Department for Business, Innovation & Skills

Guide to the simplified student loan repayment model (June 2014)

This model is being released as an updated version of that published in December 2013. If you have any feedback on the model please send it to helen.woodward@bis.gsi.gov.uk

Contents

The RAB charge

The Resource Accounting and Budgeting (RAB) charge is the estimated cost to Government of borrowing to support the student finance system. It is based on future loan write-offs and interest subsidies in net present value terms. For convenience, we express these costs as a proportion of the initial loan outlay.

The RAB charge is calculated by taking repayment forecasts for income contingent repayment (ICR) loans, and discounting them back to the period that the loan is issued using the discount rate provided by Treasury (RPI+2.2%). This gives us a net present value (NPV) of the future repayments and the cost is the difference between the loan issued and the NPV of the repayments.

The purpose of the BIS student loan repayment model, which is called StEP, is to assist in valuing the existing ICR student loan book, and to provide forecasts for budgeting purposes.

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 B40, 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. Rows 19-21 show RPI and earnings growth. Earnings growth is shown for two points in the year, March for threshold increases in March, and September for earnings increases.

The other key parameters to experiment with are repayment policy parameters; the yellow cells in rows 4 to 14. 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 size StEP model and copied into the 'paths' worksheet. This replaces the worksheets 'paths' and 'ID' in the previous version of the simplified model. The main change from the previous HERO model to StEP is in the generation of these earnings paths, and this is the main difference in this new simplified model, along with updated macro economic forecasts.

The simplified model only calculates a RAB charge for 22year olds who have studied a 3 year first degree. Some of the more detailed modelling assumptions have been left out. 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 new financial model used by the Department for Business, Innovation & Skills (BIS) to estimate the financial cost of income-contingent student loans to Government. The main improvements compared to the previous model concern the forecasting of earnings – and in particular the trends of individuals' earnings over the repayment period. The previous model (known as the HERO model) had a reasonable repayments methodology, so many of the repayments calculations are unchanged. Although the forecast of earnings is much improved, the model will remain under scrutiny and subject to further changes in a work programme of continuous improvement. BIS is grateful for the advice of external analysts in developing and quality assuring the model.

The HERO model used a transition probability matrix to forecast earnings for an individual in a given year. The probabilities were based on the previous year's earnings for that individual, along with their age, gender and course type. By only relying on one year of previous earnings, the model had excess volatility in the earnings paths, and high earners could suddenly be switched to a low earning percentile and vice versa. Over a lifetime of earnings this meant that the earnings paths tended towards an average value; that is, the earnings distribution was too narrow. In reality we know that high earners tend to remain as high earners and the same for low earners, and the true distribution of earnings is wider than the HERO model suggested.

The overall RAB estimate is sensitive to the proportion of borrowers who repay very little (or nothing) as well as the proportion who repay early: the low and high earners over a sustained period. The previously narrow earnings distribution tended to underestimate both groups and hence underestimated the RAB charge. The new model is better at separating out the high and low earners over their lifetimes, and this improved differentiation gives a more accurate estimate of the RAB charge.

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 more years of historical earnings when forecasting future earnings, and they also make more extensive use of the Student Loans Company (SLC) administrative data where available. StEP also uses more up to date survey data than HERO.

These changes work together - the wage equation moves away from a random earnings path and gives a proper trend for individuals, at the same time accommodating several years of earnings history which can be filled with earnings data. The SLC repayments are then used in the equations to find the right starting point and trajectory for the graduate's earnings.

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 more weight to the SLC data on recent graduates. This tends to bring the early years of repayments down, and due to the increased use of earnings history, this effect persists into the forecasts.

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
- 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 the page 3 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 4 onwards.

Purpose of the simplified loan repayment model

This is an update of the simplified loan repayment model which was published in January 2014, and incorporates the substantial methodological changes described in 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 2014.

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.

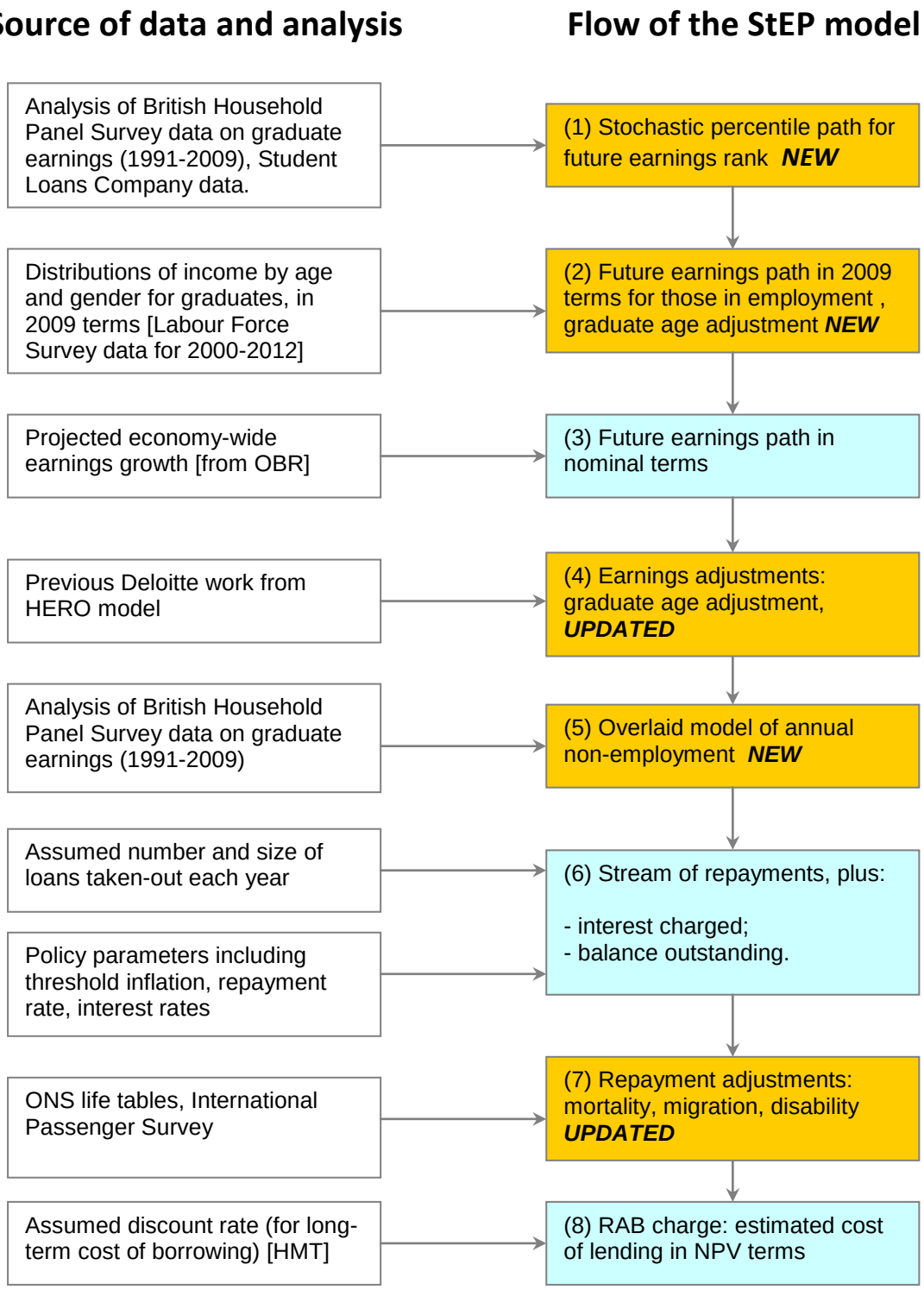
The main simplifications are the way in which this model:

- only considers entrants in 2013, not other years;
- only considers young graduates from three-year degree level courses;
- only considers a uniform amount of borrowing across every borrower;
- ignores the possibility of early repayments; and
- ignores the possibility of death or disability.

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: Forecasting earnings paths in 2009 terms

Note: The simplified model does not include the required regression models to generate earnings paths 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 four years after Statutory Repayment Due Date (SRDD)

The earnings in the first four 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 2007 or earlier)

For cohorts who are already more than four 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 2012 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 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), and the previous earnings 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 2011/12.

The 'warm start' condition (SRDD 2008 to 2011)

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

The previous model (HERO) used a percentile transition matrix based on analysis of the British Household Panel Survey (BHPS) data, even in early repayment years. Using the survey 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 SLC data allows the new model to get a more accurate starting point and trajectory for new graduates, who have been out of the labour market and may have different characteristics to the general graduate population. However, 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 4 years we assume that these graduates are sufficiently like the general graduate population that they can be modelled using BHPS data.

Earnings for the rest of the repayment period

After the first four years, the historical earnings are fed into a regression model 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, 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.

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.

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 sample is of 22 year olds already so the adjustment would have zero impact on the forecasts.

Step 3: 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 OBR forecasts.

Step 4: Applying investment income to earnings

Note: This step is not included in the simplified model. Investment income has a low prevalence among borrowers and the resulting increase in earnings is not large.

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 and gender (based on analysis of BHPS data). Then an investment income is added to their labour income from a normal distribution of investment incomes with mean ~£2,500 and standard deviation ~£2,000. 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.

Step 5: Applying full-year non-employment to earnings

Note: In the simplified model the full year non-employment has already been applied to the earnings paths in the "Paths" sheet, so the non-employment model is not explicitly included in the model.

For the purposes of this model we are interested in periods where individuals do not earn wages. To define this state we group unemployment and economic inactivity together as 'non-employment'. In the HERO model the non-employment was modelled as the bottom end of the earnings transition matrix, with low percentiles being assigned zero wages. This tended to mean that a strong connection was created between low earnings and non-employment. Whilst we know that a connection exists it was modelled too strongly in the previous model as those in the top end of the percentile distribution effectively had no chance of moving into non-employment.

In the StEP model a separate process of non-employment is modelled using logistic regressions with 3 lagged dependent variables to provide employment history. After earnings paths have been generated the probability of being non-employed is calculated and the appropriate zero earnings periods are overlaid onto the earnings path.

For cold/warm starts only: For the first four years after SRDD (when models based on SLC administrative data are used to calculate earnings), the probabilities of being non-employed for a full year are derived from the prevalence of full-year non-employment in the SLC data. These probabilities are dependent on any previous history of non-employment, gender and age. In the early years of repayment non-employment occurs at a much higher rate because of borrowers going on to further study.

The employment history is important as we know that the probability of being non-employed is dependent on employment history.

Steps 6 & 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.

Note: The probability of death or disability is very low particularly for younger graduates and this is excluded from the simplified model.

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.

Note: The probability of a current borrower being subject to any of these frictions is very low and frictions are excluded from the simplified model.

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.

Note: The simplified model assumes zero migration.

Prepayments (voluntary repayments)

In addition to the obligatory repayments, borrowers may choose to make a voluntary repayment (known as a prepayment). The probability of making a prepayment is dependent on the magnitude of the debt outstanding and the number of years into the repayment period. The majority of prepayments come from borrowers with low amounts of debt in the first few years of repayment. If a borrower is due to make a prepayment in the model, a percentage of the debt outstanding is paid as a prepayment.

Note: The simplified model does not include prepayments.

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 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/search/index.html?newquery=awe

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 is based on a combination of:

- Administrative data from the Student Loans Company
- Labour Force Survey data covering the period 2000-2012
- British Household Panel Survey data covering the period 1991-2009

Non-employment rates are also based on analysis of British Household Panel Survey data.

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.

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