depend on the borrower's stage in life. The selection of the regression model is based on the borrower's earnings in the previous year.

Once the earnings path is calculated, the BHPS earnings are converted into LFS cash values using percentiles based on the borrower's age and gender. Due to differences in survey design, the LFS is a more reliable source of actual earnings, while the BHPS gives a better indication of individual earnings trends over many years.

Non-employment forecasting

The non-employment modelling follows a similar structure to the earnings modelling: the SLC administrative data is used for the first three years and the BHPS is used for the rest of the repayment term. Logistic regression models are used to derive the probability of non-employment, which depends on non-employment history, earnings history, and age and course level (and course subject and institution for the SLC-based models). The actual non-employment status is used in the warm and hot start conditions.

The graduate age adjustment

As mentioned previously, a recent graduate aged 30 for example is unlikely to have the same earnings or earnings growth as a borrower aged 30 who graduated aged 22. The use of the SLC data deals with this issue in early repayment years, but this issue continues to have a significant impact on the borrower's earnings for a number of years after graduation.

To adjust for this the model calculates a second earnings path for the borrower with their age at SRDD set at 22 (or 21 if the qualification is a sub degree), then effectively takes a weighted average of the two earnings paths to come up with an overall earnings path for the borrower. The weights start with a high dependency on the "typical graduation age" earnings path and shift towards the "actual age" earnings path over time.

Note: The graduate age adjustment is not built into the simplified model. The earnings paths provided have already been adjusted using the age adjustment methodology.

Step 4: Forecasting earnings paths in nominal terms

The individual earnings paths are adjusted into nominal prices using outturn macro-economic data (sourced from ONS Average Weekly Earnings series) and Office for Budget Responsibility (OBR) forecasts.

Step 5: Applying investment income to earnings and frictions

Note: The modelling of investment income is not included in the simplified model. Instead the model contains 20,000 investment income path simulations generated using the regression models. The methodology behind the forecasts is explained below.

A borrower is expected to include investment income above £2,000 in their annual earnings. The full model calculates a probability that a borrower will earn investment income based on their age, labour income, gender and whether they declared investment income in the previous year (based on analysis of BHPS data). Then investment income is added to their labour income using a regression model based on the same set of characteristics. If the model determines that the borrower has earned an income from their investments but it amounts to less than £2,000, then this is treated as zero investment income.

Repayment frictions

"Frictions" are defined as factors which prevent the correct amount of repayment being collected by an individual. These relate to technical issues such as National

Insurance numbers not matching their student loan account to their HMRC record, or the borrower not having a National Insurance number. The probability of any of these frictions occurring is based on analysis of SLC data, and if the individual is subject to any repayment friction then their repayments for that year are set to zero.

Step 6a: Calculating direct repayments

Note: The modelling of direct repayments is not included in the simplified model. Instead the model contains 20,000 direct repayments simulations generated using the regression models. The methodology behind the forecasts is explained below.

In addition to obligatory repayments collected through the UK tax system, repayments can also be paid directly to the SLC. These fall into three main categories:

1. Voluntary repayments (prepayments) – These are payments made by the individual in addition to obligatory repayments.
2. Payments from overseas – Repayments from borrowers situated overseas cannot be collected through the tax system. Overseas make obligatory repayments direct to the SLC based on their income and the earnings threshold for their country of residence.
3. Direct debits – In the last couple of years of payment the SLC offer the borrower the opportunity to repay the rest of their loan through a direct debit to prevent overpayment.

The probability of making a direct repayment is dependent on the magnitude of the debt outstanding and the number of years into the repayment period. The majority of direct repayments come from borrowers with low amounts of debt in the first few years of repayment. If a borrower is due to make a direct repayment in the model, a percentage of the debt outstanding is paid as a direct repayment.

Steps 6b & 7: Calculating repayments from earnings and deducting repayments from debt outstanding

Obligatory repayments

Once annual earnings (including any investment income) are calculated and non-employment is taken into account, the obligatory repayments are calculated according to the deterministic repayment rules for that year.

Mortality and disability assumptions

ICR loans can be written off prior to the end of the repayment term if the borrower dies or if they become disabled to an extent that they will be permanently unable to work. The probability of death in a given year is based on age and gender, and is

derived from the ONS life tables. Disability probabilities are also based on age and gender.

Migration assumptions

Borrowers who emigrate from the UK cannot make repayments through the tax system; and are required to make arrangements to make repayments direct to the SLC while overseas. The probability of emigrating comes from the International Passenger Survey (IPS), and is split between English-domiciled and EU-domiciled borrowers. A length of time away from the UK is then applied to the borrower.

Calculation of debt outstanding

Interest for the first half of the year is added to the debt outstanding at the start of the year; all repayments (both obligatory and direct) for that borrower in the year are then deducted; then the interest for the rest of the year is added. In reality repayments will be made on a monthly or weekly basis but it is a simplification to assume repayments occur in the middle of the year. If the repayment amount is greater than the loan balance outstanding (once the first half-year of interest is added), then the repayment is limited to the amount of loan balance and the loan is considered to be fully repaid.

Step 8: Calculating the RAB charge

Repayments are discounted back to a Net Present Value (NPV) using a discount rate supplied by the Treasury. The RAB charge is then deduced from the valuation and the initial loan balance.

StEP model assumptions

Macro-economic inputs
All macro-economic assumptions are taken from the Office for Budgetary Responsibility (OBR). They generally update their short term forecasts twice a year with publications alongside the Budget and Autumn Statement. They also publish a Fiscal Sustainability Report in July which focuses on the long term.

http://budgetresponsibility.org.uk/category/publications/

For outturn earnings data, we use the Average Weekly Earnings index that is published by the Office of National Statistics.

http://www.ons.gov.uk/ons/rel/lms/labour-market-statistics/may-2015/dataset--earnings.html

The discount rate used within our modelling to value future repayments is RPI+2.2% which is the rate we have been instructed to use by Treasury.

Student data
Student number forecasts are taken from an internal BIS model.

Assumptions for the age/gender mix of future students are taken from historical SLC data as are maintenance and fee loan distributions and loan take up rates.

Course lengths for new students are based on HESA data.

The model takes published ONS data to forecast loan write-offs due to death or disability.

Earnings
The modelling of earnings/non-employment is based on a combination of:

- Administrative data from the Student Loans Company
- Labour Force Survey data covering the period 2000-2013
- British Household Panel Survey data covering the period 1991-2009

Payments
The loan repayment threshold is assumed to increase annually in line with earnings for post-2012 system students. The loan repayment threshold is assumed to increase annually in line with RPI for pre-2012 system students.

Direct repayments forecasts are based on analysis of historical administrative data from the SLC.