



Institute of Actuaries of Australia

**The Higher Education Loan Programme  
(HELP/HECS)  
– Microsimulation Modelling of  
Individual Repayment Prospects**

*Prepared by Michael O'Neill & Susan Antcliff*

Presented to the Institute of Actuaries of Australia

2009 Biennial Convention, 19-22 April 2009

Sydney

*This paper has been prepared for the Institute of Actuaries of Australia's (Institute) 2009 Biennial Convention*

*The Institute Council wishes it to be understood that opinions put forward herein are not necessarily those of the Institute and the Council is not responsible for those opinions.*

© Michael O'Neill & Susan Antcliff

The Institute of Actuaries of Australia

Level 7 Challis House 4 Martin Place

Sydney NSW Australia 2000

Telephone: +61 2 9233 3466 Facsimile: +61 2 9233 3446

Email: [actuaries@actuaries.asn.au](mailto:actuaries@actuaries.asn.au) Website: [www.actuaries.asn.au](http://www.actuaries.asn.au)

**The Higher Education Loan Programme (HELP/HECS)  
– Microsimulation Modelling of Individual Repayment Prospects**

**Abstract**

The parameters of the Higher Education Loan Programme (HELP) have changed considerably since the introduction of the Higher Education Contribution Scheme (HECS) in 1989. However, the fundamental principle of income-contingent repayment on which it is predicated has remained unchanged. A body of repayment experience data has accumulated from earlier cohorts, and it has become feasible to use microsimulation models of incomes to project future repayments. In our view, the income and associated repayment histories of the earlier cohorts provide a credible basis for assessing the repayment prospects of those with outstanding debt, including those for whom we have no income history.

This paper provides an introduction to the complexities involved in quantifying the cost burden of each cohort finishing university and the challenges in establishing an appropriate modelling framework.

*Key words: higher education funding, income contingent loans, microsimulation, parameter randomisation, Markov Chains, Generalised Linear Models, binomial and Gamma distributions*

**Acknowledgements:** The authors would like to acknowledge the considerable input of Bernie Rutten in the design and implementation of the microsimulation model discussed in this paper. We would also like to thank Tim Higgins for his helpful comments on earlier drafts of the paper. All errors and omissions are the responsibility of the authors.

**Disclaimer:** Permission to present this paper has been sought from both the Commonwealth Department of the Treasury and the Department of Employment, Education and Workplace Relations. The content of the paper does not, however, in any way represent an official Commonwealth view.

# **The Higher Education Loan Programme (HELP/HECS)**

## **– Microsimulation Modelling of Individual Repayment Prospects**

### **Introduction**

The Higher Education Contribution Scheme (HECS) was announced in the 1988/89 Budget and represented a radical new approach to social policy. There were two key elements to the scheme. The first was that university students should meet some part of the costs of their study given the private benefits which they would potentially accrue. This represented a return to the situation which had applied prior to the Whitlam Government's abolition of university fees in 1972.<sup>1</sup> The second element was that students would have the option to defer their charge, effectively borrow the equivalent amount from the Commonwealth Government, and only have to repay this debt if and when their income exceeded a specified threshold. There was no requirement to demonstrate an ability to repay the deferred charge; on the contrary, one goal of the policy was to ensure that those who did not have the ability to repay should not be required to do so.

The effect of the policy was that the Commonwealth Government built up a portfolio of debts owed by individual students with extremely variable repayment prospects. At the time the policy was considerable speculation about the proportion of debt that would not be repaid and figures as high as 30 percent were mentioned.

In subsequent years the Scheme has been repeatedly modified and extended. However, the two fundamental design features of optional deferral of a charge and income contingent repayment have remained.

The office of the Australian Government Actuary was first asked to make an assessment of the amount of debt that was unlikely to be repaid in 1994. At that time, the typical debts were small, there was very little repayment experience and only limited aggregated data could be provided. Accordingly, a simple model was adopted based largely on subjective judgments about salary increases, voluntary repayments and exits from the system.

Over time a significant body of income and repayment data has become available, allowing the construction of more sophisticated models.

This paper presents the latest iteration of the microsimulation model, which uses individually fitted Generalised Linear Models (GLMs) to simulate the lifetime income profiles for each individual holding a debt.

The paper consists of the following seven sections:

- a very brief survey of microsimulation modeling and income modeling;
- a discussion of the essential features of the HELP scheme and the implications for modeling the profile of repayments;
- a review of the modeling approaches that have been used over time and the rationale for the current approach;
- an overview of the available data and its salient features;
- a description of the model and the challenges associated with its implementation;
- an analysis of the potential shortcomings and possible means to overcome them; and

## **The Higher Education Loan Programme (HELP/HECS) – Microsimulation Modelling of Individual Repayment Prospects**

- reflections on the application of the approach to other actuarial problems

### **1. Microsimulation modeling of incomes**

Simple systems in the physical world can often be represented by exactly solvable models. The trajectory of a cannon ball, fired with a known direction and velocity, for example, can be predicted with some accuracy. If, however, there are thousands of cannons all firing in different directions with different velocities it becomes considerably more difficult to specify an exact solution of the system. In this case, it is simpler to go back to each individual cannon and simulate its behaviour to reach a numerical solution. This kind of modeling of large and complex systems by simulating the behaviour of individual units within the system is known as microsimulation modeling.

The application of microsimulation modeling techniques to socioeconomic modeling was pioneered in the late 1950's by Guy Orcutt in the United States<sup>2</sup>. Broadly speaking, microsimulation models can be classified along two dimensions. The first dimension is whether they are static or dynamic. Static models take a cross section of the population of interest at a particular point in time and model the instantaneous effects of policy or program changes. Dynamic models age the initial population over time, allowing the characteristics of the original population to be updated to reflect evolving demographic economic structures. The second dimension is whether the models are deterministic or stochastic. In a deterministic model the relationships are fully determined by the parameters defined within the model – the trajectory of the cannon ball is fully determined once we know the velocity and angle of projections. A stochastic model incorporates random processes, either to reflect the random nature of the underlying mechanisms being modeled – the cannon misfires occasionally – or to account for random influences due to incomplete model specification – we don't know the quality of the gunpowder used which will affect the initial velocity.

At much the same time as the first microsimulation models were being developed, there was increasing interest in understanding the dynamics of earnings mobility. The policy responses to poverty are likely to be very different depending upon whether it is seen as a permanent or transient state. Much of the initial research explored the use of Markov chains to estimate the probability of transition between different income classes (see for example Champernowne (1953)<sup>3</sup>). In the 1970s, an alternative approach was developed which used econometric models to estimate individual earnings functions based a range of personal characteristics and environmental variables. Lillard and Willis (1978)<sup>4</sup> produced the seminal paper on this approach, reporting on a model estimated using seven years of longitudinal data from the Michigan Panel Study on Income Dynamics. This was effectively a dynamic, stochastic microsimulation model, although the interest was still in earnings mobility and the results reported reflected this focus. Lillard and Willis' work has formed the basis for most of the income modeling work that has followed.

The fundamental structure of these models is a combination of a standard earnings function which depends upon measured variables such as education, gender, age, work experience and so on, a time varying factor and an error term comprising unmeasurable permanent differences between individuals and a transitory component. Since the initial work was done, there has been considerable discussion around the appropriate form of the error function, particularly whether it should be a random variable (and if so, which distribution is appropriate) or a time series process, and what assumptions should be adopted for the variance.

## **The Higher Education Loan Programme (HELP/HECS) – Microsimulation Modelling of Individual Repayment Prospects**

### **2. The HELP Scheme**

When HECS was introduced, the Commonwealth Government determined the amount which could be charged by public universities for undergraduate study as a flat amount per unit of equivalent full-time study load. Subsequent changes provided for differential charges depending on the course of study and, with the 2003 Higher Education Reforms, discretion for universities in setting their charges (up to a maximum specified by the Commonwealth for *Commonwealth supported places*<sup>5</sup> and without restriction for full fee paying places). Full fee paying places for undergraduate courses have now been abolished and there is no longer a minimum charge, but most universities charge at or close to the maximum for Commonwealth supported places.

Students enrolling at university are faced with a choice between paying their fees up front (with a 20% discount for fees on Commonwealth supported places) and deferring the charge and establishing a debt with the Australian Taxation Office. Any debt that has been outstanding for more than 11 months on the 1<sup>st</sup> of June each year is subject to indexation at the rate of increase in the CPI over the 12 months to the preceding March.

Repayments are assessed at the time a tax return is lodged based on the repayment schedule in force for the relevant financial year. Repayments can also be made on a voluntary basis at any time (with a 10% bonus for payments of \$500 or more). The compulsory repayment schedule for 2008-09 is shown in Table 1 below. Note that, unlike personal income tax, the repayment rate applies to all income, not just the income above the threshold.

**Table 1: Compulsory Repayment Threshold and Rates**

<b>Income Range</b>	<b>Repayment Rate</b>
Below \$41,595	Nil
\$41,595–\$46,333	4.0%
\$46,334–\$51,070	4.5%
\$51,071–\$53,754	5.0%
\$53,755–\$57,782	5.5%
\$57,783–\$62,579	6.0%
\$62,580–\$65,873	6.5%
\$65,874–\$72,492	7.0%
\$72,493–\$77,247	7.5%
\$77,248 and above	8.0%

Over recent years, up front payments have comprised about 30% of total revenue, voluntary payments around 10-15% with the remaining 55-60% of revenue having been derived from compulsory repayments.

## **The Higher Education Loan Programme (HELP/HECS) – Microsimulation Modelling of Individual Repayment Prospects**

Debt is only written off on the death of a debtor, meaning that debt can potentially stay in the system for 80 years.

Initially, the primary interest was in the face value of debt that would not be repaid. In more recent years, there has been increasing interest in the pattern of repayments and the discounted value of those repayments – what might be thought of as the real value of the debt.

The design of the scheme has three important implications from a modeling perspective:

- the non-linearity of the income-contingent repayment system means that it is not possible to rely on the average repayments across all debtors at a point in time, or for a particular cohort across time;
- the time frames are potentially very long and repayments can be spread over many years<sup>6</sup>; and
- the continuing modification of the scheme since the time of its introduction means that it needs to be feasible to incorporate changes to policy parameters.

There was also a hierarchy of client needs for the model to meet:

### *1. Estimates for inclusion in their Financial Statements which would satisfy the Auditor-General.*

Initially, the statements reported on the face value of the debt which was not expected to be repaid. More recently, changes in accounting standards changed the focus to the real value of the receivable as measured by the present value of expected repayments. This measure is sensitive to the timing of repayments and emphasises the need to take adequate account of periods of low income.

### *2. Financial metrics with sufficient granularity.*

Sound financial management practices demanded that our client have an understanding of a broader range of metrics relating to the scheme and how these indicators were changing over time. The model outputs thus needed to be sufficiently finely grained that the data could be “sliced and diced” in multiple ways.

### *3. Ability to test changes in policy parameters.*

Our client required a tool which they could use in a policy exploration context which would allow them to analyse the financial implications of changing scheme parameters. They were looking for the flexibility to change the policy parameters with access to a wide range of measures of the impacts.

## **The Higher Education Loan Programme (HELP/HECS) – Microsimulation Modelling of Individual Repayment Prospects**

### **3. A History of HELP Models**

In structural terms, there have been four distinct models used at various times by AGA to estimate the outstanding HELP debt which might not be repaid. In each case, the model framework has been driven by the available data.

#### *Speculative Model (1994)*

AGA first undertook an analysis of what was then HECS debt in 1994. At that time, the scheme had only been in operation for four and a half years. Individual level data was not available and even had it been, would have been of limited use since repayments to that date amounted to less than 7% of the debt that had been accumulated.

The model adopted was therefore based entirely on judgement, informed by publicly available information on participation rates, unemployment rates and graduate salaries. It was a deterministic model that aged a hypothetical cohort of debtors through time using decrements for death and unemployment and allowing for promotional and general salary increases and inflation of the outstanding debt in line with the assumed CPI.

#### *Cell Based Model (1995)*

In the following year, AGA first obtained access to unit record data relating to the scheme. This data provided comprehensive details on all debt transactions, together with a small number of demographic variables, most notably age and sex, for each individual who had ever incurred a debt. With the availability of this data, supplemented by the results of a microsimulation model developed by Professor Ann Harding<sup>7</sup>, we were able to develop a model which took account of the observed repayment experience.

The fundamental premise of the approach taken was that while HECS debtors could be expected to differ quite markedly in their lifetime salary progression and hence their propensity to repay their debt, it would be possible to identify groups which in aggregate would exhibit a relatively stable repayment pattern. After extensive analysis of the data, we decided to divide the data into six sub-groups based on gender and age when the study was completed: those aged less than 30 at the time of incurring their last debt, between 30 and 55 inclusive, and over 55.

The outstanding debt within each group was then further broken down by repayment history (whether a repayment had been received and at what rate) and the number of years since the study was completed. Three repayment categories were identified and annual transition probabilities and repayment rates calculated directly from the data for the years covered by the ATO data. For the years beyond the extent of the data, the simulation results from the Harding model were used. Smoothed factors were then fitted for each group over a period of 50 years allowing the outstanding debt to be projected over the same period and a doubtful debt percentage to be calculated for each gender/age group/repayment category/years since completion combination. This percentage was applied to the actual outstanding debt in each equivalent combination.

This model had the significant advantage that some of the assumptions were based on actual experience and as the experience increased, greater credibility could be assigned to real data. However, it relied upon stable policy parameters for its continuing validity – if the income level at which repayments were required was increased or decreased at a different rate from general wage growth, for example, the transition probabilities and repayment rates derived from the earlier data was no longer reliable.

## **The Higher Education Loan Programme (HELP/HECS) – Microsimulation Modelling of Individual Repayment Prospects**

This is exactly what happened when the repayment thresholds were substantially reduced in 1997 and a new approach was required. We started work on a microsimulation model in 1998.

### *Microsimulation Model Version 1 (1998)*

At that time we were able to obtain three years of reasonably reliable income data from the ATO and our income model was constructed as a two stage year by year Monte Carlo simulation. The first stage predicted the probability of having a non-zero income. This was done separately for each of the four possible combinations of zero and non-zero incomes in the previous two years with the probability functions depending upon one or more of gender, current age, age at completion of study, years since completion and number of years of study. The second stage then applied only to those simulated to have an income and assigned an income based on a regression model which depended upon the previous two incomes as well as age at completion of study and year of completion. Other variables which might have been expected to improve the explanatory power of the model such as course of study, occupation, age of children, part-time status, receipt of welfare payments were not available.

This model had the substantial advantage of predicting incomes rather than repayments so that compulsory repayments could be calculated under any combination of thresholds and repayment rates. However, the very limited amount of data, particularly in relation to young recently completed students where income data was often missing, meant that relatively little confidence could be placed in the fitted functions for this group. This was a major shortcoming, given that the bulk of the debt was held by those currently studying, which could only be addressed by waiting for more data to become available.

This Monte Carlo simulation model was retained for a further five years with the additional data which became available each year being incorporated in the regression modelling process. While this improved its performance somewhat, it became apparent that the model suffered from two significant deficiencies. The first was the onerous process involved in updating the model. The second was that the memoryless nature of the two state income model was resulting in too many transitions between the earning and non-earning states which in turn led to excessive volatility in incomes. While in any cross-sectional snapshot of the data the overall distribution of incomes generated by the model was reasonable, individuals exhibited considerable movements from year to year with too many debtors simulated to have zero incomes at some time over their working life. In combination with the non-linear repayment system, this meant that both the time to repayment and the level of doubtful debt were overestimated. With the increasing size of debts flowing from the introduction of differential HECS in 1997, the excessive variability produced by the model had the potential to significantly overestimate the debt and deferral subsidies<sup>8</sup> and this had become a serious threat to the reliability of the model.

### *Microsimulation Model Version 2 (2004)*

In 2004, we had 10 years of incomes and could start to see the differences between the incomes we were simulating and those we were observing. We needed a method which would more closely reproduce the income patterns evident in the data. The primary objective was to reduce the volatility we were seeing in the incomes projected over an individual's lifetime to a level more commensurate with that observed in the data.

The focus changed from modelling income on an annual basis, taking account of incomes in the previous couple of years and the limited demographic factors available, to modelling incomes for a person's entire lifetime. It relied on being able to capture the characteristics of the long term income profile in a relatively small number of parameters governing the progression and



## **The Higher Education Loan Programme (HELP/HECS) – Microsimulation Modelling of Individual Repayment Prospects**

variability of incomes. It met our primary objective of generating income profiles which more closely resembled experience. Importantly, it also addressed another problem with the previous model which was the heavy human input required to update the model. While it is computationally intensive, the level of intervention required has been considerably reduced and advances in computing power means that it runs in less time than the earlier model.

The new model is able to meet each of the three client objectives outlined in the previous section; it produces more robust estimates of the headline required for financial reporting as well as providing a facility to interrogate the results in more detail if required and model alternative program designs in a policy consideration context.

Unforeseeable changes in the underlying processes will always remain a challenge in modeling a complex system. However, the evidence over the four years (presented later in this paper) suggests that the life-time income projection approach is giving more robust results than the previous modeling framework.

The criteria set out in the previous section necessitate that the HELP system model project incomes at an individual level. The initial lack of longitudinal income data (or indeed any income data) meant that this was not feasible at the time we commenced modeling. As income data become available, we were able to move towards a microsimulation model, with an increasing focus on optimal computational approaches to improve the accuracy of model projections.

### **4. The Income Data**

The ATO has provided details of the HECS/HELP assessable income<sup>9</sup> for all those who have ever had a HELP debt for each financial year since 1993/94. Unsurprisingly, there are significant lags in tax lodgement, particularly for those still studying (who may not be required to submit a tax return) and it takes a number of years after completion of the relevant financial year before we have materially complete data. Some individuals are identified as not being required to submit a tax return but more often we don't know whether blank income records will at some time in the future be replaced with a recorded income. We receive updated income records for a particular financial year for a further three years after the initial data is provided.

In terms of the debt which won't be repaid, there are two groups of particular interest – those who never have an income above the repayment threshold and hence never make a payment; and those who have incomes above the threshold only intermittently and may not complete repayment during their working life. When estimating the discounted value of repayments, the incomes simulated from year to year also become important.

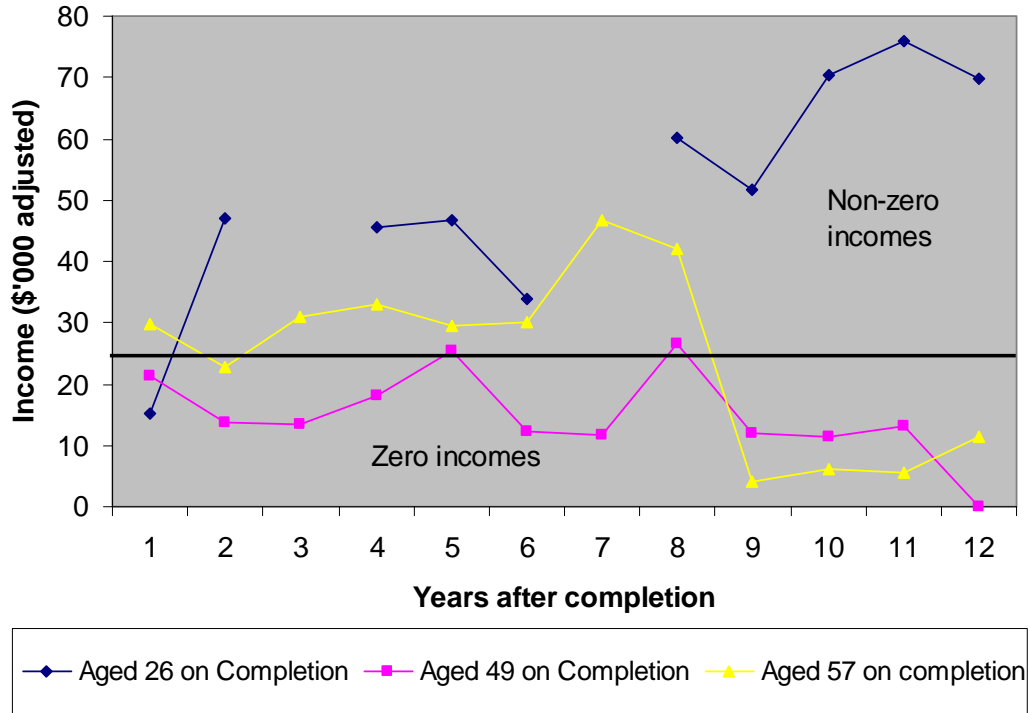
While the original microsimulation model focused on whether a person had an income or not, for version 2, we changed the concept of a non-zero income to refer to an income which might potentially give rise to a repayment<sup>10</sup>. Conversely, a zero income denotes an income which lower than the point at which the lowest repayment threshold might reasonably be set. This provides the capacity to model the impact of reductions in the repayment threshold.

After examining ten years of income history, it became apparent that there were three main groups of income profiles: a small but significant group of debtors had never had an income above the repayment threshold and, after ten years, was increasingly unlikely to ever commence repayment; a second group had a non-zero income (as now defined) every year from the time they started to repay and the third group oscillated between zero and non-zero incomes. The following two charts show some typical examples of intermittent earners of each gender. The incomes in

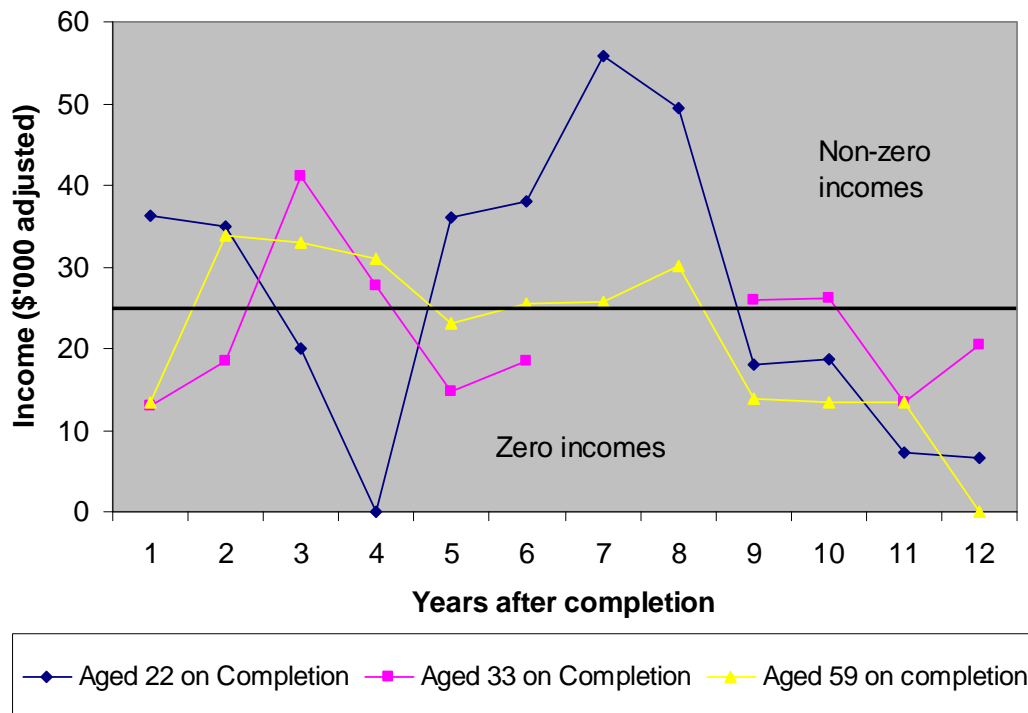
## The Higher Education Loan Programme (HELP/HECS) – Microsimulation Modelling of Individual Repayment Prospects

these charts are real incomes, that is, the incomes reported on the ATO data have been adjusted for general wage growth. Note also that these are not the income profiles of any actual individuals.

**Figure 1: Examples of intermittent income profiles – Males**



**Figure 2: Examples of intermittent income profiles – Females**

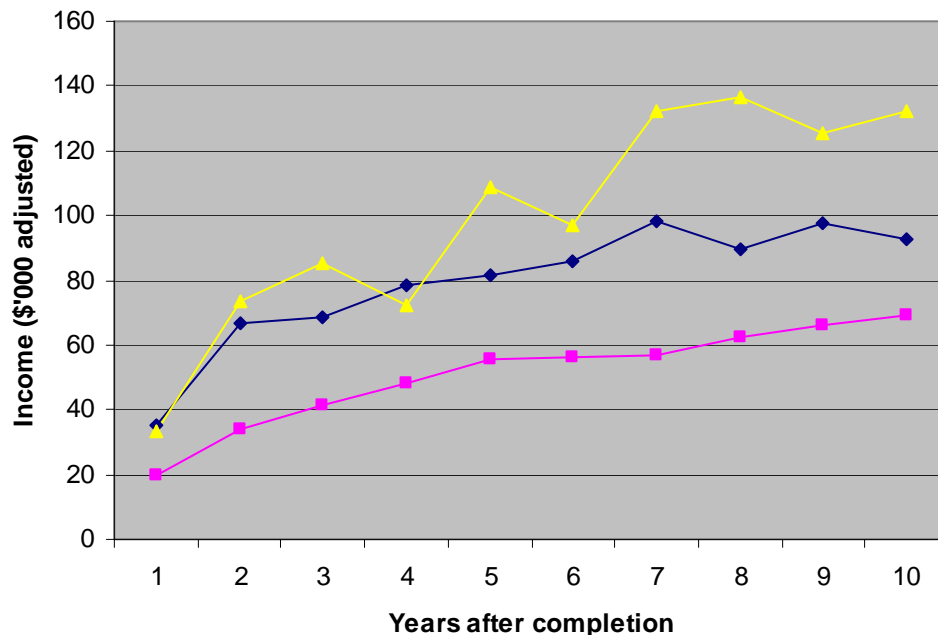


## The Higher Education Loan Programme (HELP/HECS) – Microsimulation Modelling of Individual Repayment Prospects

Intermittent earnings patterns are particularly relevant to females who are likely to have periods out of the workforce while raising children.

The model initially had to assign debtors between the non-earners, consistent earners and intermittent earners and, for the intermittent earners, allocate non-zero incomes appropriately over the person's lifetime. The model then had to project income profiles for those who had non-zero incomes. Individuals could be broadly assigned to one of three sub-groups based on the shape of their income profile. The first group exhibited the typical logarithmic earnings profile that professionals such as actuaries would expect to experience, with income rising fairly rapidly at first but with slowing rate of increase over time. Figure 1 shows the type of income profiles observed for this group.

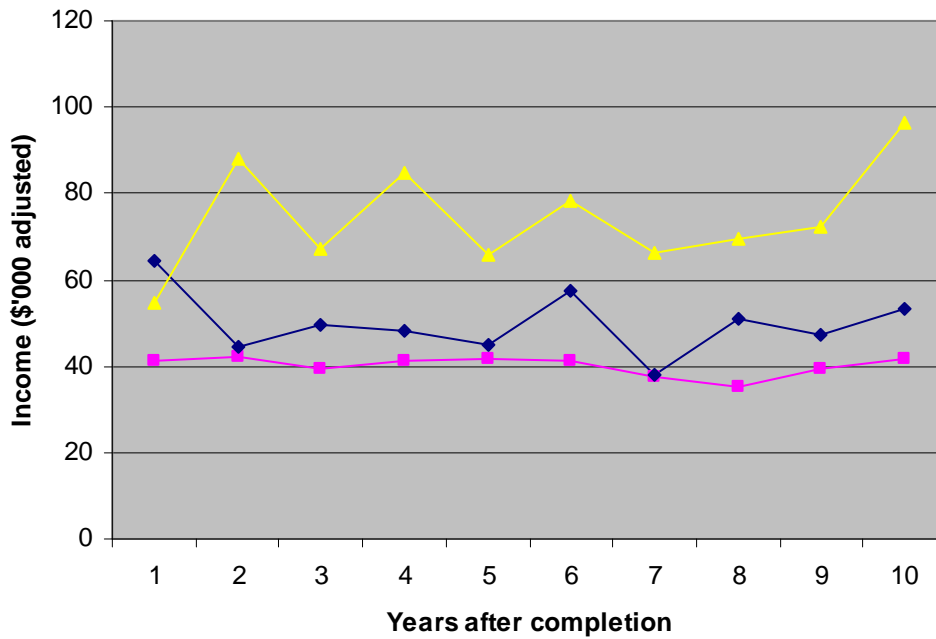
**Figure 3: Examples of promotional income profiles**



Our expectation, before looking at the data, was that most of the debtors would have this type of income pattern. However, there was a significant group who, when general wage growth was removed from the data, had income profiles that were essentially random variation around a mean.

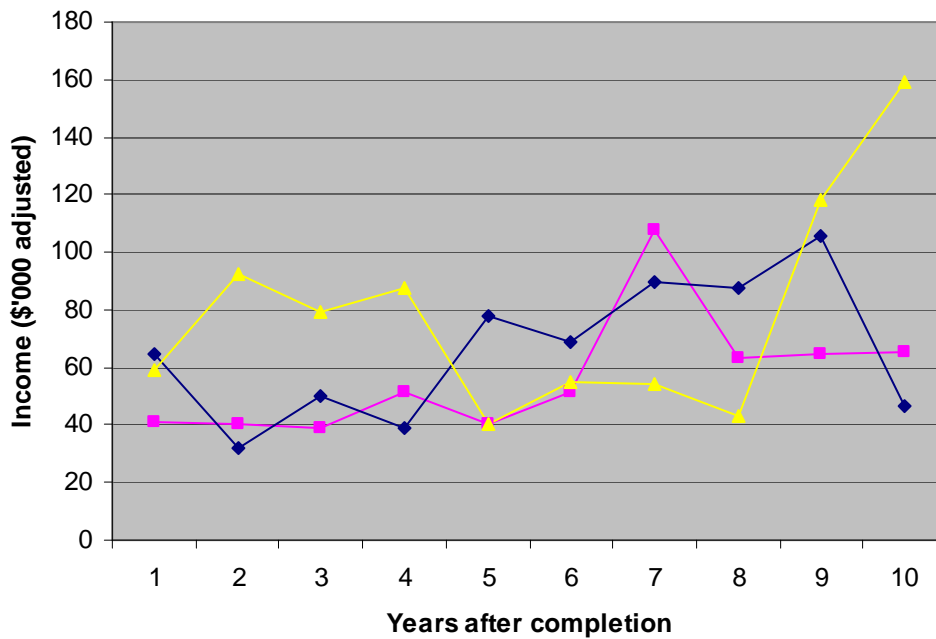
## The Higher Education Loan Programme (HELP/HECS) – Microsimulation Modelling of Individual Repayment Prospects

Figure 4: Examples of flat income profiles



The third and smallest group showed extreme variations in income from year to year and needed to be handled separately.

Figure 5: Examples of highly variable income profiles



Debtors were assigned to these three groups, based on their demographic characteristics and income data, to the extent that it was available, parameters were selected for simulation of lifetime income profiles.

## **The Higher Education Loan Programme (HELP/HECS) – Microsimulation Modelling of Individual Repayment Prospects**

The process was designed such that the model could be updated when new income data became available, with minimal reliance on subjective judgement.

### **5. The model and challenges in implementation**

The model is designed to project the amount and timing of compulsory and voluntary repayments over the next 45 years. It is assumed that virtually all repayments will have occurred within this period. Compulsory repayments are entirely driven by the underlying income module, while voluntary repayments depend partly upon incomes (but also on the time since study was completed, the amount of outstanding debt and previous repayment behaviour). The income module is therefore the single most critical element in the model. In order to capture the observed features of the data discussed above, it consists of two stochastic components merged with interaction effects:

- an income incidence model which projects when an individual will have a non-zero income, based on their income history (if any) and other demographic characteristics; and
- an income progression model which determines the amount of these incomes based on historical incomes and projected income incidence, along with demographic characteristics.

#### *Income incidence*

Income incidence is modelled using a conditional binomial distribution, with the probability of an income in a given year dependant on demographics and earning histories. Individual earning history is the most significant predictor of future earning capacity.

The model reflects the three distinct patterns of income incidence identified in the data – those who rarely earn, those who consistently earn, and those who earn intermittently. As noted above, a two state income model was unsuitable since without memory, the model results in too many individuals shifting between the zero and non-zero income states too frequently.

Demographic predictors were included in the model based on analysis using classification trees. Sex and age were found to be the most important variables. Ideally, males and females would be modelled together to retain a consistent error structure. However, analysis showed that the relationship between income and age is fundamentally different for males and females due to the maternity effect. This effect is difficult to model and project both because drops in income cannot be unequivocally attributed to maternity and because a sufficient volume of data on workforce exit to allow analysis is only now emerging, with virtually no reliable data on workforce re-entry.

Without adjustment, simulations using the conditional binomial distribution result in a spreading of incomes between individuals, with too many people having intermittent incomes and too few having no incomes. To overcome this problem, the model has been ‘zero-inflated’ to increase the number of people who never earn and avoid the artificial spreading of simulated incomes. Each individual who is yet to earn an income is assigned a non-zero probability of never earning an income in the future, given their income history and demographics.

Generalised linear models (GLMs) were used to estimate the probability of falling into the group who never have an income and, for those who didn’t fall into this group, the probability of having an income in the first, second and third years after completion. Separate models were used for the first 3 years following completion to capture transition to work. A single GLM was then used for the fourth and subsequent years after completion, with the age term sufficiently capturing the

## The Higher Education Loan Programme (HELP/HECS) – Microsimulation Modelling of Individual Repayment Prospects

life-time income profiles. The probability that an individual has not previously earned an income decreases rapidly for the first few years since completion and tapers off asymptotically to a level which corresponds to the proportion of the relevant population that will never earn an income. Thirteen years (that is, what is now the maximum income history of the earliest cohorts) without an income is considered to be an absorbing state. That is, beyond that point the probability does not depend upon the number of years since study was completed. Incorporation of serial correlation and heterogeneity in the variance of incomes was limited by the length of the income dataset.

The parameters fitted are sensitive to the proportion of the population projected to never earn an income which will trigger a repayment. Even after applying zero-inflation, the model will simulate a very small proportion of individuals having a low but non-zero probability of commencing earning many years later. This is an unavoidable consequence of fitting a model based on data which covers less than a third of the projection period. However, it also appears to be borne out by the data where we are still having a small number of people from the earliest cohorts commencing repayment.

The parameterisation of the GLMs involved balancing the conflicting requirements of parsimony and explanatory power. The model needed to be somewhat parsimonious, involving limited use of polynomials and interaction terms, in order to avoid nonsensical outcomes. At the same time, it needed to be powerful enough to explain the significant trends in income incidence. This balance was difficult to attain with age, particularly for females. Females have a starting-work effect at early ages, followed by a maternity effect, followed by a return-to-work effect, followed by a retirement effect. Additional terms were therefore required in the GLM. The projected incomes for males and females over the next 45 years were examined to ensure that the polynomial resulted in a sensible age-profile of incomes. Most importantly, unstable GLMs which resulted in increasing income incidence at the oldest ages were avoided.

The probability of never earning a non-zero income was specified as follows, conditional on number of years without an assessable income.

$$\Pr\left(\sum_{j=i+1}^{\infty} Inc_j = 0\right) = f(I(x \leq 25), I(x \geq 45), x, x^2 | i)$$

where  $x$  is the age at completion

$i$  is the number of years since completion

$Inc_i$  is the projected income in year  $i$

$f(x/i)$  is used to denote a generic function of  $x$  conditional on  $i$

Individuals who are yet to make a repayment are randomly assigned to the group projected to never earn, with individual probabilities being based on age and the number of years of income history without a repayment. The process is not iterative and the assignment is performed once, as at the valuation date.

The probability of earning a non-zero income  $i$  years following completion of university was specified as:

**The Higher Education Loan Programme (HELP/HECS)  
– Microsimulation Modelling of Individual Repayment Prospects**

$$\Pr(Inc_i \geq 0) = \begin{cases} f(y, y^2), & i = 1 \\ f(y, y^2, I(Inc_{i-1} > 0)), & i = 2 \\ f\left(y, y^2, i, (x+i), I(i > 4), I(i > 8), I(Inc_{i-1} > 0), I\left(\sum_{j=1}^{i-1} Inc_j > 0\right), \left(\sum_{j=1}^{i-1} Inc_j > 0\right) / (i-1)\right), & i \geq 3 \end{cases}$$

where  $y$  is the age in projection year  $i$

That is, the probabilities of earning incomes at 3 or more years post completion are based on the number of years post completion, interacting age effects as well as 3 income variables:

- an indicator variable specifying whether the individual is yet to earn an income
- the proportion of incomes earned to date; and
- the previous income.

Note that the functions also include some variable interactions. In additions, the age effects for females were necessarily more complicated, involving an indicator variable for whether the person is between age 25 and 34 (to capture the majority of the maternity effect).

The age and age squared terms capture the transition from full-time to part-time employment to retirement.

*Income amounts*

The income incidence model simulates the occurrence of future non-zero incomes. The income progression model builds on this information to simulate a lifetime income profile for all individuals with an outstanding debt.

The income progression model determines the amount of real income (that is, with general wage growth removed) received by an individual for a particular year. Our initial hypothesis was that most individuals would exhibit a significant upward trend in incomes reflecting promotional advancement with increasing employment experience and competence (as illustrated in Figure 3). However, the data showed that this was only a subset of the non-zero income group and the first step was to identify this subset. We refer to this subgroup as the category with a significant income trend.

For those with an income history, a binomial model was used to allocate individuals to the significant income trend category, based on the significance and magnitude of the slope parameter of a log linear regression fitted to their incomes. The probability of assignment to the regression group was calculated conditionally on the number of years since completion, using the following generic function with normally distributed errors, along with income statistics where available.

**The Higher Education Loan Programme (HELP/HECS)  
– Microsimulation Modelling of Individual Repayment Prospects**

$$\Pr(\text{Regression}_i) = \begin{cases} f(g, y, y^2, d), & i = 0 \\ f(g, y, y^2, d, \text{Inc}_1), & i = 1 \\ f(\hat{R}_i^2, \hat{\beta}_i, \hat{\mu}_i, \hat{\sigma}_i^2, g, y, y^2, d | i), & i \geq 2 \end{cases}$$

where  $y$  is the age in projection year  $i$

$g$  is an indicator variable for gender

$d$  is the duration of study

$\hat{R}_i^2, \hat{\beta}_i, \hat{\mu}_i, \hat{\sigma}_i^2$  are the mean, standard deviation, significance, and slope of the log-linear regression.

The classification is not iterative. Where an individual had no income history, the allocation was performed based on demographic variables alone, drawing probabilities from the historic distribution of incomes by age and gender. For example, an individual who has just completed university and has no income history might be randomly assigned regression parameters. This approach is simpler than simulating 2 years of incomes using point estimates, and randomly assigning to the regression group in projection year 3.

Where an individual debtor had no income history, the appropriate parameter distributions for the category to which the debtor has been allocated are used to assign the requisite parameters which are then used to generate an income profile.

Where we had some incomes, but not a full history, the information available from the actual incomes were used to predict the parameters for the income profile, simulating required parameters using conditional Gamma distributions. The model thereby applies increasing credibility to observed regression statistics as the length of income history increases. Note that the classification is made only once and not recursively.

These proportions were then used to randomly assign those individuals without little or no post study incomes to the significant income trend category based on their demographic characteristics. Overall, around half the debtors were assumed to have a trend growth in incomes. The remaining debtors were assumed to have no significant trend in incomes.

For those with a complete history, regression classifications were based on the significance of the regression and whether the parameters fell within certain bounds (ie high standard deviation, negative or very high slope or intercept).

The next step was to model the incomes for each group. For those with a significant income trend, a log linear regression was used to fit a model to the incomes of those with non-zero incomes in all of the available years. The basic functional form for individual  $j$  in year  $i$  was:

$$\text{Inc}_{i,j} = \alpha_j + \beta_j \times \ln(i + \lambda_j) + \varepsilon_j$$

The ' $\lambda$ ' term allowed differences in the steepness of the initial trend and how quickly it tapered off to be captured in the fit. While in theory the ' $\lambda$ ' variable could be fitted as part of the regression process, in practice, the use of two values of ' $\lambda$ ' appeared to capture the main elements of the data and in the interests of simplicity it was treated as a constant, taking the value of either 1 or 10. The smaller value of  $\lambda$  was used where the income profile showed a steeper progression of incomes.



**The Higher Education Loan Programme (HELP/HECS)  
– Microsimulation Modelling of Individual Repayment Prospects**

The remaining debtors who were predicted to have no significant trend in incomes were modelled with a point estimate, with random variation around their observed mean income.

In each case, these functional forms were used to fit the data for those with non-zero incomes in all years. For those fitted to a point estimate, an estimate of the mean and standard deviation of each profile was derived. The set of observed means and standard deviations from these individual regressions were then themselves fitted to a gamma distribution (noting that neither the mean nor standard deviation could take on a negative value).

The point-estimate parameters were then projected using the following generic functions for mean and variance, with Gamma distributed errors.

$$\sigma = \begin{cases} f(g, x, x^2, d), & i = 0 \\ f(g, x, x^2, d, Inc_1), & i = 1 \\ f(g, x, x^2, d, \hat{\mu}_i, \hat{\sigma}_i), & i \geq 2 \end{cases}$$

$$\mu = \begin{cases} f(g, x, x^2, d | \sigma), & i = 0 \\ f(g, x, x^2, d, Inc_1 | \sigma), & i = 1 \\ f(g, x, x^2, d, \hat{\mu}_i | \sigma), & i \geq 2 \end{cases}$$

where  $\hat{\mu}_i, \hat{\sigma}_i$  are the estimated mean and standard deviation fitted to the data.

$$\hat{\mu}_i = \sum_{j=1}^{i-1} Inc_j / (i-1)$$

$$\hat{\sigma}_i = \sqrt{\left( \sum_{j=1}^{i-1} Inc_j^2 - \hat{\mu}_i^2 \right) / (i-1)}$$

That is, the mean and standard deviation point estimate for a *point-estimate projection* for an individual is modeled as a function of gender, age, duration of study, and the mean and standard deviation for the individual's historic income profile. The model structure means that variability in the projected income amounts is captured by not only the error term, but also by variation in the parameters.

For the population whose profile incorporated a significant trend element, parameters for the intercept and slope of the income curve and the associated variability were also required. Again the values derived from the individual regressions were used to fit a gamma distribution for the parameters themselves.

The regression parameters were projected using the following generic functions, with Gamma distributed errors.

$$\alpha = \begin{cases} f(g, x, x^2, d), & i = 0 \\ f(g, x, x^2, d, Inc_1), & i = 1 \\ f(g, x, x^2, d, \hat{\alpha}_i, \hat{\beta}_i, \hat{\sigma}_i), & i \geq 2 \end{cases}$$

**The Higher Education Loan Programme (HELP/HECS)  
– Microsimulation Modelling of Individual Repayment Prospects**

$$\sigma = \begin{cases} f(g, x, x^2, d | \alpha), & i = 0 \\ f(g, x, x^2, d, Inc_1 | \alpha), & i = 1 \\ f(g, x, x^2, d, \hat{\sigma}_i | \alpha), & i \geq 2 \end{cases}$$

$$\beta = \begin{cases} f(g, x, x^2, d | \alpha, \sigma), & i = 0 \\ f(g, x, x^2, d, Inc_1 | \alpha, \sigma), & i = 1 \\ f(g, x, x^2, d | \alpha, \sigma), & i \geq 2 \end{cases}$$

where  $\hat{\alpha}_i, \hat{\beta}_i, \hat{\sigma}_i$  are the estimated intercept, slope and standard deviation for the log-linear regression fitted to the data.

That is, the parameters for a *regression projection* for an individual is modeled as a function of gender, age, duration of study, and the parameters of the regression fitted to the individual's historic income profile. Again, variability in the projected income amounts is captured by the error term as well as by variation in the parameters.

Lifetime income patterns are captured in terms of real (promotional) income increases approaching an asymptote. The decline in real incomes as full-time gives way to part-time employment is captured by the influence of the age and age squared terms in the income incidence model.

*Interaction between income incidence and amount*

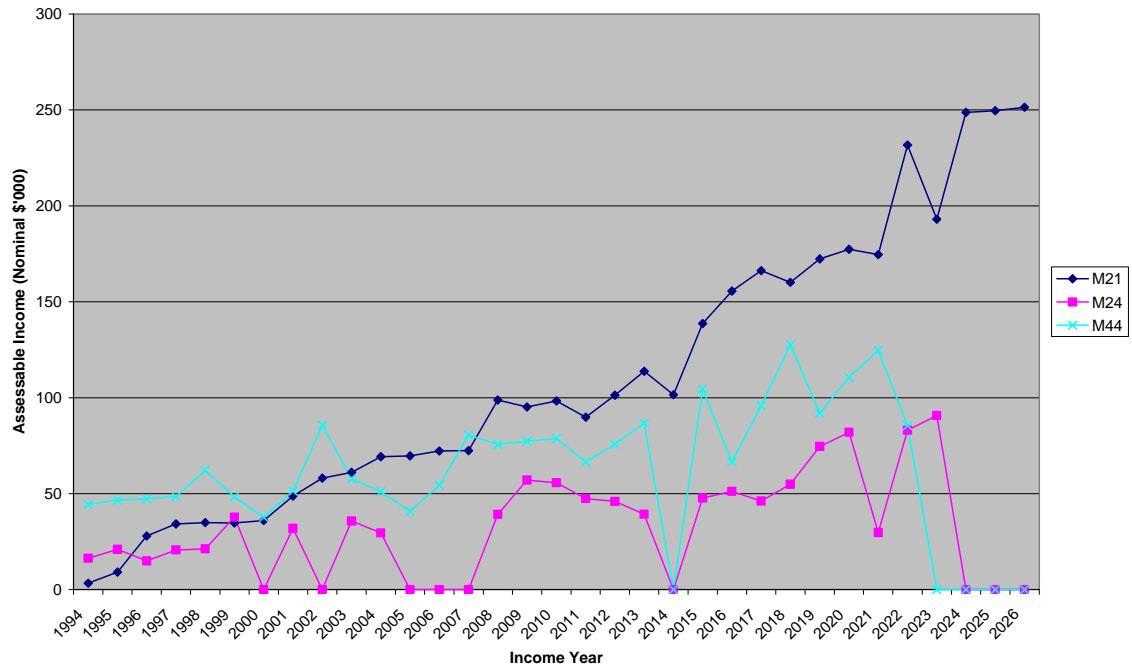
The future incomes simulated from the income progression model are assumed to occur in succession, in years where the income incidence model indicates that an income will occur. That is, a basic assumption is made that those who miss an income will resume work at the same stage in the progression. For example, if an individual is simulated to have a continuing income trend and miss the next three incomes, the income in three years time is taken as the next income in the progression (the next income in the progression is deferred for three years).

However, people with infrequent incomes will usually have low income amounts. For those simulated to have an income progression, the intercept parameter was scaled based on the incidence of incomes observed and/or simulated in their first twelve years following completion of university to reflect this depression of incomes.

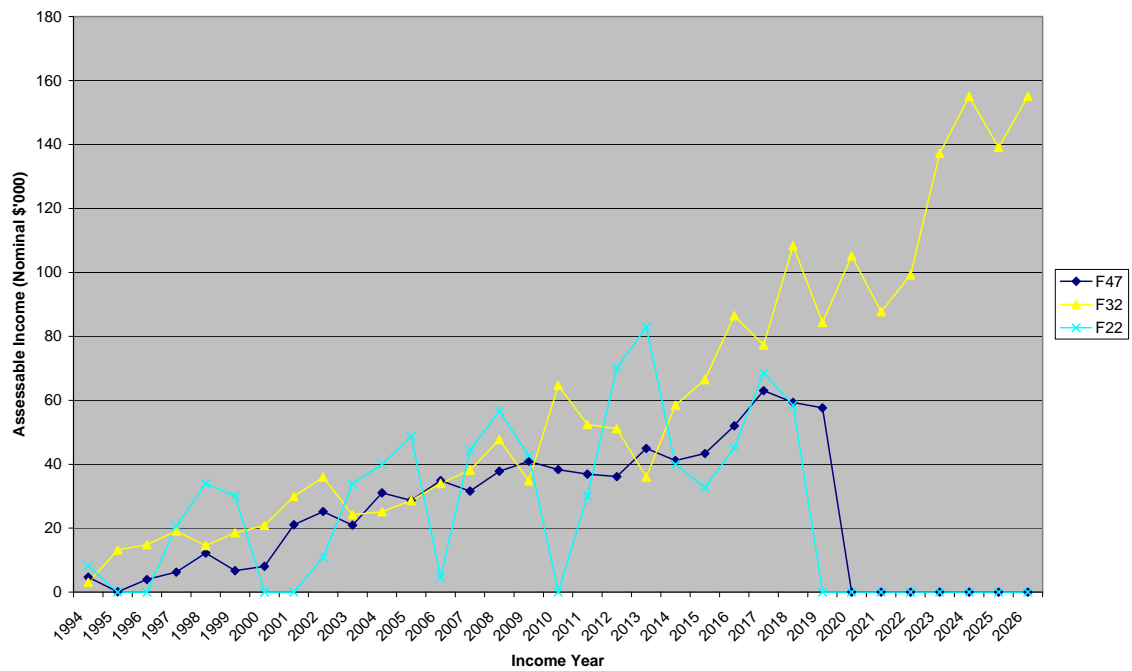
The following three charts provide some examples of income profiles generated by the model. Note that unlike the earlier charts these are nominal incomes. The first chart shows three males who have income records since 1994. One was an intermittent earner over that period and the other two weren't. The numbers in the legend indicate the age of completion.

## The Higher Education Loan Programme (HELP/HECS) – Microsimulation Modelling of Individual Repayment Prospects

**Figure 6: Generated income profiles – males with an income history**

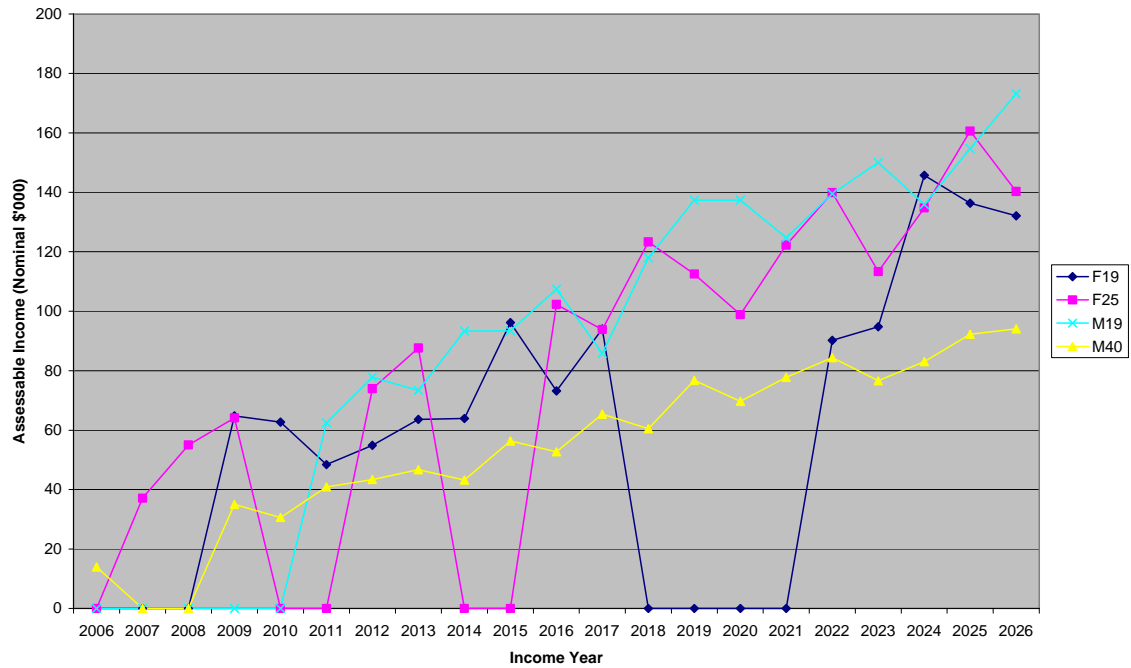


**Figure 7: Generated income profiles – females with an income history**



## The Higher Education Loan Programme (HELP/HECS) – Microsimulation Modelling of Individual Repayment Prospects

Figure 8: Generated income profiles – individuals with no income history

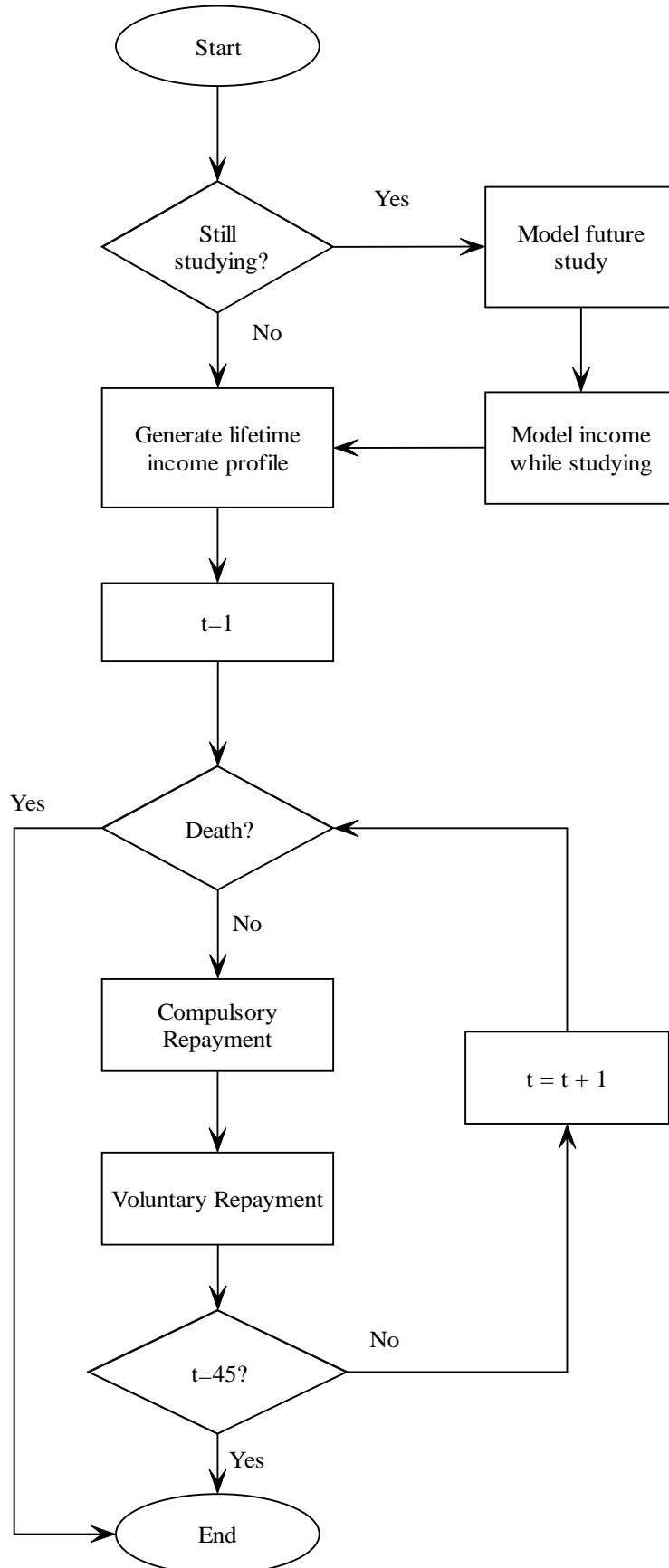


### *Debt repayments*

The primary function of the model is to estimate future repayments and, once the simulated income profiles have been generated, this is a relatively simple process. The following flowchart illustrates how the income module fits into the broader model framework.

The Higher Education Loan Programme (HELP/HECS)  
– Microsimulation Modelling of Individual Repayment Prospects

Figure 9: Flowchart of model processes



**The Higher Education Loan Programme (HELP/HECS)  
– Microsimulation Modelling of Individual Repayment Prospects**

The following table briefly describes each of the processes identified in the flowchart.

**Table 2: Processes incorporated in the model**

<b>Process</b>	<b>Description</b>
Still studying?	Incurred a debt in the semester immediately preceding the valuation date $\Rightarrow$ still studying
Model future study	Stochastic process modelled on a semester by semester basis with the load in each successive semester dependent upon:  Type of debt last incurred (HECS-HELP, FEE-HELP or OS-HELP) Years of study to date Study load in previous semester Gender  Once a load of zero is simulated, no further study is assumed to occur
Income while studying	Stochastic process, depending upon gender and current age
Lifetime income profile	As described in this paper
Death?	Stochastic process depending upon gender and current age
Compulsory Repayment	Income and compulsory repayment thresholds and repayment rates
Voluntary Repayment	Voluntary repayment is modelled as a three stage stochastic process: firstly, whether a repayment is made; if yes, whether the entire outstanding debt is repaid; and if only part is repaid the proportion repaid. The probabilities depend upon:  Amount of outstanding debt Income Whether any previous voluntary repayments have been made in the preceding ten years.

Three macroeconomic parameters are incorporated in the model:

- the annual growth in the consumer price index (CPI) which is used for indexing the outstanding debt at the end of each year;
- the general growth in wages, with income transitions derived using incomes adjusted to remove the effect of general wage growth based on Average Weekly Earnings (AWE), and growth in average wages superimposed on the increases predicted by the income module;

## The Higher Education Loan Programme (HELP/HECS) – Microsimulation Modelling of Individual Repayment Prospects

- the discount rate used to value future repayments of debt

Using a discount rate set equal to the inflation rate gives the amount which is not expected to be repaid in current nominal dollars. This can then be directly compared with the figure for current outstanding debt to give a percentage of debt that will not be repaid. Setting a non-zero real discount rate allows the deferral cost of applying indexation at the CPI rate to be quantified. It also enables a present value of the repayment income stream to be calculated.

### 6. Model performance

Considerable attention was given to testing the accuracy of the income progression model in projecting aggregate levels of repayments and residual debt. Accuracy was judged against actual experience over the past twelve years. It is important to note that the outcomes for an individual are likely to vary considerably from the model projections. The critical test is whether the model achieves a distribution of incomes which results in aggregate simulated repayments which correspond to actual experience. There was a reasonably accurate fit to actual repayments.

The other important test was the power of the model to accurately project future revenue. The following table shows the difference between the revenue projected by the model for the year after the valuation date and the revenue actually received. It has been difficult to take a longer timeframe because of the multiple policy changes which have impacted on repayments. Note that the original speculative model did not model projected repayments.

**Table 3: Deviation between revenue projected by the model and actual outcome**

Financial Year	Model	Percentage Deviation
1995-96	Cell based	11.2%
1996-97	Cell based	10.3%
1997-98	Cell based	-4.5%
1998-99	Micosimulation v.1	-18.5%
1999-00	Micosimulation v.1	1.2%
2000-01	Micosimulation v.1	-5.4%
2001-02	Micosimulation v.1	7.1%
2002-03	Micosimulation v.1	-4.9%
2003-04	Micosimulation v.1	12.5%
2004-05	Micosimulation v.2	13.4%*
2005-06	Micosimulation v.2	0.9%
2006-07	Micosimulation v.2	0.5%
2007-08	Micosimulation v.2	-2.1%

\* The 2004-05 result was substantially affected by the reduction in the discount available on voluntary repayments from 15% to 10% with effect from 1 January 2005. This resulted in a large bring forward of voluntary repayments to the second half of the 2004 calendar year. Our hypothesis, which was subsequently borne out by experience was that this would be a transient effect. There were also problems in the initial specification of the model which were corrected in the following year.

## **The Higher Education Loan Programme (HELP/HECS) – Microsimulation Modelling of Individual Repayment Prospects**

While it is still relatively early days, the performance of the second version of the microsimulation model appears to be significantly better than its predecessors in predicting revenue outcomes. The fact that these outcomes have been achieved with all changes to the model being internally generated from the introduction of new data is particularly pleasing.

Another obvious check is the cross sectional income distributions generated by the model compared with sample based cross sectional data from other sources such as the ABS Income Distribution Survey. However, there are a number of reasons why this is not a simple matter. Our population is a subset of the Australian population with particular characteristics. They have attended university as a Commonwealth supported undergraduate between 1989 and 2007, or as a postgraduate between 2002 and 2007, or as a full-fee paying student between 2005 and 2007, and they have incurred a debt for at least one of their semesters of study. The exclusion of those who pay for all of their study up front rather than deferring some part of the charge is likely to significantly bias the results.

Over time we will be able to compare the income distributions generated by the model with the actual income distributions. The lags in the data becoming available from the ATO meant that it was not possible to do this for the version of the model we are currently using.

### **7. Potential shortcomings with the approach and further work:**

There are a number of areas where the model is obviously vulnerable.

#### *Scheme maturity*

Although the HECS scheme commenced in 1989, it has by no means reached maturity in actuarial terms. We still have a relatively short sequence of data when measured against the time scale of the scheme. For example, it is impossible to analyse repayment patterns for an individual who completed study 20 years previously as there are no such debtors. The relatively large debts now being incurred mean that repayment may extend over many years and the incomes in the distant future may be an important driver of the debt not expected to be repaid (the present value of repayments is less sensitive to these long term repayments).

#### *Stationarity/stability of the conditional distributions*

The model is applied to all cohorts regardless of the source of their debt or the years in which study was undertaken. Cohort experience is unlikely to be consistent; it is possible that there have been significant changes between cohorts from different completion years which are not captured by the explanatory variables in our model.

The most reliable data that we have is in relation to those who completed their study in the years 1990 to 1993 inclusive. These cohorts have had the longest time to make repayments and thus provide the best evidence of long term repayment patterns. At the same time, however, their debts are relatively small compared with the debts currently being incurred and this raises some questions about the transferability of their experience to those who are currently completing. The model simply attempts to fit the observed behaviour over the period for which we have data as accurately as possible.

Changes in scheme such as the impact of the increased thresholds could be expected to have a significant impact on cohort experience, affecting both the amount and timing of repayments. For those completing in the last few years, the much higher levels of debt on completion mean that we are unlikely to have a reasonable degree of certainty until well into the 2020s.



## **The Higher Education Loan Programme (HELP/HECS) – Microsimulation Modelling of Individual Repayment Prospects**

### *Inappropriate functional forms*

The model generates income profiles using one of two broad functional forms. In practice, there is no reason for incomes to comply with a parametric function. It is apparent, in comparing the observed data with the generated profiles that discontinuities in historical incomes will lead to greater variance. This variance will show up as year to year deviations rather than intermittent discontinuities. At this stage, this does not seem to be causing problems but when increasing numbers of debtors have repayment periods extending over many years more thought may need to be given to this issue.

### *Macro-economic feedback*

The model does not attempt to take into account the effects of the macroeconomic environment on incomes, and hence on repayments, but this will obviously have major effects. For example, there is a possibility that those completing their study at a time of full employment are likely to have a significantly better lifetime earning potential than those entering the labour market during a recession. Our income data series commences in 1994 and thus encompasses the longest sustained period of economic growth since Federation. Thus our data reflects a particularly benign period for graduate income prospects.

In early 2009, the economic outlook for the world, including Australia, appears very grim. The possibility of sharp and sustained downturn in the Australian economy cannot be ignored and would pose particular challenges for projecting incomes. The current modelling framework relies entirely on historical income profiles for generating future incomes and this means that projected incomes will reflect the economic circumstances that have prevailed over the past decade. Without adjustment, the model will assume that the strength of the Australian economy over recent years will persist into the future. We may need to incorporate more conservative assumptions, particularly in relation to graduate salaries and the rate of transition to repayment for the most recent cohorts and this will not occur automatically within the model. There is scope to incorporate a lower short to medium term general wage growth assumption which would allow some of the effect of a recession to be incorporated. It would also be possible to manipulate the income profiles generated by the model to depress repayment patterns or decrease the probability of being categorized as a frequent earner. Judgements on the size and duration of such an intervention will unavoidably involve a considerable subjective element.

In the long-term, we will build up a range of income profiles observed over different economic conditions and it may be possible to incorporate expectations around the macroeconomy in the generation of future income projections. This would, however, add significantly to the complexity of the model and a cursory examination of economic history highlights the difficulty of projecting macroeconomic trajectories.

## **8. Broader applications**

The primary advantage of the model from the HECS perspective was that it was able to simulate a wide range of outcomes using a relatively parsimonious specification. After the initial judgment about the classification of income profiles into the three broad groups and the functional forms to be adopted for each group, the model is able generate profiles with a mix of stable and highly variable profiles. This is essential where we need to capture the effect of variability in individual outcomes under a non-linear repayment system.

The same approach could be used in modeling incomes for other purposes where there is a non-linear element and longitudinal outcomes are of interest, such as income tested social security payments or across year income averaging provisions in the tax system. It needs to be

## **The Higher Education Loan Programme (HELP/HECS) – Microsimulation Modelling of Individual Repayment Prospects**

emphasized that we are talking about the extension of the approach, not the actual output of this model which relates only to a particular unrepresentative subset of the Australian population.

It could also be expected that a similar approach might have benefits in other areas where results are sensitive to variability of the experience. For example, in looking at usage of health services, it could be expected that there would be an underlying trend towards increased usage with age, but some individuals will have sustained high usage, some will have consistently low usage and there will be a group with highly variable usage. The design of a health insurance product which includes an excess would need to take account of this variability in health status across individuals rather than just average levels of usage.

---

1 The 1986-87 Budget had introduced a Higher Education Administration Charge (HEAC) but this was a relatively small charge (\$250 per annum) specifically to meet administration costs rather than making a contribution to total education costs. The HEAC was abolished when HECS was introduced in the 1988-89 Budget.

2 Orcutt, G.H., "A new type of socio-economic system", *Review of Economics and Statistics*, Vol. 39, No. 2 (May, 1957), pp. 116-123

3 Champernowne, D.G., "A model of income distribution", *The Economic Journal*, Vol. 63, No. 250 (Jun., 1953), pp. 318-351

4 Lillard, L., Willis, R., "Dynamic aspects of earnings mobility", *Econometrica*, Vol. 46, No. 5 (Sep., 1978), pp. 985-1012

5 A Commonwealth supported place is a higher education place for which the Commonwealth makes a contribution towards the cost of a student's education.

6 The average debt of a student at the time of completing their study is now around \$14,000 and could be expected to grow as the higher charges which took effect in 2005 work their way through the system (students enrolled before that time were subject to grandfathering arrangements).

7 The model from which this dataset was derived is described in Harding, Ann, *Lifetime Income Distribution and Redistribution: Applications of a Microsimulation Model*, 1993, New Holland, Amsterdam.

8 The deferral subsidy is the subsidy provided by the Commonwealth indexing the outstanding HECS debt at a rate lower than its borrowing cost. It is calculated as the difference between the present value of repayments discounted at the rate used to index outstanding debt (currently the CPI) and a discount rate which represents the real cost of funds to the Commonwealth.

9 Assessable income for the purposes of determining compulsory repayments is defined as taxable income plus net rental losses on rental properties to the extent that they have reduced taxable income.

10 In terms of what we are trying to model an income of \$5,000 is effectively equivalent to an income of \$20,000 since neither of them will give rise to a repayment. Not trying to simulate these incomes considerably simplifies the model. At the same time, once we approach the repayment threshold, modelling income becomes much more important since an income just below will give rise to a very different result from an income just above. Currently we are using an income of \$25,000 (in 2007 dollars) as the non-zero income level.