Department for Education

Student loan forecasts, England: 2019 to 2020

Quality and methodology information, September 2020

Contents

1. Definitions

Academic year	The year from 1 August to 31 July. Throughout the publication this is denoted in the format '2012/13' to describe the year from 1 August 2012 to 31 July 2013.
Advanced Learner Loan (ALL)	A fee loan payable to Further Education (FE) providers on behalf of FE learners who meet the eligibility criteria and started a FE course on or after 1 August 2013.
Cancelled loans	The borrower no longer has any liability to repay, as provided for in the loan's regulations.[1] A borrower's liability is cancelled: • On the death of the borrower; • On reaching the age or length of time cancellation criteria for their loan (which varies by loan product); or, • If borrower is in receipt of a disability related benefit and permanently unfit for work.
Capitalised interest	The interest accrued on student loans is added to a borrower's loan balance, rather than requiring repayment at the time it is accrued.
Doctoral loan	Loans issued to students on doctoral courses, on the Plan 3 repayment system. They are paid directly to students and can be used to cover fees or living costs.
Domicile	The usual residence of a student in the period prior to commencement of study. The financial support available to students from Government can vary for students from different domiciles. This publication includes forecasts of entrant numbers for English and EU domiciled students. Wherever 'EU domiciled' students are referred to this includes students domiciled in countries other than the UK that count as EU domiciled for funding purposes.
Entrants	Students in their first year of study. Defined as those starting a course in the academic year who have not been active at the same broad level of study at the same provider in either of the two previous academic years.
Face value of loan book	The total outstanding balance of the loan book. This will include all previous loan outlay and accrued interest, less any repayments or loan cancellations.
Financial year	The year from 1 April to 31 March. Throughout the publication this is denoted in the format '2012-13' to describe the year from 1 April 2012 to 31 March 2013. Some aspects of the student loan system are based on tax years (the 12-month period starting on 6 April), but as a simplification the student loan models assume that this is the same as the equivalent financial year.
Fully repaid loan	The borrower has repaid the loan in full during their repayment term without it being cancelled.
Higher education full-time loan	Loans available to students on full-time higher education courses, including first degrees, sub-degrees and certain postgraduate courses (e.g. Postgraduate Certificate in Education or PGCEs) that are eligible for the undergraduate loan system

[1] The Education (Student Loans) (Repayment) Regulations 2009

Higher education part-time loan	Loans available to students on part-time higher education courses with an intensity of 25% or higher.
Income Contingent Repayment (ICR) loan	Loans for which the required repayments are based on the borrower's income. The type of student loan that has been available to students since 1998.
Liable to make repayments	The borrower has a remaining loan balance and has reached their Statutory Repayment Due Date (SRDD).
Maintenance loan	Maintenance loans are loans to cover living costs, paid directly to the student.
Master's loan	Loans issued to students on master's courses, on the Plan 3 repayment system. They are paid directly to students and can be used to cover fees or living costs.
Plans 1, 2 and 3	The ICR loan scheme has been separated into different repayment arrangements called Plans 1, 2 and 3. While they operate in a similar manner, they differ in some ways such as the repayment thresholds, interest rates and the length of borrowers' repayment terms. Plan 1 is the loan system for undergraduate students that started courses before September 2012, Plan 2 the system for undergraduates since September 2012 and for Advanced Learner Loans, and Plan 3 the system for postgraduate loans introduced in 2016.
Resource Account Budgeting (RAB) charge	Used in the DfE annual accounts, this is the proportion of loan outlay that is expected to not be repaid when future repayments are valued in present terms.
Repayment term	The period for which a loan borrower is liable to make repayments based on their income. At the end of a borrowers' repayment term any remaining loan balance is cancelled.
Repayment threshold	The annual income threshold above which borrowers are required to make repayments on any eligible income. Plan 1 and Plan 2 loan borrowers are required to pay 9% of any earnings above the threshold and Plan 3 borrowers will be required to repay 6%.
Statutory Repayment Due Date (SRDD)	The point a borrower becomes liable to begin repaying a loan, normally the start of the tax year (6 April) after graduating or otherwise leaving their course. After their SRDD, borrowers are required to make repayments if their income is above the repayment threshold.
Stock charge	Used in the DfE annual accounts, this is the proportion of the total outstanding face value of the loan book that is expected to not be repaid when future repayments are valued in present terms.
Tax year	The 12-month period starting on 6 April. As a simplification, the student loan models assume that this is the same as the equivalent financial year running for 12 months from 1 April. Repayment thresholds are fixed for the duration of each tax year and borrowers' SRDDs are at the start of the tax year after they graduate or otherwise leave their course.
Transfer proportion	Under the partitioned loan transfer approach,[2] student loan outlay is partitioned into loaned and transferred funds. Conceptually the transfer proportion is the

[2] Student loans in the public sector finances: a methodological guide - partitioned loan transfer approach

	fraction of student loan outlay identified at loan inception as government expenditure, in recognition that this portion of the loan is unlikely to be repaid.
Tuition fee loan	Tuition fee loans are loans to cover all or part of the cost of tuition. They are paid directly to the learning provider.
Voluntary repayment	A borrower can at any time choose to repay some or all of their loan balance early, in addition to any repayments they are liable to make based on their income.

2. Background

Introduction

Income Contingent Repayment (ICR) student loans are provided by Government to higher education students and some further education students to cover course fees and living costs while they are studying. They were first introduced in the UK for new undergraduate students in 1998, at the same time as tuition fees. Prior to 1998, university students were provided funding by Government through a mixture of grants and, from 1990, mortgage style loans that were available to help with living costs. Mortgage style loans are not covered in this publication.

Each of the four constituent countries of the UK now have their own student loan policies, but only English student loans are considered in this publication. These are loans issued to English domiciled students that attend any learning provider in the UK and EU domiciled students that attend learning providers in England. A summary timeline of income contingent repayment loans is available in Table 2.1 below.

Table 2.1: Income Contingent Repayment loan timeline: England

1998	• Plan 1 loans introduced for new UK domiciled undergraduate students, to cover living costs • Annual tuition fee of up to £1,000 also introduced in 1998 Teaching and Higher Education Act.
2006	• Maximum annual tuition fee increased to £3,000 for new entrants • Tuition fee loans introduced • EU domiciled students became eligible to take out tuition fee loans • Repayment term changed to 25 years for new entrants, rather than ending at age 65
2012	• Plan 2 loans introduced for new entrants • Maximum annual tuition fee increased to £9,000 for new entrants • Part-time undergraduates become eligible for tuition fee loans
2013	• Advanced Learner Loans introduced for students aged 24+ on some further education courses, on the Plan 2 system
2016	• Plan 3 loans introduced for new students taking master's courses, who could borrow up to £10,000 over the length of their course • Maintenance grants replaced by additional maintenance loans for new undergraduate entrants on lower incomes • Advanced Learner Loans extended to students aged 19-23
2017	• Nursing, midwifery and most allied health students become eligible for student loans, in place of receiving NHS bursaries
2018	• Plan 2 repayment threshold increased from £21,000 to £25,000. It had previously been announced that it would remain at £21,000 until April 2021 • Doctoral loans of up to £25,000 across the length of a borrower's course introduced, on the Plan 3 system • Maintenance loans introduced for part-time undergraduates, for on-campus, degree-level students only
2019	• Plan 2 repayment threshold increased to £25,725 for tax year 2019-20, after which it will increase in line with average earnings growth figures published by the Office for National Statistics (ONS).

Student loans are issued by and administered by the Student Loans Company (SLC) on behalf of the Government and the devolved administrations in the UK. The Department for Education produces forecasts for its outlay on, and the repayments it expects to receive from, the English student loans that it is responsible for. These are for use in financial planning, policy development and to value the loans that have been issued in its annual accounts. The forecasts presented in this publication are produced by four models, as follows:

- Student numbers model – this model forecasts the number of full-time undergraduate entrants eligible for Office for Students funding in England. These forecasts are used in the student loan outlay and repayment models to estimate the future growth in full-time loan borrower numbers.
- Student loan outlay model – this model produces forecasts for loan outlay on higher education ICR loans, including those issued to undergraduates and postgraduate students.
- Student loan earnings and repayments model – this model produces forecasts for the future repayments that will be made by higher education ICR loan borrowers.
- Advanced Learner Loans model – this model produces forecasts for loan outlay and repayments that will be made on Advanced Learner Loans, which are available for some further education courses.

This document provides information on these four models, including the methodology, data sources and assumptions used in producing the forecasts.

These forecasts are based on currently announced government policy. Any changes to student loan eligibility, quantum or terms and conditions, if implemented by Government, could affect the forecasts presented in this publication.

These forecasts consider limited impacts of Covid-19. Economic forecasts are based on the *Central scenario* forecast by the OBR in their Fiscal Sustainability Report of July 2020.[3] However, the impact of (now removed) temporary Student Number Controls and the recent change to use Centre Assessed Grades (CAGs) for A Levels on the student finance entrant population, and the subsequent impacts of such changes in student numbers on student loans outlay and repayments, have not been considered.

[3] Office for Budget Responsibility - Fiscal Sustainability Report, July 2020

3. Student entrants model

Introduction

The student entrants model forecasts the full-time undergraduate entrant population eligible for tuition fee loans,[4] which is forecast as a subset of full-time undergraduate English domiciled entrants to UK providers and EU domiciled entrants to providers in England. The model assumes a constant proportion of loan-eligible entrants based on a three-year average of the proportion of loan-eligible entrants in HESA Student Record (for this publication, 2016/17 to 2018/19).[5]

Entrant outturn (inclusive up to 2018/19) includes entrants to HEIs only. However, to account for the new categories under the Office for Students (OfS) registration effective from 2019/20, the forecast from 2019/20 largely includes Approved (fee cap) providers. More specifically, the Approved (fee cap) forecast is based on HEIs (where the vast majority of which have registered as Approved (fee cap)) and former designated APs registered as Approved (fee cap), as of March 2020.[6] The full-time undergraduate students of these providers are eligible for tuition fee loans ranging between £9,000 and £9,250 per year, so the forecast aims to capture entrants eligible for these maximum tuition fee loans.

In summary, the forecast generally includes entrants to providers eligible to charge maximum tuition fees of £9,250 per year, making their students eligible for the higher tuition fee loan (of £9,250 per year). However, there are some further nuances around Access and Participation Plan (APP) and Teaching Excellence Framework (TEF) award which impact the maximum fees providers are able to charge under each registration category. Therefore, the forecast does not include *all* providers that could charge maximum fees and may also include *some* providers that cannot.

Since academic year 2017/18, nursing, midwifery and allied health profession (NMAH) entrants have been eligible to apply for student loans for tuition fees and maintenance costs. This funding policy change means that these loan-eligible entrants are also forecast alongside English and EU domiciled loan-eligible entrants.

In summary, the student entrants model forecasts:

- Full-time undergraduate entrants to HEIs and former designated APs registered as Approved (fee cap) in the OfS registration. These entrants are specifically English domiciled to UK providers or EU domiciled to providers in England.
- A subset of entrants which are eligible for tuition fee loans from Student Finance England.

Due to the registration process being an iterative process, with some providers having registered under the new categories and some currently remaining as HEIs and APs, there are a number of groups not included in the model's forecast. To summarise, the model does not forecast:

- Students registered for HE courses at Further Education Colleges (FECs), regardless of registration status with the OfS.[7]
- Students of:
    - Designated or non-designated APs not yet registered as part of the OfS registration.
    - Providers registered as Approved in the OfS registration.[8]
    - Former designated APs which as of March 2020 were unregistered in the OfS registration and may have registered as Approved (fee cap) since.

[4] Eligibility criteria is relative to the policy context for the academic year. For detail of eligibility per academic year, see the government legislation website.
[5] For this publication, eligible tuition fee loan entrant proportions have been applied from academic year 2016/17 to all forecasting years. Due to a change in eligibility for nursing, midwifery and allied health profession entrants from academic year 2017/18, these entrants have been included separately to the forecast from 2017/18.
[6] Analysis of these providers was conducted in March 2020, at the time of the production of this forecast. This means that former designated APs which have since registered as Approved (fee cap) have not been included.
[7] HE level entrants to FECs are not included in the forecast as the coverage for HESA data includes HEIs and APs only.
[8] Instead, assumptions about the number of entrant loan borrowers are made in the student loans outlay model. This is because these students are eligible for tuition fee loans of a maximum of £6,125 per year.

    - Former non-designated APs registered as Approved (fee cap) in the OfS registration.[9]
- Part-time students of undergraduate level study.
- Postgraduate students – both Level 7 (taught) and Level 8 (research).
- Non-English UK domiciled students.
- EU domiciled students at providers in Wales, Scotland and Northern Ireland.
- Full-time enrolments of undergraduate level study (i.e. continuing students).

For these students, assumptions are made about the number of loan borrowers in the student loans outlay model to forecast outlay for total full-time undergraduate tuition fee and maintenance, part-time undergraduate tuition fee and maintenance, masters and doctoral loans – see Section 4. These assumptions are based on historic SLC data of the number of loan borrowers and yearly growth.

Loan-eligible entrant growth rates from the student entrants model are applied to the latest year of SLC data in DfE's student loans outlay model, which inform the department's financial accounts regarding outlay of student loans through the SLC. The forecasts are also used by the Office for Budget Responsibility (OBR) in the Economic and Fiscal Outlook which forecasts public spending, including student finance over a five-year period.[10]

Methodology

The student entrants model includes the following data: ONS population estimates up to 2018, 2018-based ONS population projections,[11] UCAS undergraduate applicant statistical releases for 2020/21 cycle (15 January deadline),[12] UCAS end of cycle report of acceptances[13] 2019/20 and HESA Student Record 2018/19. ONS data apply to the English domiciled model only, all other data apply to both English and EU domiciled models.

The model forecasts entrants and loan-eligible entrants over a six-year period in three distinct modelling stages:

1. English domiciled entrants
2. EU domiciled entrants
3. Loan-eligible entrants (including the impact of policies for NMAH entrants).

English domiciled entrants model

English domiciled entrants are forecast by age and gender breakdowns before aggregating to a total. Generally, a linear regression model is used to forecast entrants (exceptions are detailed in the relevant sections). English domiciled entrants are generally calculated as

$$N_{Entrants} = R_{Entry} \times \left[\left(population \times R_{applicant} \times R^{ms}_{acceptance}\right) + N^{nms}_{acceptance}\right]$$

[9] The forecast also does not include former non-designated APs registered as Approved (fee cap) as very little is known about these providers and their student numbers. Also, the forecast does not consider future behavioural changes around the OfS registration for former designated APs not yet registered. It assumes that existing HEIs will continue to register as Approved (fee cap) but does not assume former designated APs (that have not yet registered) will register as Approved (fee cap). This is because the distribution of former designated APs among registration categories is currently unknown, especially at the provider level. The model therefore includes an updated version of the register at regular intervals to monitor changes in registration.
[10] Office for Budget Responsibility - Economic and Fiscal Outlook, March 2020
[11] Office for National Statistics - All data related to population projections
[12] UCAS undergraduate applicant releases for 2019 cycle
[13] UCAS 2018 end of cycle report - Summary of applicants and acceptances

where $N_{Entrants}$ is the number of English domiciled entrants, R_{Entry} is the entry rate, $R_{applicant}$ is the applicant rate, $R^{ms}_{acceptance}$ is the main scheme acceptance rate and $N^{nms}_{acceptance}$ is the number of non-main scheme acceptances.

There are some minor exceptions for some age groups depending on the data that is available, and these are discussed below. To forecast these entrants, the model calculates the volumes of four stages: population, applicants, acceptances, and entrants.

Population

ONS population estimates and projections by gender are aggregated by age groups 18, 19, 20 and 21-25 to provide yearly population estimates and projections for England. These estimates and projections are used as the base population for each linear model of these age groups. Subsequently, population data are not used in the models for age 17 and under and 26 and above.

Applicants (main scheme)

Main scheme applications refer to those submitted prior to the UCAS June deadline so do not include non-main scheme routes such as clearing. UCAS main scheme applicant statistical releases are aggregated by age groups 17 and under, 18, 19, 20, 21-25 and 26 and above by gender. For age groups 17 and under and 26 and above, these data are used as the base population.

For each age and gender breakdown for each academic year, the model calculates the historic proportion that have applied by the January and March deadlines compared to the June total of applicants. A three-year average of the growth in this proportion is applied to estimate the January and March proportions in relation to the June deadline for the current cycle (of which only January deadline data are available for). From this, total applicants for the remainder of the current cycle are calculated.

Applicants aged 17 and under are forecast by applying a three-year average of applicants to all forecast years. Applicants of all other age groups are forecast using a linear regression model on log-transformed values. For applicants age 18-25, these values are applicant rates (defined as applicants divided by population), from which, applicants are forecast for each academic year by multiplying population by the forecast applicant rate. For applicants age 26 and above, these values are applicant numbers.

The historical time series used in the regression for age groups 18 to 20 begins in 2014, and for age group 21-25 begin in 2013. The start year is different for the different age groups so that the data included in the time series are more reflective of recent applicant trends.

Acceptances

UCAS end of cycle report of acceptances is separated by acceptance type (main scheme and non-main scheme) and are aggregated by age groups 17, 18, 19, 20 and 21 and above by gender. This means that yearly applicants age 21-25 and 26 and above are combined for the remainder of the model.

Main scheme acceptances age 17 and under are forecast by applying a three-year average of main scheme acceptances to all forecast years. Main scheme acceptances of all other age groups are forecast using a linear regression model on a log-transformed acceptance rate (defined as main scheme acceptances divided by applicants), from which main scheme acceptances are forecast for each academic year by multiplying applicants by the forecast acceptance rate.

The historical time series used in the regression for age groups 18 to 20 begins in 2013, for age group 21 and above begin in 2015. The start year is different for the different age groups so that the data included in the time series are more reflective of recent main scheme acceptances trends.

For all age groups, the model calculates the historic main scheme and non-main scheme acceptances proportions of total acceptances and applies the proportions from the latest year of historic data to all

forecast years to forecast non-main scheme acceptances. Total acceptances per year are the sum of main scheme and non-main scheme acceptances of a given year.

Provider behavioural response

UCAS undergraduate applicant releases show that 18-year-olds account for around 50-55 per cent of English domiciled main scheme applicants each year.[14] Due to projected decline in the 18-year-old population up to and including 2020 (with the knock-on effect experienced for 19- and 20-year-olds in subsequent years), and because forecast (main scheme) acceptances are restricted by the applicant population, the model forecasts volatility in acceptances, and thus entrants, throughout the period 2019/20 to 2022/23.

A manual adjustment is applied to mitigate this, which assumes that providers will respond to any declines in applicants by increasing their offer rates (which filters through to acceptance rates). This assumption is based on historic UCAS data that suggest that each year accepted applicants grow more than applicants (Table 3.1).

Table 3.1: Historic UCAS applicants and acceptances annual growth rates
Full-time undergraduate English domiciled main scheme applicants and acceptances, academic years 2013/14 to 2019/20

	Annual growth						
Main scheme population	**2013/14**	**2014/15**	**2015/16**	**2016/17**	**2017/18**	**2018/19**	**2019/20**
Applicants	2.9%	3.4%	0.8%	-0.3%	-4.7%	-3.7%	-0.6%
Accepted applicants	6.5%	3.7%	2.7%	0.3%	-2.2%	-3.2%	-0.4%

Source: UCAS[15]

Although there is uncertainty regarding the extent of any provider behavioural response to a decline in applicants, historic trends in the UCAS data suggest that providers strive to grow, or at least maintain, student numbers each year.

The adjustment assumes linear growth from 2019/20 to 2022/23, with the end result having smoothed the volatility in the original forecast. This intervention is owned by OBR and adopted based on OBR practice. It is assumed sensible given how providers have increased their student numbers throughout a period of 18-year-old population decline (period of decline from 2015 to 2020 with HESA student numbers up to 2018/19 evidencing growth from 2015/16). Further research with providers would be required to confirm this judgment, however, providers have increased their student intake throughout the 18-year-old population decline, and as a market driven sector, growth (or at least strive at that) is highly likely.

Entrants

HESA entrants by gender are aggregated by the same age groups as acceptances and include both first degree and other undergraduate. This therefore includes Levels 4, 5 and 6.

Entrants aged 17 and under are forecast by applying a three-year average of entrants to all forecast years. Entry rates of all other age groups are forecast by applying a three-year average of growth in entry rate as constant to each forecast year, from which entrants are forecast by multiplying acceptances by entry rate. The use of a three-year average over a linear regression for this element of the model is a legacy function and is continuously being reviewed as part of the model's development.

Impact of Office for Students (OfS) registration

Prior to 2019/20, full-time undergraduate students of HEIs were eligible for tuition fee loans of £9,250 per year and full-time undergraduate students of APs were eligible for tuition fee loans for £6,125 per year.

[14] UCAS 2020 cycle applicant figures - 15 January deadline
[15] UCAS undergraduate end of cycle reports

Since 2019/20, English institutions offering higher education courses who desire designated status have been required to register with the OfS under one of two categories: Approved (fee cap) or Approved.[16] The maximum fees institutions registered as Approved (fee cap) are permitted to charge depends on whether they have a Teaching Excellence and Student Outcomes Framework (TEF) rating and whether they have an Access and Participation Plan (APP) that has been approved by the OfS.[17] Currently, Approved (fee cap) providers can charge a higher fee amount if they have an approved APP: up to £9,250 with and £9,000 without a TEF award, for full-time students. Approved (fee cap) providers can only charge a basic fee amount if they do not have an approved APP: up to £6,165 with and £6,000 without a TEF award, for full-time students. Providers in the Approved category are not subject to fee limits, but students attending these institutions are only eligible for student support up to the basic fee amount.[18]

Approved (fee cap) providers

Generally, HEIs have registered as Approved (fee cap), but there is some movement for APs, with some registering as Approved (fee cap). If a former designated AP registers as Approved (fee cap) it means they become subject to the tuition fee limit and their students eligible for maximum fee loans of £9,250 per year.

The impact assessed in the model is purely of former designated APs registered as Approved (fee cap) as of March 2020. This means that any former designated APs that have since registered, or are yet to register, as Approved (fee cap) have not been included. We do not model any additional students of former non-designated APs which may have already or may in future register as Approved (fee cap) as very little is known about these providers, particularly in terms of their student numbers. Therefore, this is not the total potential impact of entrants to Approved (fee cap) providers.

To estimate these entrants, the model relies on the latest version of the OfS register (March 2020) and matches Approved (fee cap) providers to HESA's AP Student Record 2018/19 to calculate the number of full-time undergraduate English or EU domiciled (to providers in England) entrants. Growth rates from the core entrants forecast are then applied to the 2018/19 total to forecast growth across the forecast period.

Whilst this methodology does not assume any future changes in registration behaviour, it is likely that more former designated APs will register as Approved (fee cap). However, it is not yet known which providers will and any estimation of which would be purely judgment based. Given that the number of entrants currently estimated in the impact is small (around 9,000 per year), there is little risk in not assuming changes in registration behaviour, as changes will likely be small as more providers continue to register. We are continuously updating this estimate as new versions of the OfS register become available to capture providers registered since.

Approved providers

Assumptions about the number of entrant loan borrowers of Approved providers are made in the student loan outlay model. These assumptions are based on equivalent analysis for Approved providers but do allow some additional capacity for former non-designated APs which may have already or may in future register as an Approved provider with the OfS. Formerly non-designated, the students of these providers were ineligible for student finance, so registration means that the entrants of these providers become eligible and will be new students to the student finance system. The entrants of these providers are accounted for in these assumptions because generally APs are expected to register under the Approved provider category, so it is a more sensible assumption that there will be additional entrants under this category than Approved (fee cap). It is important to reiterate that the number of students at these providers is unknown as non-designated providers are not required to submit data returns to HES. Proportionally, however, the assumed number of entrants to former designated vs. former non-designated providers roughly corresponds to the proportion of former designated to non-designated providers known, so is assumed a suitable assumption.

[16] Office for Students - Securing student success: Regulatory framework for higher education in England
[17] Student Finance England - Student Finance Memorandum
[18] An overview of basic and higher fee limits can be found at: Office for Students - Access and participation plans: Fee limits

EU domiciled entrants model

EU domiciled entrants to English providers are forecast as a total because the data included within the model are unavailable by age and gender breakdowns.

Applicants

UCAS applicant data are only available as applications (not applicants) for EU domiciled applicants. Applications are estimated for the current cycle (where only January or March deadline data are available) but are not forecast long term. To estimate total applications for the current cycle, the model applies a three-year average of the growth in the proportion of historic January and March applications in relation to the June total of applications and uses this proportion to estimate total applications for the cycle.

Acceptances

UCAS acceptances data are only available for total acceptances for EU domiciled acceptances, meaning main scheme and non-main scheme acceptances are combined. Total acceptances are forecast for the latest year of data that are available across the source data input to the model (i.e. the cycle year which UCAS applications are estimated for) but are not forecast long term. To forecast acceptance rate, a three-year average of acceptance rate growth is applied to the latest year of data across the model, from which acceptances are forecast by multiplying applications by the forecast acceptance rate.

Entrants

To forecast short term entry rate, a three-year average entry rate is applied to the forecast years between the latest year of HESA entrants data and the latest year of data across the model, from which entrants are forecast by multiplying acceptances by the forecast entry rate. Long term, entrants are forecast using a linear regression model on entrant numbers. The historical time series used in the regression begin from 2015.

Brexit and EU funding

Due to the uncertainty regarding future EU funding policy, the EU domiciled entrant forecast was held constant from 2020/21. In June 2020, the Government announced the removal of home fee status (HFS) and loan access in England for EU students from 2021 (there will be some entrants who are still eligible, depending on when they arrived in the UK and when they start their course). The forecast does not reflect this announcement.

Loan-eligible entrants

Loan-eligible entrants are forecast using a three-year average of the proportion of loan-eligible entrants, by domicile, in HESA Student Record (for this publication, 2016/17 to 2018/19).[19] The proportion for each domicile is assumed constant for all forecast years. As a proportion of entrants, both the English and EU domicile loan-eligible forecasts apply the yearly growth in forecast entrants (by respective domicile).

Nursing, midwifery and allied health profession (NMAH)

Forecasts of these entrants are not available by domicile. They are forecast separately to English and EU domiciled entrants to monitor the impact of the 2017/18 tuition fee reform and additional NMAH policies on entrant numbers.

[19] The latest entrant data used to inform this forecast are from 2017/18, meaning that no changes in eligibility criteria after 2017/18 are modelled to forecast entrants for subsequent academic years. The main implication of this is exclusion of postgraduate pre-registration healthcare courses, which became eligible for undergraduate support in 2018/19. However, the numbers of postgraduate students are much lower than undergraduate.

Baseline NMAH forecasts are a product of information provided by the Department of Health and Social Care (DHSC) on total clinical places funded by DHSC since the 2017/18 reform, and a judgment by the OBR on the take-up of these clinical places – informed by HESES/HEIFES data published by the Office for Students (OfS) on uptake of relevant courses in 2017/18 and 2018/19. The forecasts also include policies aimed to increase the uptake of NMAH entrants from 2020/21.

The NMAH forecasts are combined with English and EU domiciled loan-eligible forecasts to calculate the total loan-eligible entrants per academic year.

Long-term student numbers forecast

Beyond the six-year forecasting period, loan-eligible entrant growth rates are forecast in a long-term student numbers model (to 2100) as an input for the student loans repayment model. This model projects growth rates for loan-eligible entrants only. This specifically includes undergraduate (both full-time and part-time), masters and doctoral students.

In the long-term model, growth rates beyond the six-year period are calculated by weighting ONS principal population projections by age frequencies derived from the latest year of SLC data, where available, or by HESA data for the masters and doctoral loan.

The sum of male and female population projections per age group per year are calculated and compared against the reference year 2021 to derive population growth per age group per year. Age group weights, derived from SLC or HESA data, are then used to find the weighted cumulative average population growth per year per study group.

This is then converted to year-on-year change for forecasting growth rates and numbers beyond the six-year forecast period for each student population.

Data quality

The nature of any forecast is inherently uncertain and dependent on the quality of the source data, modelling methodology and assumptions made throughout. The forecasts use published data from ONS, UCAS and HESA to forecast full-time undergraduate entrants to HEIs and former designated APs registered as Approved (fee cap) only.

ONS Population Estimates and Projections

ONS population estimates are forecast on a calendar year basis. The model forecasts all subsequent components on an academic year basis (August to July). Although this misalignment may slightly impact applicant rates, it is assumed that ONS population forecasts are accurate enough for these purposes as the majority of full-time undergraduate entrants will commence study in the first calendar year of the academic year (thus matching the population data).

ONS publish data quality guidance on national population projections,[20] which are primarily the population forecast used by Government. Short-term principal projections are largely considered reliable; given that the student entrants model only forecasts over a six-year period, this increases confidence in the base population which entrant forecasts are modelled from. ONS do not make any predictions of future political or economic changes that could affect population numbers.

Post-2012/13 higher education data

Given the lack of suitable years' worth of UCAS and HESA data, and as a methodological limitation of linear regression, the model is sensitive to any significant anomalies to the trends in these data (such as

[20] Office for National Statistics - National population projections (quality and methodology information)

those caused by significant policy and funding changes). Due to the 2012/13 tuition fee reform to fees of £9,000 per year, the model only uses data from 2013/14 onwards to inform the forecast. The lack of data points increases the level of uncertainty in the model, particularly due to the complexity of the higher education environment (with a semi-regularly changing policy and funding context), which could result in volatile data.

Assumptions

It is uncertain how universities will respond to the knock-on effects of a declining 18-year-old population up to and including 2020, or whether future applicants will increasingly secure a place through non-main scheme routes (such as clearing), both of which will affect the model's applicant and acceptance rates. The model draws upon a limited number of data points (2013/14 to 2019/20) to inform its assumption regarding a provider behavioural response, and throughout a changing HE policy and funding context, so is unable to infer more robust conclusions.

Additionally, students may withdraw from a course after accepting a place but before they are classed as an entrant. The model assumes no changes in the proportions of applicants that have accepted a place withdrawing prior to entering a provider, although this could change in future.

Entrants where gender is recorded as 'Other' are combined with female entrants because females account for the largest group of entrants. This assumption requires review, particularly with an increase in non-binary gender identification, but it has been applied for this publication as a legacy assumption with minimal impact to the forecast.

EU domicile data

EU domiciled entrant forecasts are particularly uncertain because:

- UCAS applicant statistical releases are only available as applications. A single UCAS applicant can make up to five applications per academic year, and the average number of applications can vary over time, causing increased uncertainty over EU domicile forecasts. Given this constraint, the model assumes that EU applications are a suitable proxy for EU applicants for forecasting EU entrants.
- The forecast for EU domiciled entrants has been assumed flat from 2020/21 onwards. Although an unlikely outcome, without definitive funding policy (at the point of production of this forecast) the model is unable to make a more informed and robust forecast.
- Future EU domiciled applicant behaviour is particularly uncertain given the UK's withdrawal from the European Union and impacts from Covid-19. These factors have not been modelled in EU domicile forecasts.

4. Student loan outlay model

Introduction

The student loan outlay model forecasts loan amounts that the Department for Education (DfE) expects to pay higher education students (and their providers) via the Student Loans Company (SLC).

A range of sub-models are used to capture the various loan types available to students on higher education courses. The loan products that outlay forecasts are produced for are:

- Full-time undergraduate loans (Plan 2) – the loan system for students on full-time courses that started since September 2012 that are eligible for undergraduate student support funding, consisting of fee loans and maintenance loans.
- Part-time undergraduate loans (Plan 2) – the loan system for students on part-time courses that are eligible for undergraduate student support funding. These first became available in September 2012, consisting of a tuition fee loan. From August 2018 maintenance loans were also available to some part-time students.
- Master's loans (Plan 3) – loans available to master's students to help cover fees and living costs. They were introduced in August 2016 and are on the Plan 3 repayment system.
- Doctoral loans (Plan 3) – loans available to doctoral students from August 2018 to help cover fees and living costs. They are on the Plan 3 repayment system.

Outlay forecasts for Plan 1 loans are no longer produced due to the very small number of recipients expected in 2019/20, because most students who started a course before September 2012 have now completed those studies. The Student Loans Company recorded less than £500,000 of Plan 1 loan outlay in the tax year 2019-20.[21]

Documents detailing the availability and the student finance package for each loan product can be found at the Student Finance England Practitioners' Website.

The higher education student finance package for undergraduates is covered in detail in the Student Finance Package document and all rates of support are detailed in the Financial Memorandum. The eligibility criteria can be found in the Assessing Eligibility Guidance document.

Eligible English domiciled students are entitled to fee and maintenance loans for courses that are eligible for undergraduate funding. Eligible EU domiciled students are entitled to fee loans only. Both are entitled to the same amount for postgraduate loans.

Full-time and part-time undergraduate loans

Since 2019/20, English institutions offering higher education courses have been required to register with the Office for Students (OfS) under one of two categories: Approved (fee cap), or Approved.[22] The maximum fees institutions registered as Approved (fee cap) are permitted to charge depends on whether they have a Teaching Excellence and Student Outcomes Framework (TEF) rating and whether they have an Access and Participation Plan (APP) that has been approved by the OfS.[23] Currently, Approved (fee cap) providers can charge a higher fee amount if they have an approved APP: up to £9,250 with and £9,000 without a TEF award, for full-time students. Approved (fee cap) providers can only charge a basic fee amount if they do not have an approved APP: up to £6,165 with and £6,000 without a TEF award, for full-time students. Providers in the Approved category are not subject to fee limits, but students attending these institutions are only eligible for student support up to the basic fee cap of Approved (fee cap) institutions.[24]

[21] National Statistics: Student loans in England, 2019 to 2020
[22] Office for Students - Securing student success: Regulatory framework for higher education in England
[23] Department for Education - Loan, grant and tuition fee rates for academic year 2019/20
[24] An overview of basic and higher fee limits can be found at: Office for Students - Access and participation plans: Fee limits

Maintenance loans for eligible students depend on their location and household income (where a borrower applies for a means tested loan). The maximum maintenance loan for full-time Plan 2 borrowers living away from home and studying outside of London, in 2019/20 is £8,944.[25] Table 1A of the Student Loans Company statistical publication Student support for Higher Education in England presents the maximum rates of maintenance loans and tuition fee loans for full-time students domiciled in England.

Maintenance loans became available in 2018/19 to part-time, on-campus, degree students. These loans mirror the full-time maintenance loan, with the intensity of study taken into account alongside means testing and location. Students studying courses at less than 25% intensity are not eligible for part-time maintenance loans. The student loan outlay forecasts released in June 2019[26] assumed that maintenance loans would additionally be extended to distance-learning and sub-degree students in academic year 2019/20.[27] This will not be the case, and the model has been adjusted to reflect this. The forecast loan outlay for part-time maintenance loans decreases significantly compared with the figures published in June 2019 as a result of this amendment.

Postgraduate master's loans

The postgraduate master's loan was introduced in 2016/17. Eligibility for a master's loan depends on the duration and intensity of the student's course, their age on the first day of the first academic year of their course, and their nationality or residency status. The course must also be provided by a university or college in the UK, which is either publicly funded or a designated private provider. From 2019/20, English providers are required to register with the OfS as Approved (fee cap) or Approved to be eligible for student support funding.

Unlike undergraduate loans, master's loan entitlement for eligible students depends on the start date of their course, rather than location or household income. The maximum master's loan amount for a course starting in 2019/20 is £10,906 across the length of the course.

Postgraduate doctoral loans

The postgraduate doctoral loan was introduced in 2018/19. Eligibility for a doctoral loan is based on the duration of the student's course, their age on the first day of the first academic year of their course, and their nationality or residency status. The course must also be provided by a university or college in the UK, which is either publicly funded or a designated private provider. From 2019/20, English providers are required to register with the OfS as Approved (fee cap) or Approved to be eligible for student support funding.

Doctoral loan entitlement for eligible students depends on the start date of their course. The maximum doctoral loan amount in 2019/20 is £25,700 across the length of the course.

Methodology

Student loan outlay, for undergraduate higher education loan products (and including postgraduate Initial Teacher Training, which is also funded under Plan 2), is forecast based on historical data from the Student Loans Company. For postgraduate loan products, where historical information is limited, an alternative forecasting method is required.

Undergraduate higher education loan products

The student loan outlay forecasts for higher education loan products that are eligible for Plan 2 Higher Education funding use current and historical anonymised data on individual loan borrowers from the

[25] Loan for living costs for new full-time students and full-time students starting their courses from 1 August 2016 onwards who are continuing their courses in 2017/18 who are living away from home and studying outside London.
[26] Student loan forecases, England 2018 to 2019
[27] Consultation outcome: Part-time undergraduate maintenance loan: government response

Student Loans Company. This is a change from the methodology in previous years, where loan averages and information on aggregated groups by cohort were used. This year, individual-level data on loan borrowers were provided by SLC in April 2019 providing nearly complete information on student loans up to and including 2019/20. Using this and other information from SLC and growth rates from the student entrants model, forecast students were generated and allocated loans according to announced loan caps or OBR RPIX forecasts. Note that fee and maintenance loan levels available to students are typically already known for the first two or three academic years for which the model produces forecasts, currently 2019/20, 2020/21 and 2021/22. Maximum fee amounts in 2019/20 and 2020/21 have been set at the same levels as in 2017/18, while maintenance loan entitlements were increased in line with the July 2020 Office for Budget Responsibility (OBR) RPIX central forecast for the Jan-Mar quarter in each year.

This modelling method assumes that the distribution of characteristics of future loan borrowers, such as means-testing, gender and degree subject, is the same as for loan borrowers from the recent past. Internal analysis of historic trends indicates that this assumption is sound based on recent SLC data. However, it is possible that Covid-19 and the altered university admissions process resulting from the A-Level grading system used in 2019/20 may impact on the distribution of the characteristics of loan borrowers and on their loan amounts in the short-term. This is being monitored by DfE, and if needed, or where future policies are announced, an assessment will be made on the impact of the policy on the borrower numbers and loan amounts and the forecasts may be adjusted accordingly.

When the individual-level data was received from SLC in April 2019, it contained nearly complete data for 2018/19. To use this most recent year of data for modelling, estimates of the missing information until the end of the academic year, from April to August 2019, were made and the dataset amended accordingly. There are three main types of missing information between April and August, and these were quantified to produce a mini forecast of a full year of data for 2018/19. These three types of information are:

1. Missing borrowers. By April, most loan borrowers in the academic year have been paid a loan amount by SLC. However, some students may not yet appear in the dataset because their first loan payment is late in the year, for example if their course starts between April and August, or if a student's circumstances change and they require student finance when they did not earlier in the year. The number of missing borrowers was calculated using SLC data on the total number of students at the end of August 2019, split by institution type, study mode, student domicile, and loan product. This was compared to the number of students in the individual-level SLC data, and a number of students equal to the difference in the April and August totals was added to the individual-level dataset by randomly sampling and duplicating pre-existing entrants and amending the unique identifiers of the duplicated records. It was assumed that all missing borrowers after April 2018/19 were entrants (first year students).
2. Missing withdrawals. Some students will withdraw from university between April and August of an academic year, and this information will not have been contained in the April 2019 extract for 2018/19. In addition, there is a (non-negligible) time lag between a student's withdrawal and the HE provider notifying SLC of that event, so some information on withdrawals up to end April 2019 is missing too. To calculate the number of extra withdrawals in 2018/19, the 2019 extract was compared with the 2018 extract, with the 2019 extract providing an updated view on withdrawals in 2017/18; concretely, the proportion of missing withdrawals by study mode, institution type and loan product received was estimated. Within each of these combinations of characteristics, the expected number of extra withdrawals in 2018/19 was then randomly sampled and their records amended to add withdrawal information.
3. Missing loan outlay. Most loan outlay is paid prior to April in each academic year and is therefore recorded in the April cut of the SLC data. However, students who start later in the academic year, or whose circumstances change, may receive loan payments later than April and not all borrowers are paid their requested loan amount. An evaluation of the payments against the requested amount in 2017/18 (i.e. the most recent full academic year of data in the 2019 extract) is made to estimate the proportion of borrowers that will not get paid their requested amount after April; these proportions are calculated by study mode, institution type and loan product. For each of these combinations of characteristics, the April SLC data was searched for students who had received less than their requested loan amounts and had not withdrawn in 2018/19. A random sample of these students had their loan amounts increased to their requested amounts, based on the aforementioned proportions.

Once this missing 2018/19 data had been imputed, the now-full year of data was carried forward into the model.

For borrowers who have not finished their course yet, their original course length is not always representative of the number of years of loans a borrower will go on to receive. For example, a borrower may repeat a year, switch courses to include or remove a placement year, transfer onto a new course entirely (starting in year 1 again) or withdraw. To reflect this in the outlay forecasts, once the full year of 2018/19 data had been imputed, continuation rates reflecting average course lengths were applied to the latest and previous cohorts of students who were studying in 2018/19. These continuation rates were derived from 2014/15 and 2015/16 HESA student record data for full-time students, whereas for part-time students the continuation rates are based on 2016/17 and 2017/18 aggregate SLC data on part-time students taking up tuition fee loans. This adjustment was applied randomly within each study mode and for each original course length.

This data could now be used to generate students starting in future academic years. Entrant borrowers in 2018/19 were sampled, duplicated and renamed with new date information to generate entrants in future years up to 2024/25. The number of students that needed to be generated in each future year was calculated by applying the student entrants forecast growth rates to the number of 2018/19 entrants from aggregate SLC data. Different growth rates are calculated for part and full-time students and England and EU domiciled students, so the new students were generated for each combination of these characteristics. Some subsections of the student population were separated out and generated using a set entrants forecast, for example where a student numbers cap was in place. These sets of students included NMAH students, students studying Medicine and Dentistry, postgraduate Initial Teacher Training students, and students studying at alternative providers. These students were identified by a combination of JACS 3.0 code, course level and text mining course names. A small portion of each number forecast was set aside to generate students studying two or more funded courses.

Some students may be eligible to receive funding for more than one HE Plan 2 course; examples include postgraduate teacher training students, students completing a foundation degree before continuing to study for a bachelors-level degree, or students studying an ELQ-exempt course. Students studying more than one funded course therefore usually already have a loan balance when they start their second course. The entrants generated by the method above did not have pre-existing loan balances. In order to generate a realistic loan borrower population where some entrants each year already have a loan from a previous course, a random sample (by study mode and domicile) of previous loan borrowers were chosen and allocated to start new courses, with the proportion starting at different institution types, course levels and subjects (NMAH, Medicine and Dentistry, Initial Teacher Training, and all other courses) determined by an analysis of historic SLC data.

At this point in the modelling process, the SLC data consisted of a row per student per course with their loan outlay recorded up to the end of 2018/19 (some of which is estimated), as well as future students generated from current and previous students, and renamed and with their date information shifted forwards. With this data, future outlay can be generated.

To generate future outlay, the outlay of each student in the most recent year (2018/19) was uprated by the OBR forecast of RPIX or to the relevant loan cap, whichever was lower. Separate tuition fee caps were applied to students at Approved and Approved (fee cap) providers, set to the maximum loan amount for each type of institution. Maintenance loan caps were applied separately to medical and dentistry students in their fourth and fifth years (who are eligible for lower loans because they receive an NHS bursary in these years), and to all other students. Additionally, a random sample of students on courses of length 4 years were chosen to be sandwich placement students and were given lower fee and maintenance loans in their third years, where their third year was later than 2018/19. A random sample of entrants starting in 2018/19 who had fee loans below the cap in that year were chosen to be studying courses with fee waiver students, and allocated fee loans at the cap in their second year, even where this was an above-inflation increase. From 2020/21, a random sample of nurses was chosen to take out a reduced maintenance loan as a result of the re-introduction of NHS bursaries for NMAH students. The proportions of students who have sandwich placement, fee waiver or nursing bursary loan adjustments were determined by an analysis of SLC data or by agreement with the OBR.

The academic year loan outlay forecasts were converted into financial years. This was done based on an analysis of academic year payments and which financial year they fall in, using SLC data. This analysis took into account at which point within an academic year a student started. Each academic year straddles two financial years, and students starting later in an academic year will have a larger portion of their loan paid in the second financial year. At this point, forecasts of total loan outlay were produced.

Master's loans

Due to their recent introduction in 2016/17, historic master's loan borrower data is available at an aggregate level. Given the differences in the loan products, with the masters' loans entitlement for the course rather than each year, the method for forecasting future loan outlay uses a different approach to undergraduate loans. Instead, estimates of the total number of students who are likely to take up a loan each academic year are derived by assuming that the number of loan recipient entrants in 2018/19, based on SLC data, grows annually by 2% from 2019/20 to 2021/22, 1% from 2022/23 to 2023/24, and 0% in 2024/25.

Table 4.3: Core master's loans model parameters by course duration

Course duration	Proportion of loan recipient entrants	2016/17 average loan (per year)
1 year	80%	£9,300
2 years	17%	£4,400
3 years	3%	£2,700

Annual academic year loan outlay is calculated using a cohort approach, based on start year and the proportion of students within each course duration. Our model parameters are shown in Table 4.3, where average loan amounts are rounded to the nearest £100. The parameters are derived from SLC management information data for the 2016/17 borrowers; since last year's publication, the proportions of loan recipient entrants have been updated to bring the model forecast closer to outturn. The expected number of loan borrowers in each cohort is multiplied by a corresponding average loan amount and uprated by OBR forecast RPIX for entrants only, in each year. The sum of the outlay from each cohort is aggregated to produce a final academic year outlay figure. Financial year outlay is then calculated using the assumption that two thirds of the yearly loan amount each academic year is given in the first financial year it overlaps with (covering loan outlay from August to March), and the remaining third in the following financial year (covering April to July).

Doctoral loans

As doctoral loans were introduced in academic year 2018/19, historic doctoral loan borrower data is available for 2018/19 at an aggregate level. To qualify for a doctoral loan, the borrower's course must last between 3 and 8 academic years. For each course duration, estimates of the total number of students who are likely to take up a loan each academic year are calculated by assuming that the number of loan recipient entrants in 2018/19, based on SLC data, grows annually by a constant growth rate based on OfS estimates. Assuming that the characteristics of students taking up doctoral loans do not change year-on-year, the parameters for loan recipient entrants are displayed in Table 4.4. Here, we used provisional SLC management information data, as at the beginning of March 2019, to estimate the loan take-up split by course duration in 2018/19.

Table 4.4: Core doctoral loans model parameters by course duration

Course duration	Proportion of loan recipient entrants	Annual growth rate
3 years	43%	3%
4 years	33%	3%
5 years	6%	0%
6 years	10%	0%
7/8 years	8%	0%

The average loan (for the whole course) taken out is estimated to be £24,375 in 2018/19, assuming that 95% of borrowers receive £25,000 (which is the maximum maintenance loan entitlement in 2018/19) and the remaining 5% receive half of that amount. It is assumed that for new students the average loan will increase by forecast RPIX each year. Annual academic year outlay is then calculated using a cohort approach, based on continuation rates estimated from HESA cohort data and the proportion of students estimated to take up a doctoral loan as outlined above. The expected number of loan borrowers in each cohort is multiplied by the corresponding average loan amount, with the sum of the outlay from each cohort aggregated to produce a final academic year outlay figure. Like in the master's loans model, financial year outlay is then calculated using the assumption that two thirds of the yearly loan amount each academic year is given in the first financial year it overlaps with and the remaining third in the following financial year.

Long-term outlay forecasts

The methodology for the outlay forecast for the next five financial years for full-time (FT) and part-time (PT) undergraduate, master's and doctoral loans is documented above. After this, an alternative method is used to forecast the long-term outlay.

The proportion of total borrowers from each cohort, by product, shown in Table 4.5, is multiplied by the growth rate of that cohort (found by taking the entrant growth from the long-term student numbers model for the cohort when they started) to give a total student loan borrower growth rate for the academic year. This student loan borrower growth rate is then multiplied by the previous academic year outlay forecast and multiplied by forecast RPIX. The financial year forecast is then calculated by using the assumption that two thirds of the yearly loan amount of each academic year is given in the first financial year it overlaps with and the remaining third in the following financial year, for undergraduate maintenance and postgraduate loans; for undergraduate fee loans, the assumption is that the yearly loan amount of each academic year is split evenly between the two financial years that the academic year overlaps with.

For example, the master's loan outlay forecast for academic year 2025/26, $O^{AY}_{25/26}$, is calculated as:

$$O^{AY}_{25/26} = \left(1 + \frac{3}{4}G^{E}_{25/26} + \frac{1}{5}G^{E}_{24/25} + \frac{1}{20}G^{E}_{23/24}\right) \times \left(1 + RPIX_{25/26}\right) \times O^{AY}_{24/25}$$

where $G^{E}_{XX/YY}$ is the entrant growth for academic year XX/YY and $RPIX_{XX/YY}$ is the RPIX for academic year XX/YY. The outlay forecast for financial year 2025-26, O^{FY}_{25-26}, is then calculated as follows:

$$O^{FY}_{25-26} = \frac{2}{3}O^{AY}_{25/26} + \frac{1}{3}O^{AY}_{24/25.}$$

Table 4.5: Proportion of total borrowers by product and cohort

Cohort	Proportion of total borrowers (FT undergraduate)	Proportion of total borrowers (PT undergraduate)	Proportion of total borrowers (master's)	Proportion of total borrowers (doctoral)
1st year	35%	45%	75%	35%
2nd year	30%	30%	20%	30%
3rd year	25%	15%	5%	25%
4th year	10%	5%	–	10%
5th year	0%	5%	–	–

Data quality

Producing forecasts is inherently uncertain and they are very dependent on the data sources, modelling techniques and assumptions used in the model. In particular, the model assumes that the characteristics and behaviour of future borrowers will be similar to historic ones derived from SLC administrative data, which may not necessarily be the case.

The model is dependent on the July 2020 central OBR macroeconomic forecasts that it uses to uprate fee and maintenance loans. Any significant changes to the economy from these forecasts could affect the outlays that will be made on student loans.

The model uses SLC administrative data to determine borrower numbers and loan amounts. The DfE receives data extracts from the Student Loans Company on an academic year basis that are used in the student loan outlay model. This data is consistent with the data published in the SLC Student Support for higher education in England publication.

SLC publishes a statement on its administrative sources as well as data quality guidelines for its publications.

The model uses the growth in forecasted entrants from the DfE student entrants model; see Section 3 – Data Quality. The student entrants model largely forecasts full-time undergraduate entrants to Approved (fee cap) providers eligible for tuition fee loans from Student Finance England. This specifically includes former HEIs and designated APs registered as an Approved (fee cap) provider with the OfS, see Section 3 for more detail. This does not include students studying higher education courses at FECs registered as Approved (fee cap). The outlay model assumes that the growth in students studying higher education courses at FECs is the same as former HEIs and that the growth in entrants eligible for tuition fee loans from Student Finance England is the same for maintenance loans. If the growths in these populations are not consistent then this will have an impact on the number of new entrants who are taking out loans in the model.

The model assumes that fees and maintenance loans will be uprated by forecast RPIX in future years for which fee and maintenance loan levels have not yet been announced. The model also assumes that the current student finance policies will remain unchanged. Generally, once policies have been announced they are incorporated into the loan outlay models; however, the announced 2021/22 loan caps and entitlements have not been included yet, nor has the policy that EU domiciled students will no longer be eligible for home fee status, undergraduate and postgraduate financial support from Student Finance England for courses starting in academic year 2021/22. Therefore, any changes to the student finance policy will affect future forecasts. In particular we have been considering the Augar Report and its recommendations carefully. We plan to respond to the report alongside the Spending Review, providing the sector and students with certainty about the future of post-18 education and funding. Any changes to student loan

eligibility, quantum or terms and conditions, if implemented by Government, could affect the forecasts presented in this publication.

The forecasts for new loan products such as postgraduate loans and part-time maintenance loans are more uncertain as there is limited historical SLC data on which to base the forecasts. Instead, assumptions have to be made about the likely uptake of these products, which are more uncertain because the introduction of these products is likely to change student behaviours. The postgraduate borrower number forecasts do not capture any impacts from Covid-19.

5. Student loan earnings and repayments model

Introduction

The DfE student loan earnings and repayments model is the financial model used to estimate the financial cost of income contingent student loans to Government. It forecasts the repayments that the Department expects to receive on its expenditure on student loans.

The model is a micro-simulation model. It forecasts student loan repayments by estimating future earnings for individual student loan borrowers, and applies the loan repayment policy to each borrower, before aggregating the results to estimate totals for the population as a whole. For each loan borrower, it predicts their next year's earnings, and when this is repeated it generates an earnings path. Where historical information on earnings is available the model makes use of this. Earnings predictions are based on the borrower's level of study, gender, years since SRDD, age and other information. This allows the model to capture individual changes in earnings over the borrower's working lifetime.

Once a borrower's earnings have been forecast their repayments, interest and loan balances are calculated year by year for the length of their repayment term, or until they finish repaying their loan. Further adjustments are made to some borrowers' repayments to allow for investment income, voluntary repayments, overseas repayments, direct debit repayments, incorrect amounts being repaid, and loans being cancelled.

The model forecasts repayments for English domiciled students studying in the UK and EU domiciled students studying in England. Earnings forecasts are made for undergraduates (first degrees and sub-degrees) and PGCE loan borrowers based on historical administrative data for comparable loan borrowers and survey data on UK residents with a similar qualification level. For master's and doctoral loan borrowers, earnings are modelled by applying a percentage uplift to an earnings forecast for a comparable first degree student.

The main data sources used in the model are:

- Student Loans Company (SLC) administrative data – provides details of borrowers and the loans they take out, used to forecast earnings and employment status in early repayment years. Used for modelling migration, repayment frictions and repayments made directly to the SLC.
- Longitudinal Education Outcomes (LEO) - used in earnings and employment models in early repayment years.
- British Household Panel Survey (BHPS) data – used in earnings and employment models in later repayment years.
- 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.
- Office for National Statistics (ONS) life tables – data on deaths.
- ONS Average Weekly Earnings (AWE) data – used to adjust earnings between 2014-15 earnings values and nominal terms
- Higher Education Statistics Agency (HESA) data – course completion rates, characteristics information for borrowers taking out new loan products for which there is no historical SLC data.
- Office for Budget Responsibility (OBR) macroeconomic forecasts – forecasts of earnings growth, the Bank of England base rate, RPI and RPIX.
- DfE Student numbers model – forecasts of entrant numbers.
- DfE Outlay model – forecasts of student loan outlay.

Figure 5.1: Processes and sources underlying the student loan repayment model

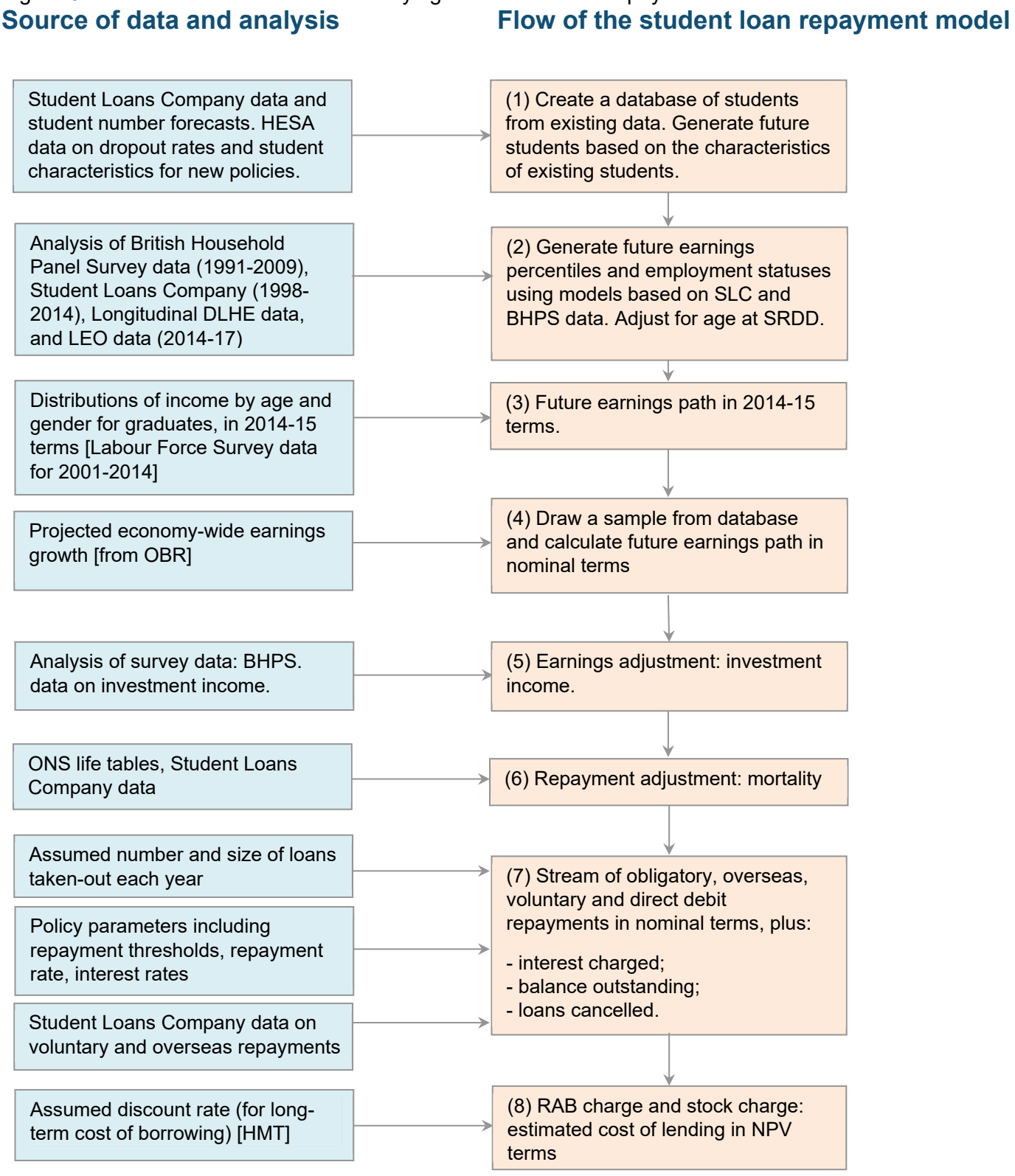


Figure 5.1 explains, at a high level, the processes that the model goes through to produce the forecasts, along with how each data source feeds into the full model.

Student loan repayment policies

ICR loans require borrowers to make repayments based on their annual income, starting from the April after they have left their course. Under each policy, borrowers are required to make repayments each tax year equal to a percentage of their income above a set repayment threshold until either they have fully repaid their loan balance, or their loan is cancelled. Loans are cancelled if the borrower dies, if they still have an outstanding loan balance at the end of their repayment term, or if they are in receipt of a disability related benefit and are permanently unfit for work. Loans accrue interest during and after the borrower's course, which is added to their loan balance.

A borrower becomes liable to repay their loan on the 6 April (start of the UK tax year) after they complete or withdraw from their course, at which point their repayment term starts on what is known as their Statutory Repayment Due Date (SRDD). There are two exceptions to this:

- Part-time loan borrowers will enter repayment at the start of the tax year after four years have elapsed since the first day of the first academic year of the course, even if they are still studying.
- When a loan product is first introduced the earliest SRDD for some borrowers may be later than it would usually be. For example, all Plan 2 borrowers that completed or left their courses before April 2016 had an SRDD of April 2016, even though under the usual rule some would have had an SRDD up to three years earlier.

A summary of the key repayment policy details for each loan product is shown in table 5.1 below.

Table 5.1: Key policy details for each loan product

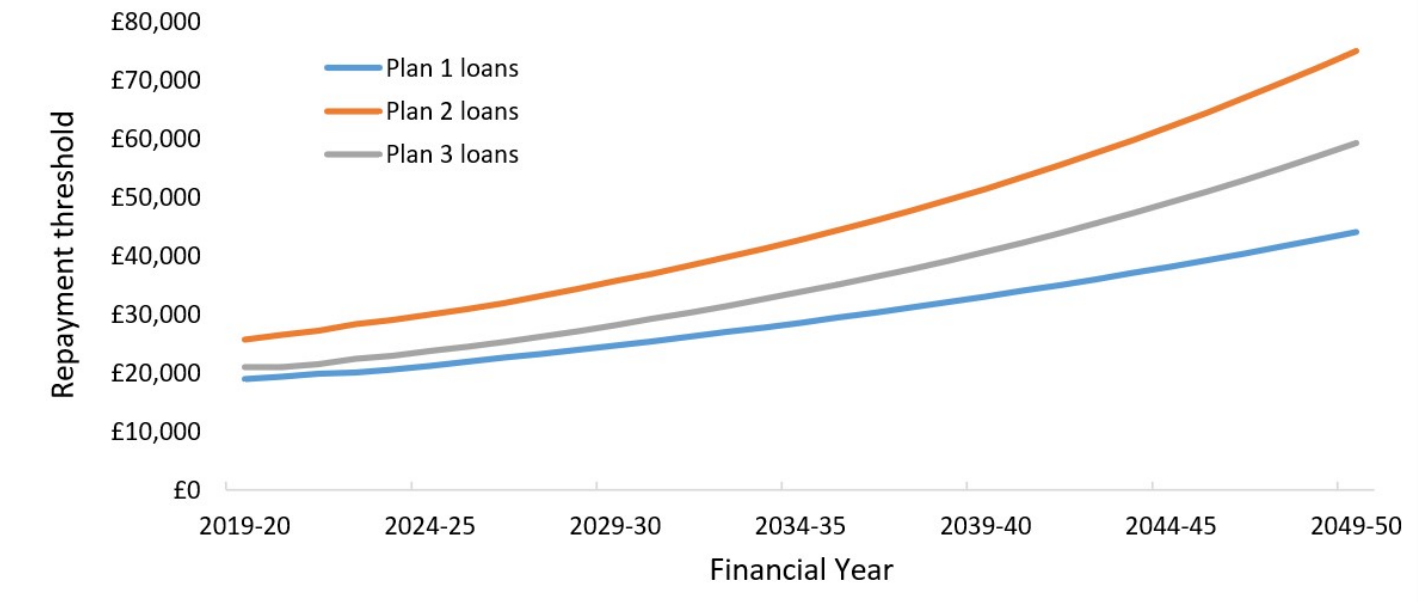
	Plan 1	Plan 2	Plan 3 (Postgraduate)
Earliest year of entrants	1998/99	2012/13	2016/17 (master's) 2018/19 (doctorate)
Earliest SRDD cohort	April 2000	April 2016	April 2019 (master's) April 2020 (doctorate)
Length of repayment term	Until age 65 (entrants up to 2005/06); 25 years after SRDD (2006/07 entrants onwards)	30 years after SRDD	30 years after SRDD
Repayment rate	9% of earnings above repayment threshold	9% of earnings above repayment threshold	6% of earnings above repayment threshold (in addition to any Plan 1 or Plan 2 repayments)
Interest rate	The lower of either RPI, or the Bank of England base rate +1%	RPI+3% during course, variable between RPI and RPI+3% after SRDD depending on earnings	RPI+3%

Each loan product has a separate income repayment threshold, above which repayments are made. Figure 5.2 shows the forecast repayment thresholds for each policy. All historic Plan 1 and Plan 2 threshold levels are published. The Plan 1 threshold is set at £18,935 for tax year 2019-20, and subsequently increases

each year based on RPI. The Plan 2 threshold was initially £21,000 for two years before rising to £25,000 in 2018-19 and was set at £25,725 for tax year 2019-20, after which it will increase in line with average earnings growth figures published by the Office for National Statistics (ONS). If a borrower has Plan 1 and Plan 2 loans, they pay back 9% of income over the Plan 1 threshold. If their income is under the Plan 2 threshold repayments only go towards the Plan 1 loan. If income is over the Plan 2 threshold, repayments go towards both loans. The Plan 3 repayment threshold is £21,000 from when the first borrowers become liable to repay in 2019-20 until 2020-21, after which it will be reviewed. To enable future repayments to be forecast, for modelling purposes it is assumed that from this point it will rise in line with Office for National Statistics (ONS) average earnings growth statistics in the same way as the Plan 2 threshold. The student loan repayment model forecasts future repayment thresholds using OBR forecasts for RPI and average earnings growth.

Figure 5.2: Forecast repayment thresholds for each loan product

England, financial years 2019-20 to 2049-50



In addition to the repayment threshold, Plan 2 also has two interest thresholds. Once Plan 2 borrowers are past their SRDD their interest rate varies depending on income. If their income is below the lower interest threshold their interest rate is RPI, above the upper interest threshold it is RPI+3%, and for anyone with an income in between it varies linearly between the two. The lower interest threshold is the same as the repayment threshold, while the upper interest threshold was initially £41,000 before increasing to £45,000 in 2018-19, and £46,305 in 2019-20. In subsequent years both thresholds will rise in line with ONS average earnings growth statistics.

Student loan borrowers resident in the UK generally make their loan repayments through the tax system to Her Majesty's Revenue and Customs (HMRC), either in-year through their employer via Pay As You Earn (PAYE) or the following year via a Self-Assessment tax return. Borrowers resident overseas are required to contact the Student Loans Company (SLC) and arrange to make repayments directly to them. Borrowers can also choose to make early repayments on their loan directly to SLC, and when a borrower is close to fully repaying their loan SLC will alert them and, to avoid over-repaying via the tax system, they can arrange to make their repayments via direct debit directly to SLC rather than through HMRC.

Methodology

Loan borrower population

A population of past and future loan borrowers is created containing information about borrowers' loan amounts, their courses, and various other information about them. To forecast a borrower's earnings the model needs data on their characteristics such as:

- Higher education provider group (see list of higher education provider groups below)
- subject group, based on subject area codes defined by HESA[28] (see list of course subject classifications below)
- course level: sub-degree, first degree and PGCE level
- age,
- SRDD, and
- up to three years of actual earnings and employment history, where available.

The higher education provider groups used in the student loan earnings and repayments model are as follows, with examples of institutions listed:

- Russell Group: Oxford, Cambridge, Leeds, Manchester, Nottingham, Birmingham, Sheffield, Cardiff, Southampton, Newcastle, Liverpool, Edinburgh, Queens (Belfast), Durham, Exeter, Bristol
- 1994 Group: Loughborough, East Anglia, Leicester, Lancaster, Sussex, Essex, Goldsmiths, Royal Holloway, IoE, SOAS, Birkbeck
- University Alliance: Manchester Metropolitan, Sheffield Hallam, Nottingham Trent, UWE, Liverpool John Moores, Northumbria, Plymouth, De Montfort, Portsmouth, Kingston, Hertfordshire
- MillionPlus: Leeds Metropolitan, Central Lancashire, Wolverhampton, Middlesex, Birmingham City, London Metropolitan, East London, Staffordshire, Derby, Sunderland
- GuildHE: Southampton Solent, Worcester, York St John, Winchester, Chichester and many arts university colleges
- Large non-affiliated: Brighton, Hull, Westminster, Kent, Edge Hill, Brunel, Strathclyde, Reading, Swansea, Roehampton, Gloucestershire, Bath, Heriot-Watt
- Small non-affiliated: Numerous small colleges

Course subject classifications, with typical subjects of study, are as follows:

- Medicine and Dentistry: Medicine, Dentistry (both pre-clinical and clinical)
- Subjects allied to Medicine: Anatomy, Pharmacy, therapies, nutrition, optometry, audiology, nursing, medical technology, environmental health
- Biological Sciences: Anatomy, Pharmacy, therapies, nutrition, optometry, audiology, nursing, medical technology, environmental health
- Veterinary Sciences, Agriculture: Veterinary Medicine and Dentistry (both pre-clinical and clinical), animal science, agriculture, forestry, food studies
- Physical Sciences: Chemistry, Materials Science, Physics, Forensic Science, Astronomy, Geology, marine sciences, physical geography
- Mathematical Sciences: Mathematics, Operational Research, Statistics
- Engineering: General engineering, civil engineering, mechanical engineering, aerospace engineering, naval architecture, electrical engineering, production engineering, chemical engineering
- Computer Sciences: Computer science, Information systems, Software engineering, Artificial Intelligence, health informatics, Games, Computer-generated audio & visual effects

[28] Further details on subject area codes are available from HESA: JACS 3.0: Principal subject codes

- Technologies: Minerals technology, Metallurgy, Ceramics & Glass, Polymers, Textiles, Materials technology, Maritime technology, biotechnology
- Architecture, Building & Planning: Architecture, Surveying, Building, Landscape design, Planning
- Social Studies: Economics, Politics, Sociology, Social Policy, Social Work, Anthropology, Human geography, Development studies
- Law: Law by area, law by topic
- Business & Administrative Studies: Business Studies, Management, Finance, Accounting, Marketing, HR management, office skills, hospitality/tourism
- Mass Communication and Documentation: Information Services, public relations, Media studies, Publishing, Journalism
- Linguistics and Classics: Linguistics, Literature, English studies, Ancient language studies, Celtic studies, Latin studies, Classical Greek studies, Classics
- European Languages and Literature: French studies, German studies, Italian studies, Spanish studies, Portuguese studies, Scandinavian studies, Russian and East European Studies, European Studies
- Other Languages and Literature: Chinese studies, Japanese studies, South Asian studies, Asian studies, African studies, Modern Middle Eastern studies, American studies, Australasian studies
- Historical and Philosophical Studies: History by period, History by area, History by topic, Archaeology, Philosophy, Theology and religious studies, Heritage studies
- Creative Arts and Design: Fine art, Design studies, Music, Drama, Dance, Cinematics and photography, Crafts, Imaginative writing
- Education: Teacher training, research and study skills in education, academic studies in education
- Combined courses and others not coded: Combined or unknown subject area

For existing loan borrowers this information comes from SLC administrative data, with some adjustments made to course lengths for those students still on courses using HESA data to simulate some dropping out or changing the length of their course in the future. For those still on courses, future loan amounts are estimated to be the same as their most recent year of loans, uprated by forecast RPIX where appropriate. Some adjustments are also made based on historical SLC data to allow for medical students receiving reduced loan amounts in the 5th and 6th years of their courses and for some “sandwich” students who receive different loan amounts in their placement year.

Some historical loan borrowers are chosen at random to start new courses in future years for which they take out student loans (e.g. a graduate may subsequently take a PGCE course for which they can receive a further loan), while the characteristics of all other loan borrowers starting on courses in future years are created by choosing borrowers at random from the most recent historical year of entrants. The number of entrants assumed to start courses each year for which they receive student loans is estimated by applying the annual growth rates from the student entrant numbers in the DfE student numbers model to the number of borrowers in the most recent year of entrants in the SLC data.

Where a new policy is being introduced and no information is available on historical loan borrowers’ characteristics, the distribution of characteristics for future loan borrowers are estimated based on historical HESA data for students on the courses for which loans will be available, with average loan amounts assumed for each borrower based on the DfE student loan outlay model.

Earnings forecasts

The part of the student loan repayment model that forecasts earnings is known as the earnings model. The earnings model estimates earnings and earnings status for each individual for 43 years, by which time

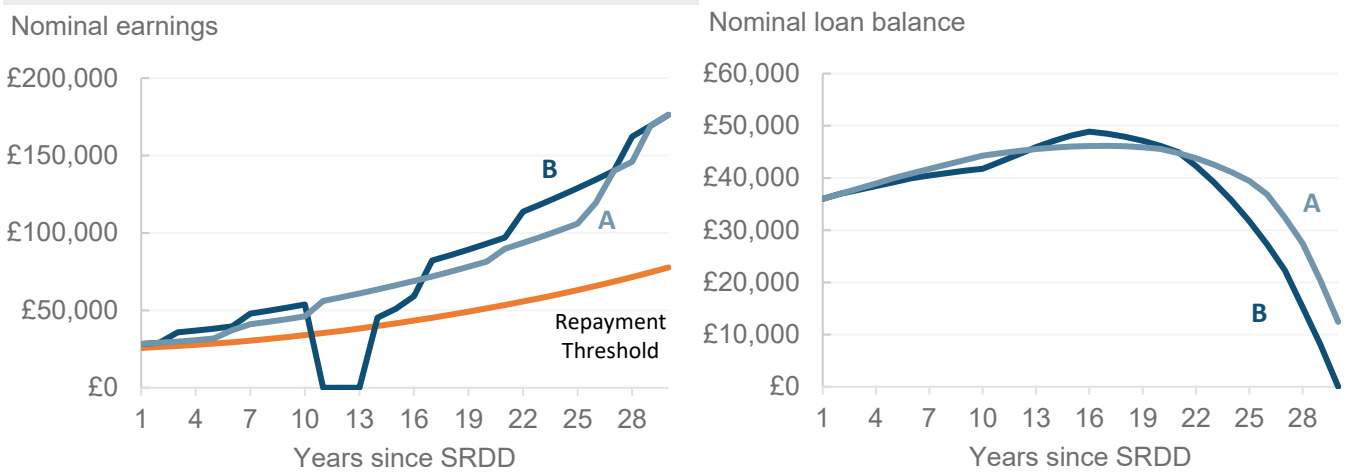
borrowers are highly likely to have reached retirement. This also allows for the different repayment periods of different loan plan types.

The earnings model is divided into two parts, based on the number of years between the forecast year and the SRDD. The first 10 years after SRDD are predicted using the 'short-term earnings model' and years 11 to 43 are predicted using the 'long-term earnings model'. Annual earnings are forecast for each borrower, and this step is repeated over subsequent years, to generate an earnings path for each individual.

An estimate of the earnings path of loan borrowers is necessary to estimate a borrower's repayments across their repayment term. As loans generate interest throughout their repayment period, the path of an individual's earnings, rather than their total earnings over the repayment period, has a significant effect on the amount of the loan the borrower will repay.

Figure 5.3 depicts two hypothetical earnings paths, A and B, for a loan borrower with an SRDD of 2019 and a nominal loan balance at their SRDD of c. £35,000. In both cases the individual has the same total nominal earnings over the 30 year repayment period however in scenario B the loan is completely repaid, whereas in scenario A part of the loan (c. £12,000 nominal loan balance) is written off.

Figure 5.3: Earning and loan balance paths for two hypothetical loan borrowers A and B



Short-term earnings forecasts

Earnings in each of the first 10 years after SRDD are estimated by matching borrowers to a training dataset of historic borrower earnings. The earnings and borrower characteristics in this dataset are derived from two sources:

- Student Loans Company (SLC) administrative data from 2001 to 2019
- Longitudinal Education Outcomes (LEO) data from 2014 to 2017

The SLC data includes earnings and employment status in each tax year following a borrower's SRDD as well as characteristics of the borrower, such as subject of study, provider and course level. SLC data does not, however, include the earnings of those who have fully repaid their loan and therefore, once high earners start to repay their loans in full, the dataset provides an unbalanced picture of the earnings of the graduate population. To overcome this, where borrowers have repaid their loans earnings are taken from the LEO dataset where available, based on fuzzy matching between the SLC and LEO datasets. The model is trained on a subset of the data, from 2014 to 2017, containing earnings and characteristics for borrowers up to 10 years after SRDD.

To create an earnings path for each borrower, the model searches the training dataset to find the most similar individual (or group of individuals) and uses their known earnings in the following year to estimate the future earnings of the target borrower. Where multiple individuals in the training dataset are matched,

the model selects one at random. The variables that define the proximity between individuals are the number of years since SRDD, their last three years of earnings (when available), Higher Education Institute type, domicile (England or EU), subject of study, gender, age and whether or not the borrower withdrew from studies before graduation. The "level of similarity" is measured by Euclidean distance.

Long-term earnings forecasts

As the combined SLC and LEO data only extends to 10 years after graduation, a different data source is needed on which to base estimates of earnings later in a borrower's career. We use the BHPS (from 1991 to 2009) for this purpose. Long-term earnings (those from the 11th year after SRDD onwards) are estimated using BHPS based regression models, adjusted to follow the LFS earnings distribution. The regression models are based on borrower characteristics and their earnings in the previous three years.

The BHPS[29] is a longitudinal study, over up to 18 years, of a representative sample of around 10,000 individuals. It interviewed each adult in a household and included questions on earnings and qualification level. The data on GB residents from this survey is used to construct regression models for forecasting employment status and earnings and mapping the earnings distribution of non-recent graduates.

Due to differences in survey design whilst the BHPS gives a better indication of individual earnings trends over many years, the LFS is a more reliable source of actual earnings of an individual given their age and gender. The final earnings estimates from the model are aligned with an earnings distribution created using aggregated LFS data.

The long-term earnings model forecasts an individual's earnings each year in a two-step process. The model first estimates whether an individual will be in employment or not, and then, for those in employment, estimates annual earnings.

Estimates of earnings are likely to be influenced by the level of qualification of individuals. As such, the long-term earnings model is split into four sub-models, based on the qualification level of the borrower:

- Sub degree qualifiers
- First degree qualifiers
- PGCE qualifiers
- Dropouts

To reflect the different qualification levels, each sub-model uses a different subset of the BHPS data when deriving equations or earnings distributions. The subsets of BHPS data used for modelling each sub-population of borrowers are as follows:

- Sub-degree qualifiers: HND/HNC or equivalent
- First degree qualifiers: First or higher degree
- PGCE qualifiers: First or higher degree
- Dropouts: A levels or equivalent.

Prior to estimating an individual's earnings, the model assigns borrowers to one of four employment states:

- Employed
- Non-employed
- Migrated
- Other zero (inactive or not repaying for a reason other than migration or non-employment, such as

[29] Further information on the BHPS can be found on the University of Essex website

an incomplete tax return or inconsistent data)

The probability of being in each of the four states is conditional on the borrower's characteristics, prior employment states and prior earnings. Up to three years of prior employment states may be used when estimating future employment states. The data is used to build binary logistic regression models predicting the likely employment state of borrowers. These regression models produce the log odds of being in each employment state, which is then transformed into a probability. The probability of being in each of the states is then scaled so that the sum of the four states is equal to 1.

A borrower is likely to be assigned to the state corresponding to the highest probability, however, to take into account the variation that cannot be explained through the borrower's characteristics random variation is also included in the employment states. This perturbs some borrowers away from the expected employment state.

If a borrower is in the "migrated" or "other zero" state, then an additional logistic regression model is run to determine whether the borrower is also employed or non-employed. Once the long-term earnings model has assigned an individual to an employment state it moves on to estimating earnings. Earnings are estimated for borrowers who are employed, including those who are employed but also migrated or in the other zero state. Borrowers who are predicted to have a non-employed state are estimated to have zero earnings in that forecast year.

Earnings are estimated through linear regression models using maximum likelihood estimation (general linear models). The regression models use a square-root transformation of earnings as this creates a more normal distribution of earnings compared to the natural log transformation. The outputs from the regression models can be squared to provide the actual earnings estimate. The earnings regression models also contain a stochastic element to perturb an individual away from the expected earnings given their prior employment history; this ensures the full distribution of earnings across the population is captured in the model.

To account for the different earnings profiles of men and women the model uses different regression models for each gender. Lookup tables are then used to adjust estimates to reflect the distribution of earnings exhibited in the LFS.

The earnings model also uses two other data sources for adjustments to the model. The Average Weekly Earnings (AWE) index, published by the ONS, is used to re-baseline survey and SLC data to earnings in 2014-15 financial year terms. The Destination of Leavers from Higher Education (DLHE) is used to estimate the impact of age at SRDD on a loan borrower's earning profile. This is to take into account that a recent graduate aged 30 for example is unlikely to have the same earnings path or earnings growth as a borrower aged 30 who graduated aged 22.

Estimates of earnings often differ between surveys, due to differing survey aims and populations. As such the data from the BHPS on graduate earnings follows a different distribution to that in the LFS. For example, the median earnings of a 25-year-old female graduate in 2014-15 values are £23,951 according to the aggregated LFS data[30] (2001-2014), but this is equivalent to only the 44th percentile in BHPS data (when aggregated from 1990-2009).

As such, the BHPS based earnings estimates are corrected to follow the LFS earnings distribution. This adjustment is accomplished by considering the percentile of the earnings estimated by the BHPS based regression models for a graduate of that age and gender and finding the equivalent earnings for the same person at the same percentile in the LFS data. For example, the BHPS regression model may estimate a female graduate, age 25, to have median earnings (for her age and gender) of £25,337 in 2014-15 values. However, we output the equivalent median earnings for a 25-year-old female graduate in the LFS data (£23,951 in 2014-15 values), rather than the BHPS based estimate.

[30] Comparisons of LFS data to administrative data on earnings from the Annual Survey of Hours and Earnings (ASHE) have shown that whilst the distribution of earnings in the LFS is noisier than in ASHE the profile is very similar when considering the whole population (Nanton & Rowling, 2017).

Age at graduation

The earnings profile of a new graduate aged 22 and a new graduate aged 45 may be quite different; typical borrowers may have more graduate level work experience than mature borrowers do once they reach the same age.

We cannot take account of this in the regression models derived from BHPS data as we do not always know the age at which a respondent graduated. Instead we take account of the impact of the age of the borrower at SRDD by deriving a second earnings forecast using a more typical age (21 or 22 depending on degree type), and a weighted average of the two earnings forecasts is taken. The weights initially result in more weight being given to the typical age earnings forecast and more weight to the actual age forecast as the borrower gains more experience (approximated by the number of years since SRDD). Weights are derived from DLHE data on the earnings of borrowers 6 months and 42 months after graduation.

Retirement

For some borrowers the 43 years of forecast earnings may include a period of retirement. Borrowers are assumed to retire at 65, after this age their earnings are modelled as zero for the remainder of the forecast years.

During the time period over which the BHPS data was collected, State Pension Age for men was 65. As such, we have very little data on the employment habits and earnings paths of individuals who remain employed after this age, and no information from the BHPS on the impact of raising State Pension Age. Given the lack of reliable data and that most borrowers will have reached the end of their repayment term before age 65, we have chosen to fix the retirement age in the model at age 65.

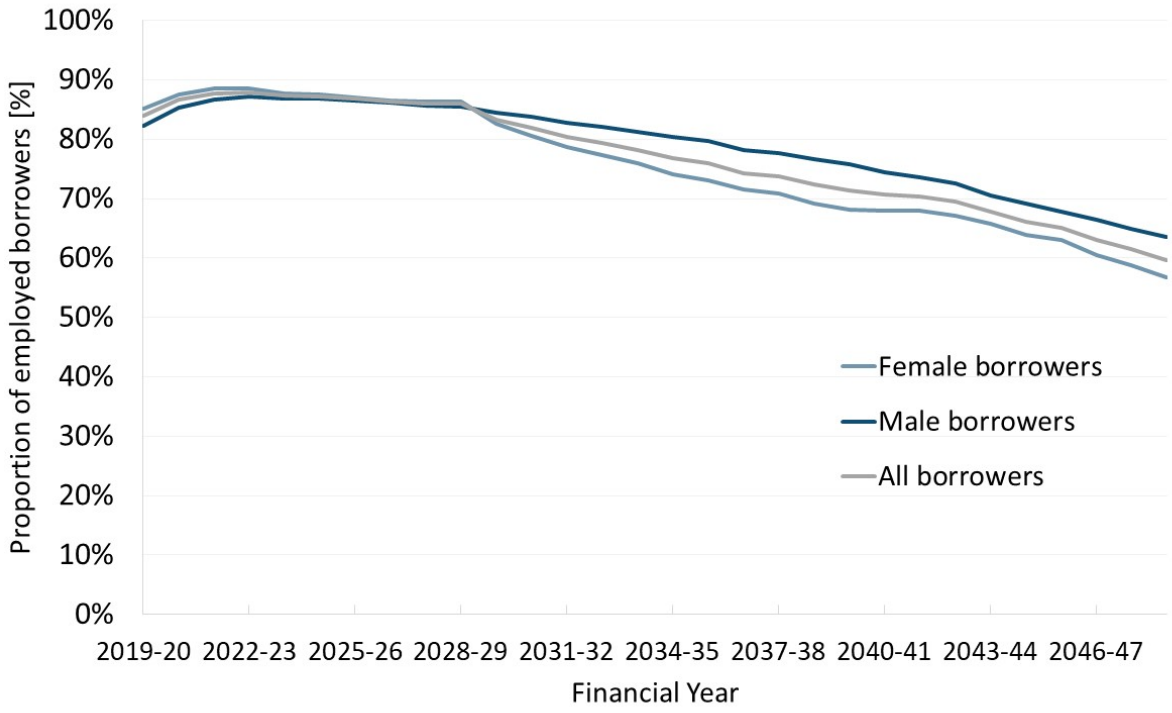
Earnings forecast outputs

At this point in the modelling we have generated annual equivalent earnings for those in employment in the UK at an individual level. These can be aggregated to provide insights to the earnings forecasts underlying the repayment forecasts.

For those borrowers with an SRDD of 2019 we forecast the proportion of borrowers employed (in the UK) over 30 years (Figure 5.4). We find that employment rates peak in the 4th year past SRDD and that during the first ten years after SRDD female borrowers are forecast to have marginally higher employment rates than male borrowers. Employment rates across all borrowers decrease after the 4th year past SRDD. The gap between male and female employment rates increases after 10 years to a long-term average of 7 percentage points. The model assumes that the employment decision of individuals in the future will be similar to those taken historically as derived from survey data and SLC administrative data. This includes historic patterns for employment by gender, which will not necessarily be the case, especially in the long term.

Figure 5.4: Forecast proportion of borrowers in employment in the UK by gender: 2019 SRDD cohort

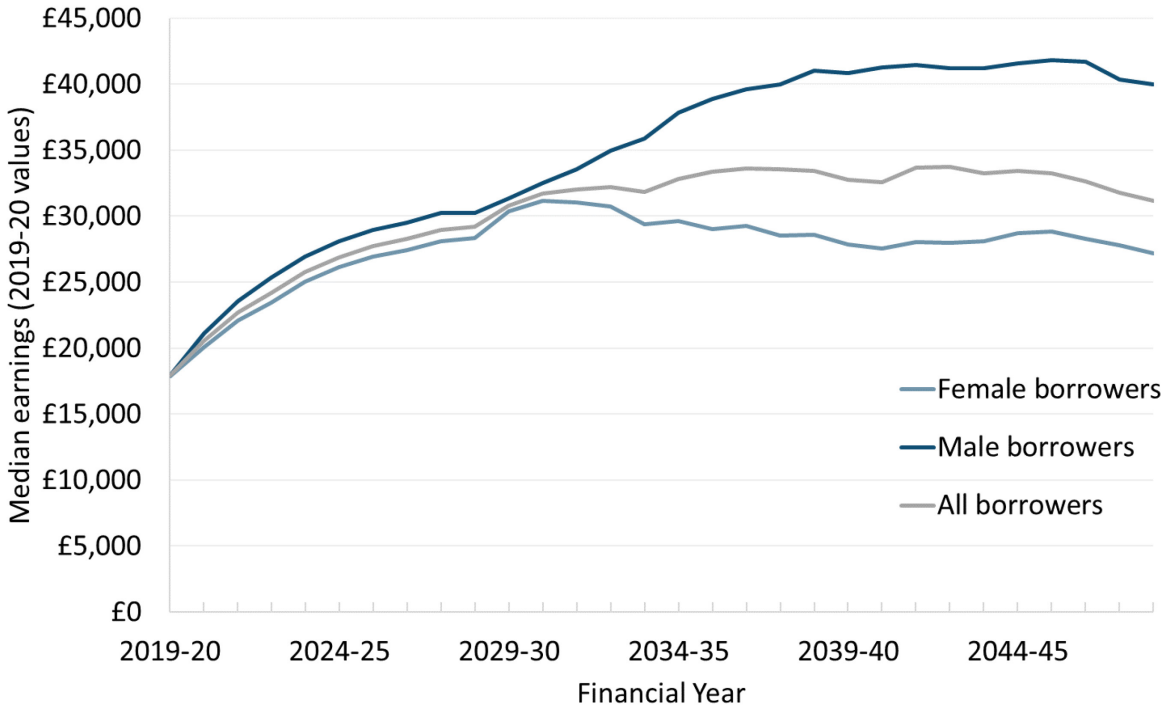
Financial years 2019-20 to 2047-48



For those with an SRDD of 2019 who are forecast to be employed, Figure 5.5 shows the median forecast earnings each year in 2019-20 values. Forecast median earnings for the first year past SRDD are £17,900 in 2019-20 values with little difference between male and female earnings. Median earnings for male and female borrowers diverge over the loan term growing to an average gap of £13,100 in 2019-20 earnings values over the last 10 years of the loan term. The model assumes that the earnings paths of individuals in the future will be similar to those taken historically as derived from survey data and SLC administrative data. This includes historic patterns for different distributions of earnings by gender, which will not necessarily be the case, especially in the long term.

Figure 5.5: Forecast median annual earnings of those in employment by gender: 2019 SRDD cohort

Financial years 2019-20 to 2047-48, earnings in 2019-20 earnings values



The distribution of forecast earnings by gender is further broken down in Figures 5.6 and 5.7. The distribution of female earnings is closely grouped around the median, with 50% of female borrowers earning within £11,000 of median earnings. In comparison the male earnings distribution is more dispersed, with the top decile of male borrowers earning over the 30-year loan term on average 1.4 times that of female borrowers in the top decile.

Figure 5.6: Forecast distribution of annual earnings of female borrowers in employment: 2019 SRDD cohort

Financial years 2019-20 to 2047-48, earnings in 2019-20 earnings values

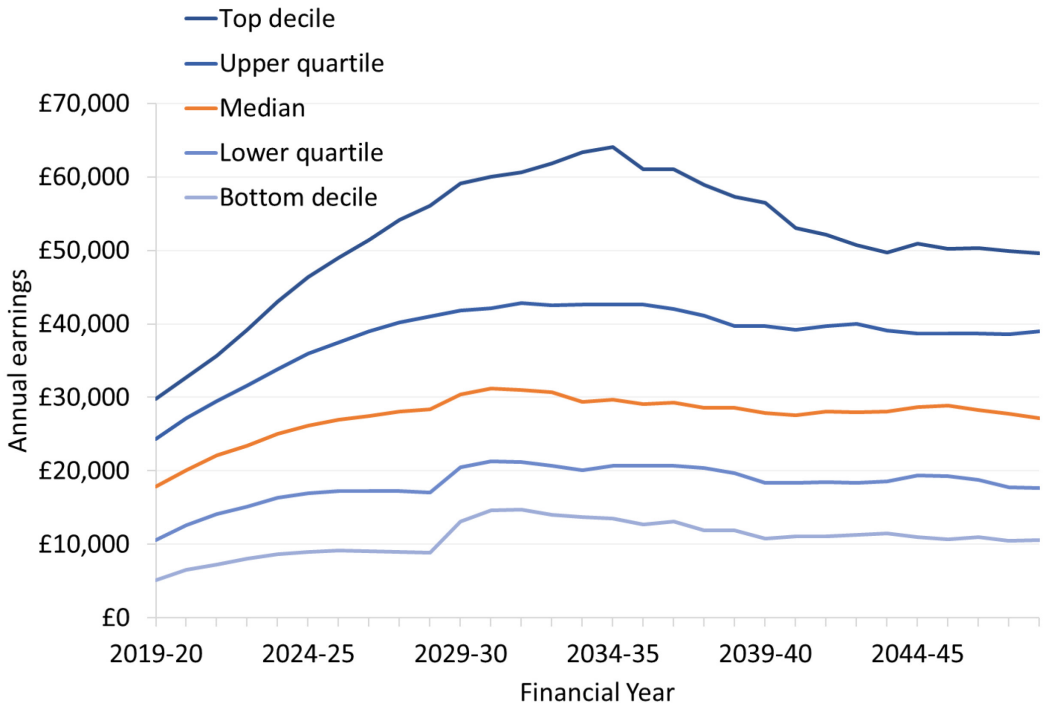
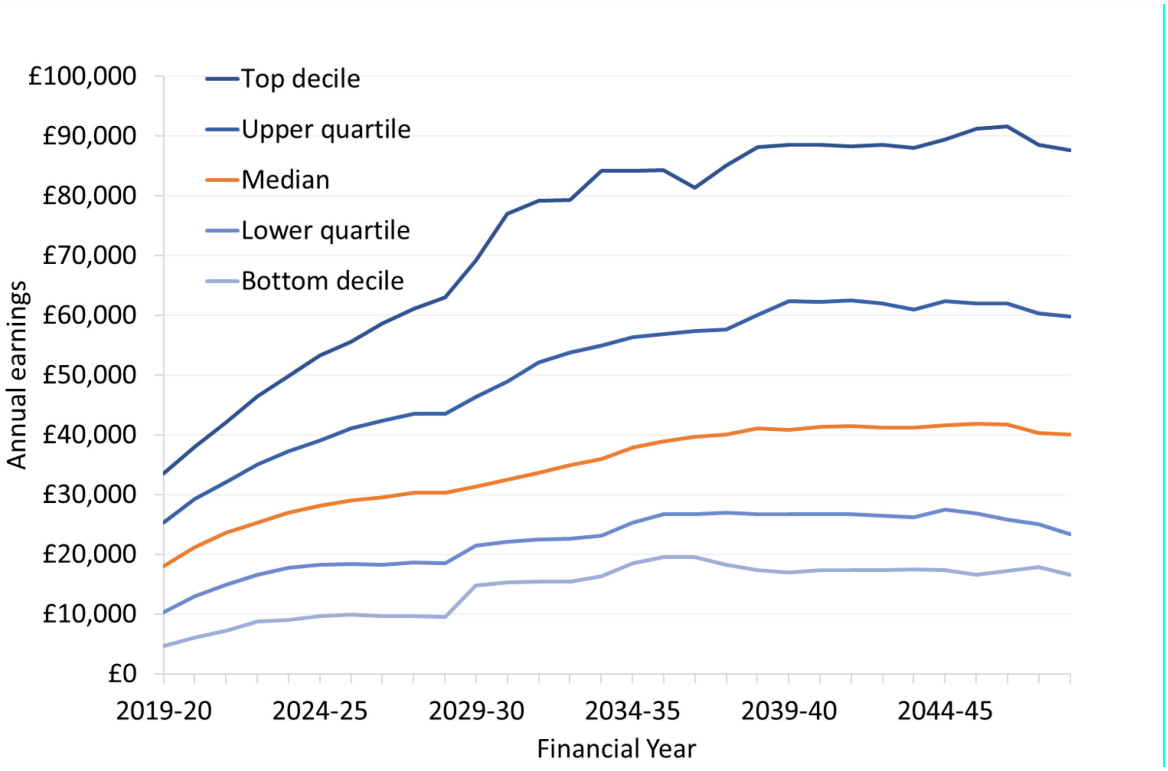


Figure 5.7: Forecast distribution of annual earnings of male borrowers in employment: 2019 SRDD cohort

Financial years 2019-20 to 2047-48, earnings in 2019-20 earnings values



Investment income

A borrower is expected to declare any investment income above £2,500 to HMRC, and this will be added to their overall earnings figure for the purpose of student loan repayments. The 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 logistic regression models generated from BHPS data). Based on these probabilities some borrowers are selected to have investment income each year.

Then an investment income amount is generated using a linear regression model based on the same set of characteristics, which is added to their labour income. If the model determines that the borrower has earned an income from their investments but it amounts to less than £2,500, then this is treated as zero investment income as the borrower would not be expected to declare this to HMRC.

Mortality

ICR loans can be cancelled prior to the end of the repayment term if the borrower dies. The probability of death in a given year is derived, based on the borrower's age and gender, from ONS life tables and SLC data. SLC data shows lower write-off levels than would be expected from ONS mortality rates, most likely reflecting that graduates have lower mortality rates than non-graduates of the same age. But the historic SLC data has little coverage of student loan borrowers aged above their mid-30s, so a weighted average of the two sets of mortality rates is taken that gives a high weighting to the SLC data at younger ages and to ONS data at older ages.

There are reasons other than death why a loan may be written off before the end of a borrower's repayment term, however, the level of these write-offs is comparatively small and as such they are not included in the model.

Repayments made directly to SLC

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:

- Voluntary repayments (prepayments) – These are (early) repayments made by the individual in addition to their obligatory repayments.
- Payments from overseas – Repayments from borrowers situated overseas cannot be collected through the tax system. Overseas borrowers make obligatory repayments direct to the SLC based on their income and the earnings threshold for their country of residence.
- 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 a borrower making a voluntary repayment each year is generated from a logistic regression model based on SLC administrative data. Voluntary repayments are particularly dependent on the magnitude of the debt outstanding and the number of years into the repayment period, as well as whether a borrower has previously made a voluntary repayment. The regression model also takes into account the age, gender, domicile, course level, subject group and provider group of the borrower. Most voluntary repayments come from borrowers with low amounts of debt in the first few years of repayment. If a borrower is due to make a voluntary repayment in the model, a percentage of the debt outstanding is paid as a direct repayment. This percentage is derived at random from a distribution based on the size of the borrower's remaining loan and the number of years since their SRDD.

Borrowers with an employment status of "migrated" can make overseas repayments. The probability of a migrated borrower making an overseas repayment each year is generated from a logistic regression models based on SLC administrative data. The regression model considers the borrower's age, gender, domicile, course level, subject group, provider group and how many years into their repayment period they are. The size of the repayment is selected at random from a distribution of repayment amounts based on the number of years since the borrower's SRDD.

In the model, borrowers are given a probability of making direct debit repayments rather than obligatory repayments if they repaid more than half of their remaining loan balance in the previous year. If selected to make direct debit repayments a borrower will repay half of their remaining loan balance in the first year of making them and the remainder of their balance in the second year.

Postgraduate loan borrowers

As they are new loan products (introduced in 2016), no historic data on postgraduate loan borrowers is available. In addition, information on postgraduate earnings and behaviours from survey data is limited, as in population surveys the proportion of the survey respondents that have postgraduate degrees is very small. Therefore, the student loan repayment model generates employment and earnings forecasts for postgraduate loan borrowers using the same earnings model as for first degree students with the same characteristics, to which it then applies a fixed uplift to earnings in all years to account for the higher earnings postgraduates are expected to have.

For master's borrowers an earnings uplift of 8.9% is applied for male borrowers and 10.3% for female borrowers. This is based on research that estimated this to be the average marginal earnings gain for master's students on top of their undergraduate degree (Conlon & Patrignani, 2011). For doctoral students, an earnings uplift of 8.0% is applied for male borrowers and 6.0% for female borrowers. These uplifts are based on results for doctoral students from the same research, though they have been adjusted down to account for trends associated with subject of study and HEI group. This adjustment aims to account for factors such as this, which were not considered directly in the research due to the available sample size but are important in student finance forecasting. For example, the population of students expected to take up Doctoral loans is not representative of all doctoral students since the availability of other funding sources (such as industry or research council funding) may differ by subject of study. These uplifts for master's and doctoral students are not directly comparable, as the other factors will also affect the average earnings for each course level.

As there is no administrative information available for them as yet, postgraduate loan borrowers are assumed to have similar behaviours as an equivalent first degree student for factors such as such as voluntary repayments, overseas repayments, mortality and investment income.

Loan borrowers on part-time courses

We do not forecast earnings for borrowers whilst they are undertaking a course, this includes part-time loan borrowers studying for longer than 4 years who have an SRDD in the fifth April after the start of their course, even if they are still studying. Similarly, we do not account for the impact of earnings prior to taking up a university course on earnings on completion. As such, estimated earnings for part-time loan borrowers may be lower than actuals on entry into the labour market. However, we do not expect this impact to be long lasting and modelled earnings for part-time loan borrowers will tend towards those of full-time loan borrowers as they move further through their career.

Repayment amounts and debt outstanding

Once annual earnings (including any investment income) are calculated and non-employment, migration, frictions, and mortality are taken into account, the obligatory repayments are calculated according to the deterministic repayment rules for that year. All obligatory, voluntary, overseas, or direct debit repayments that the borrower makes each year are summed together, up to a maximum of the borrower's remaining loan balance. Borrowers are assumed to stop repaying their loan once their loan balance reaches zero. The model does not account for borrowers making overpayments or receiving refunds after overpaying.

To calculate the size of a borrower's loan balance, borrowers are given annual outlay amounts while on their course based on the distribution of outlay amounts of historical borrowers in the SLC data, uprated in line with forecast RPIX depending on the appropriate loan policy. Capitalised interest is accumulated each year and added to the size of the borrower's debt, while any repayments are subtracted from it. The size of the borrower's debt is calculated on this basis each year until they either fully repay their loan or until their loan is cancelled, either due to mortality or because they reached the end of their repayment term.

In reality annualised repayments through HMRC are averaged out into monthly instalments by SLC, but as an approximation all repayments in the model (obligatory and direct) are assumed to be made in the middle of the financial year. Interest for the first half of the year is added to the debt outstanding at the start of the year before repayments are made, then the interest for the rest of the year is added after they have been deducted. In years where a borrower is forecast to receive loan outlay, these are assumed to occur in three instalments at the end of September, January and April. Interest is accrued on these payments and applied to the loan balance accordingly.

If a borrower's loan is cancelled this is assumed to happen at the end of the financial year, as this is the point when cancellations will occur at the end of a borrower's repayment term, which are expected to account for the large majority of cancellations.

Interest rates each year are calculated from RPI, the Bank of England base rate (Plan 1 only) and borrowers' income in line with the appropriate policy. RPI and Bank of England base rate figures are based on OBR forecasts. The interest rates for each part of the year are calculated and then combined into an annual average that is used across the financial year. For all three loan plans the RPI figure used in calculating interest rates changes each September to the March RPI figure published by ONS in the same year, but as OBR only publishes quarterly forecasts (and in the long run only annual forecasts) the model uses the forecast for the equivalent January to March quarter in the short run, and the annual figure for the same financial year in the long run. The Plan 1 interest rate can potentially vary each month between RPI and the base rate +1%, so to simulate this the OBR's quarterly forecasts for the base rate are used and compared to the RPI figure each quarter (twice in the Jul-Sep quarter when the RPI figure that is used can change).

Population totals

Forecasts for individual loan borrowers are aggregated together to estimate totals for the whole student loan borrower population. Rather than making estimates for the whole population of loan borrowers, to

make the model more efficient (i.e. quicker to run), forecasts are only made for a sample of loan borrowers, with weightings applied to these borrowers' results to "scale up" to totals for the whole population.

A random sample of 200,000 borrowers is used for each loan product, covering entrants from the first year that the loan product was introduced up to entrants in academic year 2023/24 (excluding those that have already finished repaying their loans or had them cancelled). This amounts to a 7% sample for Plan 1 loan borrowers, a 4% sample for Plan 2 full-time borrowers, a 32% sample for Plan 2 part-time borrowers and a 46% sample for master's loan borrowers.

Early in-year SLC data has indicated fewer borrowers receiving doctoral loans than previously forecast. As the doctoral loan is a new product from 2018/19, outturn characteristic data is not yet available and so borrowers are based upon historical HESA data. Until administrative data becomes available, the same sample size of doctoral borrowers is used, and the results are scaled down to match new forecasted entrant numbers.

The scaling used to increase the sample results to population totals is weighted based on several variables to reduce the sampling bias in the model. The variables used in the weighting are course start year, SRDD, domicile, gender, course level, subject group, the borrower's write-off rule and whether the loan has been sold (Plan 1 only).

The Resource Accounting and Budgeting (RAB) charge and the stock charge

The RAB and stock charges are the estimated cost to Government of providing a subsidy for the student finance system. They are the proportion of loan outlay (the RAB charge) and of the total outstanding loan balances (the stock charge) that are expected to not be repaid when future repayments are valued in present terms.

To calculate the RAB charge, the total outlay in a given year is added up and compared to the total net present value (NPV) of the repayments that are anticipated in connection with this same outlay. The RAB charge is calculated as

$$\text{RAB charge} = \left(1 - \frac{\text{NPV of repayments in respect of outlay}}{\text{value of outlay}}\right) \times 100\%$$

Similarly, the stock charge is calculated by summing all outstanding loan balances at the start of the year and comparing this to the total net present value (NPV) of the repayments that are anticipated in connection with these loans. The stock charge is calculated as

$$\text{Stock charge} = \left(1 - \frac{\text{NPV of repayments in respect of outstanding loan balances}}{\text{face value of outstanding loan balances}}\right) \times 100\%$$

The NPV of future repayments is calculated by discounting all future repayments at a rate of RPI+0.7% per year to the same point in time as the loan outlay or loan balance. This is the discount rate for financial instruments set by HM Treasury (HMT)[31] and is intended to reflect of the cost of Government borrowing.

Student loans are valued in DfE's annual accounts in line with the International Financial Reporting Standard (IFRS) 9, under which where future cash flows are discounted to measure the fair value of a financial asset, this should be done using the higher of the rate intrinsic to the financial instrument or the Her Majesty's Treasury (HMT) discount rate. For student loans the intrinsic rate would be the discount rate that gave a RAB or stock charge of 0%, so the HMT discount rate is used provided the RAB charge is greater than 0%. Should the HMT discount rate result in a RAB charge calculation giving a negative value then the intrinsic rate is used instead, meaning that that RAB charge will take a value of 0%.

In the model, RAB charges are calculated for the loan book as a whole by first calculating the NPV of individual borrowers' repayments, then for each year aggregating these together across all borrowers and

[31] Discount rates for post-employment benefits, general provisions and financial instruments: Announcement of rates - HM Treasury, December 2015.

comparing them to their total loan outlay in that year. Stock charges are calculated in the same way, aggregating the NPV of individual's repayments before aggregating them to a population total and comparing this to the face value of the loans at that point in time. Where a borrower has more than one year of outlay or has both future loan outlay and an existing loan balance that will be included in the stock charge, future repayments are allocated between each year of their loan outlay and their existing loan balance in proportion to the relative balances of each loan when valued at the same point in time (i.e. talking into account interest accrued on the earlier loan balances).

A RAB charge is no longer produced for Plan 1 loans as very few Plan 1 students are still receiving loans. It is not possible to produce a reliable RAB charge as the small numbers mean there would be a high level of uncertainty around any forecasts, particularly as these may be an atypical group of students that will not follow the same future earnings distribution as the overall population.

Data quality

RAB and stock charge estimates require earnings and repayments forecasts covering the next 30-40 years. These forecasts depend heavily on input data sources, modelling techniques and assumptions used in the model. Consequently, forecasts are inherently uncertain owing to the inevitable uncertainties associated with these sources, assumptions, and methods. For example, the model assumes that the distribution of future earnings paths will be similar to historic distributions derived from survey data and SLC administrative data, which will not necessarily be the case, particularly in the long term.

The model is dependent on the OBR macroeconomic forecasts that it uses to uprate earnings, calculate interest rates and repayment thresholds, and to discount future repayments to present values. Any significant changes to the economy from these forecasts could affect the repayments that will be made on student loans.

Tables 5.2 includes a demonstration of the sensitivity of the Plan 1 stock charge, and the Plan 2 full-time and part-time higher education RAB charge, to variations in the OBR forecasts. The variations shown are larger than the changes that would typically be seen in OBR's forecasts from one year to the next, although the OBR's Central Covid scenario, for example, sees a drop in the Bank of England base rate of almost 1pp in financial year 2019-20 relative to projections published in March (before the publication of the Covid scenarios). Table 5.2 shows that a Bank of England base rate that a 1pp lower Bank of England base rate from financial year 2019-20 onwards could increase the Plan 1 stock charge by up to 2pp. For Plan 2 RAB charges the impact of varying some key policy parameters is also demonstrated in Table 5.3.

Table 5.2: Sensitivity of Stock and RAB charge forecasts to variations of key economic inputs

Table shows the percentage point (pp) change to the forecast 2019-20 stock (Plan 1 loans) and RAB (Plan 2 loans) charges as a consequence of varying each listed macroeconomic input up or down by 1pp. All changes are assumed to begin from financial year 2019-20.

Macroeconomic change		Plan 1 stock charge	Plan 2 full-time RAB charge	Plan 2 part-time RAB charge
RPI	**-1pp**	-6pp	-5pp	-3pp
	+1pp	+6pp	+5pp	+3pp
Earnings growth	**-1pp**	+4pp	+5pp	+3pp
	+1pp	-4pp	-5pp	-3pp
Bank of England base rate	**-1pp**	+2pp	N/A	N/A
	+1pp	-2pp	N/A	N/A

Table 5.3: Sensitivity of RAB charge forecasts to variations of key policy inputs

Table shows the percentage point (pp) change to the forecast 2019-20 RAB charges as a consequence of varying each listed policy input up or down by 1pp, or in the case of the repayment threshold by £1,000. Changes of the repayment threshold and repayment rate are assumed to begin from 2020-21, since no borrowers affecting the 2019-20 RAB charge will enter into repayment until at least April 2020. The change in interest rate is assumed to begin from financial year 2019-20.

Policy change		Plan 2 full-time RAB charge	Plan 2 part-time RAB charge
Interest rate	**-1pp**	+3pp	+4pp
	+1pp	-3pp	-4pp
Repayment rate	**-1pp**	+3pp	+1pp
	+1pp	-2pp	-1pp
Repayment threshold	**-£1,000**	-2pp	-2pp
	+£1,000	+2pp	+2pp

The model uses SLC administrative data to determine borrower characteristics, loan amounts, earnings in the first three years of their repayment term and repayments made directly to SLC. Being from an administrative source the historical SLC data should be broadly accurate, although the earnings and direct repayment forecasts rely on future borrowers having similar behaviours to historic borrowers.

Where new loan products are being introduced, the forecasts are more uncertain as there is less historical information available on which to base forecasts and more uncertainty about what student behaviours will be in response to the policy. This is particularly the case for the two postgraduate loan products, for which the earnings forecasts are less well developed than for undergraduates and for which there is no historical information about loan borrowers' characteristics and behaviours.

Earnings distributions

The model currently forecasts earnings for borrowers from financial year 2019-20 onwards, as earnings in 2018-19 and earlier are available in the SLC administrative data. It can be difficult to compare earnings forecast by the model to survey or administrative data due to lags in data reporting. Instead, we compare trends in forecast earnings paths to those seen in survey data. We do this by aggregating LFS data from 2006 to 2018 and comparing the distribution of earnings for borrowers at various ages to the distribution of earnings produced in the earnings model.

As the LFS only records graduates, rather than student loan borrowers, the raw data is not directly comparable to the borrower population. We would expect borrowers to have slightly lower earnings on average than graduates, as some graduates will not take out a loan and some borrowers will not graduate. We therefore apply a heuristic to the LFS data, through reweighting, to make the LFS graduate population more comparable to the borrower population.

After the heuristic has been applied, we can compare the distribution of earnings for male and female Plan 2 borrowers, as in in Figures 5.8 and 5.9. In both male and female distributions, the model estimates agree with the trend in the LFS data. Data on graduates' earnings nearer retirement tend to be quite volatile, as the number of individuals in employment decreases; this inhibits forecasting, however most borrowers will have either repaid their loan by this point or passed the end of their loan term.

Figure 5.8: The distribution of earnings of male Plan 2 borrowers by age over the working lifetime, estimates compared to distribution observed in LFS data.

Annual earnings, 2019-20 earnings values

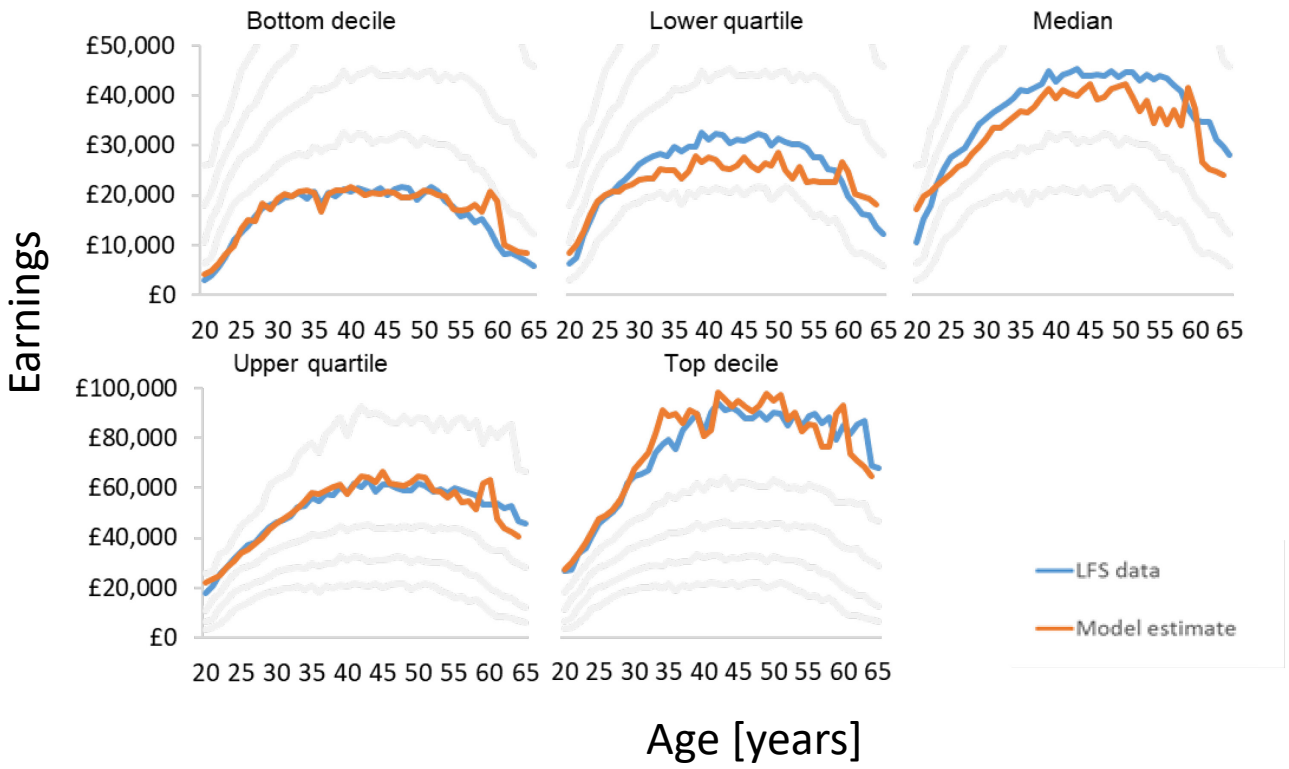
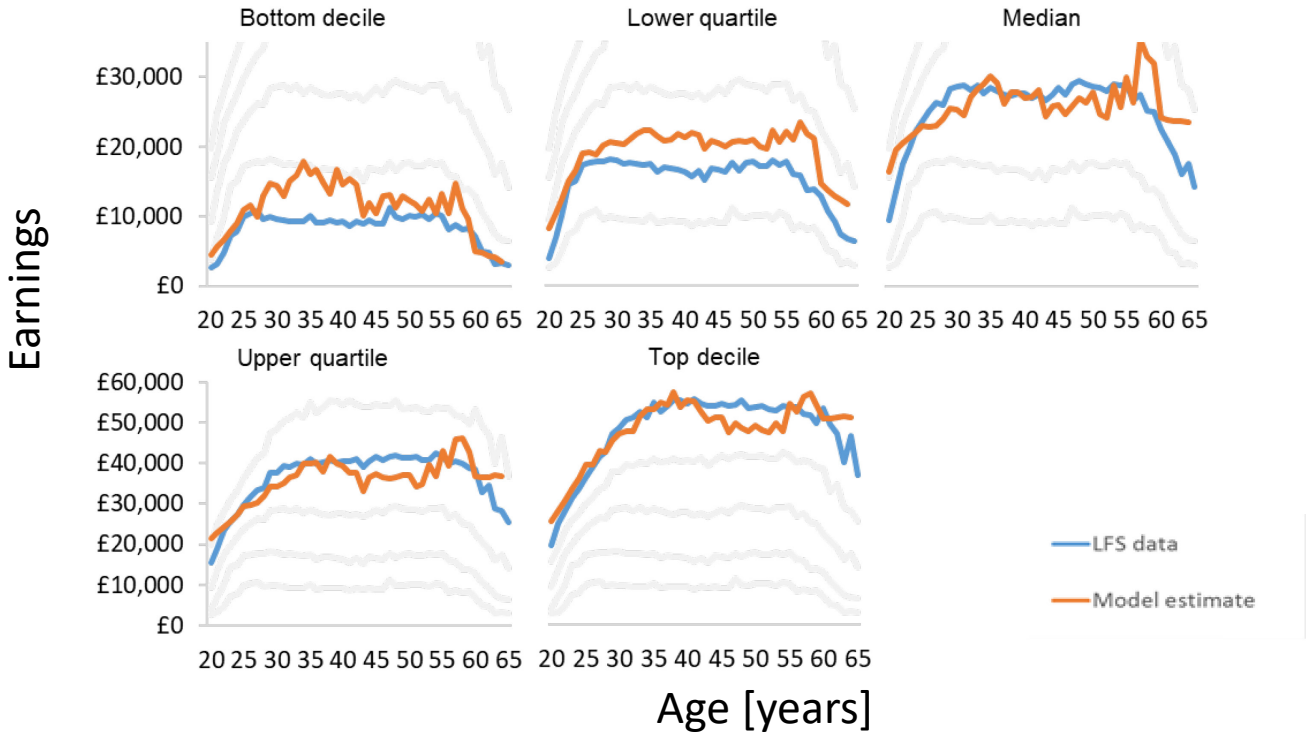


Figure 5.9: The distribution of earnings of female Plan 2 borrowers by age over the working lifetime, estimates compared to distribution observed in LFS data.



Annual earnings, 2019-20 earnings values

Comparison of forecast repayments with actuals

The student loan repayment model is designed to forecast repayments across loan borrowers' repayment terms. Comparisons between forecast repayment totals (one or two years ahead) for individual years and the actual outturn data can give an indication of how well the model is performing. Table 5.2 shows how recent outturn figures compare with the forecasts made at the time. Improvements are made to the student loan earnings and repayment model each year and the data used in it are updated, so forecasts are shown as made at both the start and end of each tax year.

Table 5.2: Forecast and outturn repayments across all higher education loan products

	Outturn		Forecast		
Tax Year	Date	£ million	Date	£ million	Difference
2013-14	30/04/2015	1,590	31/03/2014	1,630	2.5%
2014-15	30/04/2016	1,750	31/03/2014	1,870	6.9%
			31/03/2015	1,920	9.7%
2015-16	30/04/2017	1,930	31/03/2015	2,140	10.9%
			31/03/2016	1,930	0.0%
2016-17	30/04/2018	2,220	31/03/2016	2,320	4.0%
			31/03/2017	2,250	1.4%
2017-18	30/04/2019	2,340	31/03/2017	2,570	9.0%
			31/03/2018	2,470	5.3%
2018-19	30/04/2020	2,530	31/03/2018	2,700	6.3%
			31/03/2019	2,600	2.7%
2019-20	30/04/2021	:	31/03/2019	2,990	:
			31/08/2020	2,690	:

These figures include repayments made directly to SLC and PAYE and Self Assessment repayments made via HMRC. Direct repayments are recorded against the year they are received by SLC, HMRC repayments are recorded against the year of the earnings they relate to.

Repayments across all loan products are included in the data. Up to 2015-16 only Plan 1 borrowers were eligible to make obligatory repayments, though Plan 2 borrowers could make voluntary repayments. From 2016-17 the first Plan 2 borrowers became liable to make obligatory repayments, and the first Plan 3 borrowers will in 2019-20.

From financial year 2019-20 the frequency in which repayments data was provided to SLC by HMRC increased from annually (within one year of the financial year ending) to weekly. Repayments are reported within the financial year they are posted to customers' accounts, and so prior to this more frequent data sharing repayments posted to a given financial year were mostly from the year before. As a result of this change, which effectively provides a more up-to-date representation of student loan balances at a given point in time, total repayments reported by SLC for financial year 2019-20 include almost two years' worth of PAYE repayments. The impact of this transition will normalise in financial year 2020-21, where a single years' worth of PAYE repayments will be reported under the updated data sharing conditions.

Variances between forecasts and actuals will be due to a range of factors, including:

- Macroeconomic shifts and new data
- Modelling variances and random variation
- Operational factors that result in lower than expected collections.

By the time of the second forecast shown for each year the macroeconomic data for the year will largely be known, so the forecast is less dependent on OBR macroeconomic forecasts. However, modelling changes and other data updates will also have occurred so changes between the two forecasts also include other factors.

6. Advanced Learner Loans model

Introduction

Advanced Learner Loans (ALLs) are tuition fee loans to help those aged 19+ at the start of their course meet the up-front costs of regulated Further Education (FE) qualifications at Levels 3 (equivalent to 2 A levels) to Level 6 (equivalent to an undergraduate degree) in England. ALLs were introduced in 2013/14 to those aged 24+ and at levels 3-4 following a refocusing of the Adult Education Budget on adults requiring skills and learning to equip them for work, an apprenticeship or further learning.

Following a public consultation in 2014, the extension of ALLs in academic year 2016/17 to those aged 19-23 and to Levels 5-6 has been the programme's most significant change.

The Resource Accounting and Budgeting (RAB) charge for ALLs is the estimated cost to Government of borrowing to support the ALLs system. The purpose of the DfE ALLs model is to assist in valuing the existing ALLs loan book and to provide forecasts for budgeting purposes.

The RAB charge is an estimate and it is heavily dependent on assumptions around the future income of ALLs borrowers. At present, we have imperfect information about their repayment ability as ALLs are a new product and no significant number of ALLs borrowers have yet made a repayment. The first provisional repayment data was made available in summer 2018. The methodology, data sources and assumptions are presented in the section below.

Methodology

The ALLs model is a micro-simulation model. The model creates thousands of simulated borrowers with a variety of characteristics. Each borrower is assigned a debt and their earnings are projected for each of the next 30 years. Then the repayment rules are applied to each borrower to estimate their repayments, individual loan balance and interest for each of the next 30 years. The assumptions used in the simulation model fall into five main sections below:

1. Borrowers' characteristics and their loan details
2. Macroeconomic assumptions: Average Earnings Index (AEI) and Retail Price Index (RPI)
3. Loans policy assumptions
4. Annual income post learning: employment status, income and income distributions

4.1. Labour market status

4.2. Position on income distribution

4.3. Annual income in nominal terms

5. Life events

5.1. Mortality

5.2. Migration

5.3. Permanently unfit for work

5.4. Extending working lives.

1. Borrowers' characteristics and their loan details

Analysis of administrative Student Loans Company (SLC) data informs the input assumptions on the characteristics of borrowers in each academic year, the courses they study and the average loan size they take. The complete list of input parameters are:

- Total number of new borrowers;

- Proportion of borrowers on multiple courses;
- Course type (A levels, Access to HE, Level 3 Diploma, Level 4 Diploma, Level 3 Certificate, Level 4 Certificate and Level 5/Level 6 courses);
- Gender split of borrowers;
- Age distribution: 19-71 by single year of age;
- Course duration in months: 1-24 months;
- Course start month across academic year;
- Average loan size by type of course (A levels, Access to HE, Level 3 Diploma, Level 4 Diploma, Level 3 Certificate, Level 4 Certificate and Level 5/Level 6 courses).
- Non-completion rates by type of course.

2. Macroeconomic assumptions

The model uses the Office for Budget Responsibility (OBR) assumptions on future average earnings growth (AEI) and RPI projections to calculate future repayment thresholds, interest rates and discount rates for the ALLs. The model also uses the AEI projections to uprate borrowers' future earnings. Course costs are uprated by Consumer Price Inflation (CPI).

3. Loans policy assumptions

Loans are currently available to learners aged 19+ who are studying Level 3 and above qualifications. The loans are repaid at a rate of 9% of pre-tax earnings above the lower repayment threshold, as follows:

- Lower repayment threshold: £25,725 for tax year 2019-20, then adjusted annually by the AEI (Q1) thereafter.
- Higher repayment threshold: £46,305 in 2019-20, then adjusted annually by AEI (Q1) thereafter.
- Interest rate:
    - RPI + 3% up to Statutory Repayment Due Date (SRDD); then
    - RPI for borrowers with income below lower repayment threshold;
    - RPI + 3% for borrowers with income above upper repayment threshold;
    - RPI + (3% x (income – lower threshold)/(upper threshold – lower threshold))

      for those earning between repayment thresholds.

Borrowers who study Access to HE courses and complete a higher education course have any outstanding ALLs balance written off.

The impact of Regulation 25 (cancellation of fee loan) in the Further Education Loans Regulations 2012 (as amended) is not currently incorporated in the model published. The impact is estimated to be very small. It will be incorporated in the next iteration.

The presented RAB includes the impact of updating the macroeconomic determinants to OBR's Covid-19 scenario. Beyond this, there have been no adjustments to the forecast to reflect any other Covid-19 related

impacts. Potential areas of investigation for the next iteration of the model might include evaluating changes to learner volumes, learner mix (age and gender), course mix, and employment rates.

4. Annual income post learning – employment status, income and income distributions

The key assumption in the model is the future annual income of borrowers after finishing their course.

ALLs are income contingent, i.e. borrowers repay the loan only if their annual income is above the lower repayment threshold, and their repayment amount is based on their income. Annual incomes are the basis for calculating loan repayments and interest; and getting good income estimates over the working life of borrowers is critical to estimating the RAB charge.

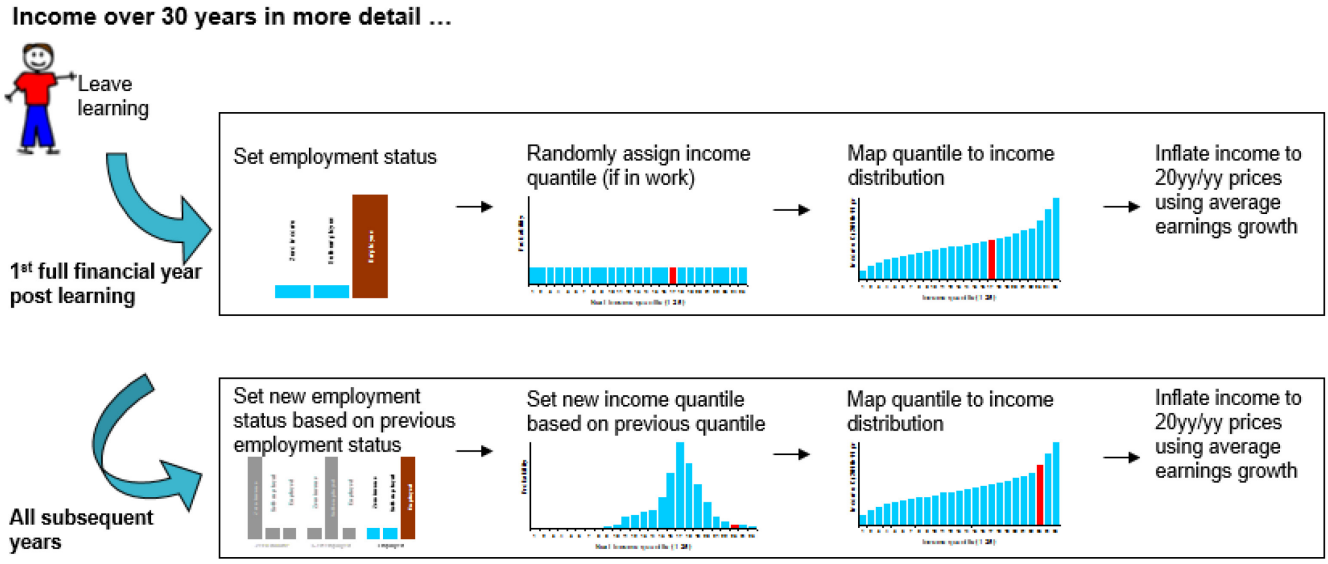
Income modelling starts for a borrower in the first financial year after completing their course. In most cases this will be the first year they are due to make repayments, except during the first years of the policy where it was legislated that the first repayments would not be due until 2016-17.

In every simulation each modelled borrower will go through processes to:

- Set labour market status – transition between 'employed', 'self-employed' and 'no income';
- Assign position on income distribution – one of the 25 quantiles[32] and
- Assign annual income based on their position in the income distributions.

The process is summarised below in figure 6.1:

Figure 6.1: Income modelling for ALL borrowers



4.1 Labour market status

Labour market status in a given year is modelled using a discrete probability matrix which supports three possible outcomes: 'no income', 'employed' and 'self-employed'. The matrix has different probabilities depending on the characteristics of the borrower:

- Course level: Level 4+, Level 3
- Gender: male or female
- Age: single year of age

[32] The income distribution is split into 25 equal quantiles, so each quantile represents 4% of the income distribution. Quantile 1 = the lowest earning. The number of quantiles was determined by considering the greatest detail that the LFS could provide whilst still providing robust estimates.

- Years since completed learning: one year, or more than one year
- Current labour market status: no income, employed or self-employed
- Current year income: one of 25 quantiles, or zero if not in employment

The data source for the analysis is 68 quarters of Labour Force Survey (LFS). The model uses two sets of labour market assumptions:

- Initial labour market status for those in the first full financial year after their SRDD.
- Probabilities of changing labour market status in all subsequent years.

In the first year after SRDD the model has no knowledge about the working history of each borrower so borrowers are assigned a labour market status only in accordance with the overall employment rates for that age, gender and course level[33]. In all subsequent years, each borrower has a probability of changing labour market status, this time taking into account their current labour market status and income quantile (if in employment).

Income data for the self-employed is not available from the LFS. This means that income based labour market changes from the employee population are used as a proxy, i.e. if an employed borrower at the bottom of the income distribution has a greater chance of becoming unemployed compared to someone at the top, then a self-employed borrower at the bottom of the income distribution also has a greater chance of becoming unemployed compared to someone at the top.

4.2 Position on income distribution

After labour market status is established in the previous stage, the model assigns the borrower to a quantile on the income distribution. As with labour market status, the model assigns future income quantiles, using different probabilities depending on the characteristics of the borrower:

- Course Level: Level 4+, Level 3
- Gender: male or female
- Age: single year of age
- If in work: current income: one of 25 quantiles
- If out of work: previous income when last in work: one of 25 quantiles or zero if never previously modelled in work.

The LFS data do not contain income information for the self-employed, and so it is assumed that transitions across the income distribution for those who are self-employed occur with the same likelihood as for those who are employed.

The LFS data only have sufficient information to compare year on year income, they do not hold robust income information from previous employment spells further back in time. As a simplification it is assumed that a borrower returns to work at the same position on the income distribution as they were in their most recent modelled income spell, e.g. if they left work in quantile 10 then when they return to work they will return in quantile 10.

The LFS data suggest that typically those entering work are on lower incomes than the equivalent group leaving work. This would suggest a natural decay in position on the income distribution between jobs, i.e. that learners should return at a lower position on the income distribution. In practice this does not produce a

[33] Ideally, the data would be further restricted to learners that are one year post learning but there is insufficient data in the LFS to support this. The initial employment status is quickly eroded by the subsequent employment transition probabilities so has a fairly limited impact on the forecast.

good modelling solution as it creates a heavy penalty for being out of work in a single year and produces unexpected lifetime income paths inconsistent with the original simplified assumption.

4.3 Annual income in nominal terms

The final step in modelling income is to convert an income quantile into actual income for a given financial year. This is achieved by matching the simulated income quantile to historic income quantile distributions. There are different distributions by:

- Course Level: Level 4+, Level 3
- Gender: male or female
- Age: single year of age
- Years since left learning bands: 1-3 years; 4-10 years; 11+ years

Income distributions in the first 3 years post training are constructed using actual ALL learner earnings from Longitudinal Education Outcomes (LEO) data. For the years 4-10 post training, we also use LEO data, but use a proxy group of learners who started before the introduction of ALLs that have similar characteristics to those that now on ALL courses. Beyond 10 years post training, we use 14 years of the Annual Population Survey (APS). Income is inflated for future growth using OBR projections of average earnings growth for the entire working population. It is also assumed that average earnings growth is even across all quantiles.

The use of LEO data is a change to the methodology relative to the 2018-19 publication. Previously only APS data was used to inform the model income forecasts, as this was the best source available at the time. It was not possible to isolate ALL-funded students in the APS data. That led to overstating the incomes of the students. Changing to LEO data has lowered the income forecasts for the ALL-funded students, meaning fewer students will reach the repayment threshold. This has led to an increased RAB and an increased transfer proportion.

5. Life events

From the time a borrower takes out a loan there are additional events that need to be factored into the modelling, such as mortality, migration, and repayments made directly to SLC. All are modelled at the individual level, but to varying levels of detail.

5.1 Mortality

The Office for National Statistics (ONS) England & Wales population projections include assumptions on mortality rates by gender and single year of age up to 2084. These are used directly to estimate the chance of death for a borrower in any year post taking out a loan, with the key assumption that ALLs borrowers have the same mortality rates as the rest of the population.

It is likely there is some difference in mortality rates for this group compared to the whole of England & Wales, particularly as there is a known link between increased wealth and improved mortality. However, any difference will be minimal and the mortality rates for the working age population are generally very low, further reducing any possible impact.

5.2 Migration

The migration assumptions are split into two components

- Probability of migrating out of the country (out-migrant rate)
- Length of spell out of the country: 3 years, 7 years or 20 years

Out-migrant rates by gender and age are derived by combining Office for National Statistics (ONS) out-migrant estimates with population projections. These are used directly to provide each borrower with a probability of becoming an out-migrant each year.

Once an out-migrant is identified the model assigns a length of stay for the duration abroad. The length and probability of duration are estimated from International Passenger Survey (IPS) data between 1978 and 2012, over which time actual length of stay figures have been relatively stable.

5.3 Permanently unfit for work

The model uses expected flows into the Support Group of Employment Support Allowance (ESA) as a proxy for the number of people expected to become unfit for work each year. Once identified as unfit for work, a simulated borrower has their remaining loan amount written off.

The ESA Support Group is not a perfect measure of the unfit for work rate, as it is a State Benefit that needs to be actively claimed so is not received by everyone that is eligible. It does, however, provide the closest match in definition to unfit for work which is sufficient for the relatively low number of people that will be affected.

5.4 Extending working lives

State Pension Age (SPA) is due to rise to 67 for men and women by 2024. For women this represents an additional 7 years of work before they will receive their State retirement benefits compared to the observed population in the model data sources.

The model includes a timetable to assign each borrower a State Pension age, from which there is an assumption around what employment rates and income to use given to reflect that people are likely to work for longer. For example, a man with SPA of 65 has the employment rate of a 66-year-old when aged 66, but a man with SPA of 66 has the employment rate of a 65-year-old when aged 66. This assumes that for men, the employment rate and income of a 65 year old is held constant between the ages of 65 and the new State Pension age; for women the employment rate and income of a 60 year old is held constant between the ages of 60 and the new State Pension age.

The effects of retirement can be seen in LFS data from as early as age 50, however, given the high uncertainty about how people will respond to the rises in State Pension age, and future employment in general, the model uses the simplified assumption.

Data quality

There is a high level of uncertainty around the ALLs RAB and stock charges estimates. At present, we have imperfect information about ALL borrowers' repayment ability as ALLs are a new product. The ALL repayment data became available in 2018. As we receive more repayment data, we will continue to compare this with the assumptions within the model and the resulting impact on the forecast.

Both administrative and survey data are used to inform the assumptions underpinning the ALLs model. SLC administrative data are used to determine borrowers' characteristics and loan amounts. Historic SLC data should be broadly accurate.

The Labour Force Survey (LFS) is used to determine employment and income state movements of ALLs borrowers and the Annual Population Survey (APS) and Longitudinal Education Outcomes (LEO) are used for the income distributions. The LFS is a survey of the employment circumstances of the UK population. It is the largest household survey in the UK and provides the official measures of employment and unemployment. The APS is a supplement to the LFS data. The APS is published quarterly, and each dataset contains 12 months of data. The sample size for each dataset is approximately 170,000 households and 360,000 individuals. LEO brings together information from the Department for Education

with employment, benefits and earnings information from the Department for Work and Pensions and Her Majesty's Revenue and Customs.[34]

The employment and income status assumptions are two of the most important assumptions within the model. We derive both of those from the LFS assuming that Further Education achievers in the LFS data are representative of ALLs borrowers.

The LFS data has various limitations and a key drawback is that only one year of earnings history is available. Also, the LFS does not capture self-assessment income, but does include a flag to identify employment status. The key assumptions arising from the data limitations are listed below:

- The model 'forgets' a borrower's past history of employment statuses prior to the current year, since a one-step transition approach is used. This means that a learner who has been unemployed for 10 years has equal chance of employment as someone who has been unemployed for just one year.
- Transitions between employment states for the employed are used as a proxy for self-employed people. Self-employed people are assumed to have the same income distribution as employed learners.
- We assume that a borrower who has left the labour market will return to work at the same position on the income distribution as they were in their most recent employment spell.

Despite these limitations, the LFS is the only data source currently available that has a sufficiently large sample to allow analysis of income by age and type of qualifications and transitions in and out of employment. We now have some actual repayment data from the Longitudinal Education Outcomes, but only for a limited number of years. We will seek to use this data alongside the LFS data to improve these assumptions.

The model is also dependent on the OBR macroeconomic forecasts that it uses to uprate earnings, calculate interest rates and repayment thresholds, and to discount future repayments to present values. Any significant changes to the economy from these forecasts could affect the repayments that will be made on the Advanced Learner Loans.

[34] More information on LEO and the data used can be found on: Longitudinal education outcomes study: how we use and share data

7. Uses of these statistics

These forecasts show how much Government outlay on student loans is expected to be in future, how much is expected to be repaid and how the student loan book may grow in the future. The Department for Education uses these models for financial planning and in the development of student funding policies.

These forecasts are also used in the Department for Education's annual accounts in the valuation of the student loan book. The stock charge, RAB charge and transfer proportion are used to impair the face value of the loan books and the value of new loans being issued respectively, to reflect that the value of the future repayments that will be received in relation to these loans is less than the long term cost of Government borrowing that would be necessary to cover its outlay on student loans.re

These models are used by the Office for Budget Responsibility as part of its estimates of public sector borrowing, including in its Economic and Fiscal Outlook that presents economic forecasts five years into the future and its Fiscal Sustainability Report that presents long term projections of UK public finances.

8. References

Conlon, G., & Patrignani, P. (2011, June). *BIS research paper number 45: The returns to higher education qualifications*. Retrieved from https://www.gov.uk/government/publications/higher-education-qualifications-returns-and-benefits

Nanton, A., & Rowlings, J. (2017, September 25). *Earnings and low pay: distributions and estimates from the Labour Force Survey*. Retrieved from ONS: www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/articles/earningsandlowpay/distributionestimatesfromthelabourforcesurvey

9. Get in touch

Media enquiries

Press Office News Desk, Department for Education, Sanctuary Buildings, Great Smith Street, London SW1P 3BT.

Tel: 020 7783 8300

Other enquiries/feedback

HE Modelling Team Higher Education Analysis, Department for Education, Sanctuary Buildings, Great Smith Street, London, SW1P 3BT.

Email: he.modelling@education.gov.uk

Department for Education

About this publication:

Enquiries: HE Modelling Team, Higher Education Analysis, Department for Education, Sanctuary Buildings, Great Smith Street, London, SW1P 3BT.

Email: he.modelling@education.gov.uk

https://www.gov.uk/government/collections/statistics-student-loan-forecasts

Reference: Student loan forecasts, England: 2019-20: Quality and methodology information