The impact of HECS debt on socioeconomic inequality and transition to adulthood outcomes

Claire Houssard, Anne Sastro, & Suzana Hardy

ECON 4106: Policy Evaluation
Session 2, 2010

Disclaimer

This paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Project was initiated and is funded by the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this paper, however, are those of the author and should not be attributed to either FaHCSIA or the Melbourne Institute.

1. Introduction

The Higher Education Contribution Scheme (HECS) has been a major fixture in Australia's tertiary education landscape. Paradoxically, little research has been conducted with respect to the impact of HECS on social outcomes. This study presents a hypothetical analysis of the impact of the 2004-05 HECS-HELP policy changes on socioeconomic inequality and transition to adulthood outcomes. We seek to remedy the technical deficiencies evident in Marks (2009) by proffering alternative methods. Summative analysis using the Household, Income and Labour Dynamics in Australia (HILDA) Survey suggests that individuals with HECS debt are less likely to own a house, and less likely to have a higher socioeconomic status, relative to their counterparts without HECS debt. However, we acknowledge the presence of a possible unobserved selection bias. We consider a difference-in-differences (DID) approach, combined with propensity score matching (PSM) as a means to construct valid counterfactual groups, and as a way to remove time and individual specific fixed effects.

2. Institutional Framework

Since the introduction of HECS, the cost of an individual's higher education in Australia has been covered partly by government contributions paid directly to the university, and partly through student contributions, which can be paid upfront or administered as an income related loan. We will now discuss in some detail the HECS policy changes since 1989. Our discussion summarizes information from Beer & Chapman (2004), Marks (2009), as well as trends observed in *figure 2.1*, which illustrates the historical government set student contribution rates. [1]

Between 1989 and 1996 the federal government set student contributions at a flat rate across all courses, and all domestic students who attended tertiary institutions did so with a HECS loan. In 1989 the student contribution amount for a typical study load was $1,800 per year, increasing on average by 4% each year to $2,448 in 1996. In 1997, a three-tiered structure was introduced, with student contribution rates determined by study area. There was also a marked increase in contribution rates of between 30% and 50%, continuing to increase on average by 2% each year. In addition to these financial changes, up to 25% of enrolments in university courses could now be comprised of full-fee paying domestic students.

In 2005, there were further substantial policy reforms, with tertiary loans for undergraduate study re-branded as HECS-HELP[2] loans. Universities, rather than the government, were now given the power to determine their own fees within a federally determined range. In 2005, the minimum of each band was $0, while the maximum of each band was 25% higher than the corresponding 2004 HECS student contribution levels. Although universities were free to set their fees anywhere within the range, the current 2010/11 student contribution rates of effectively all Australian universities are set at exactly the maximum allowable amount, and this certainly appears to have become the standard within the first 2 years of the policy change. Since 2005 these maximum amounts continue to increase by 2% each year on average.

[1] We focus only on undergraduate bachelor's degree students in our sample, and so we do not discuss postgraduate HECS policies.
[2] We will henceforth refer to HECS and HECS-HELP debts collectively as HECS debts.

Education and nursing were the only courses at the time not to experience this jump, due to their inclusion in the new “National Priority” fee band which was introduced as a fourth tier. Education and nursing course fees did however suffer the 20% jump in 2010 when they were moved from the priority band into band 1. Conversely, the student contribution rates for mathematics, science, and statistics courses fell by 75% in 2009 when these courses were moved from band 2 to the priority band.[3]

Another important policy change in 2005 was the introduction of FEE-HELP loans. Now, up to 35% of course enrolments could be comprised of full-fee paying domestic students, and these students now also had access to income related loans to pay for these fees. Individuals with these types of tertiary loans are captured by our key control variable, HECS debt status. However they make up only 10% of domestic students in total, and so we assume their impact on our results is negligible.[4]

An additional issue is that the HILDA survey does not provide information on the year an individual began their degree. This is important since the student contribution rate an individual

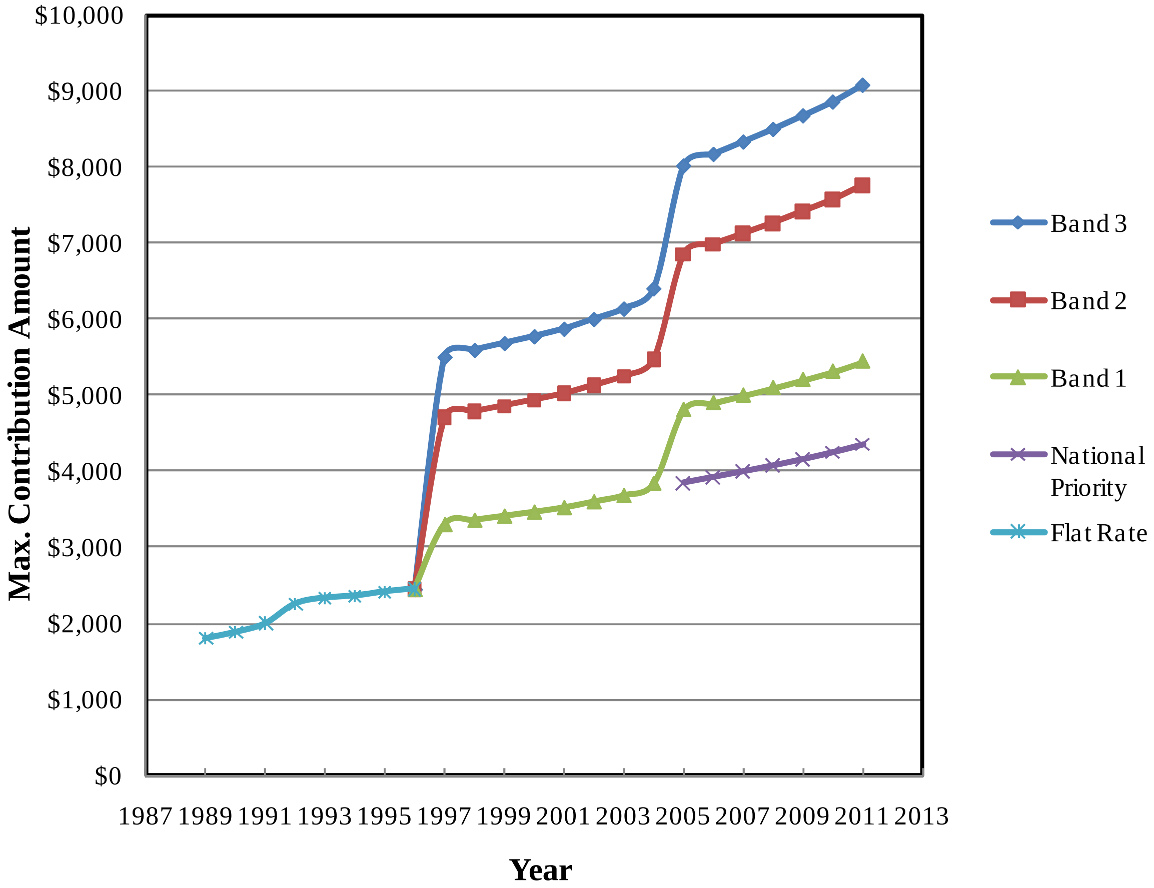

Figure 2.1: Student Contributions Over Time
Sources: DEEWR (2004-2011), DSET (2008), Jackson (2003)

---

[3] *Table A1.1* in *Appendix A1* lists all courses and their band classifications since 1997.
[4] *Table A1.2* in *Appendix A1* details the proportions of HECS-HELP, and FEE-HELP students since 2005.

pays for their entire degree is fixed according to the relevant amount from their commencing year. *Figure A1.1* in *Appendix A1* illustrates the trends in commencing student enrolments by subject area since 2003. Course fees do not appear to be correlated with the level of new enrolments. For example costly business related degrees are the second most popular for commencing students, while enrolments in the low cost national priority courses, nursing and education, are not as substantial. More crucially, there does not appear to be to any shift in commencing student enrolments into particular degree areas in reaction to the 2004-05 policy changes.

The 2004-05 policy reforms also saw a change in the debt repayment thresholds. In 1997, the repayment thresholds had been lowered, so that individuals with lower incomes were required to repay higher amounts, as well as doubling the number of repayment brackets from four to eight. In 2005, two more repayment brackets were added, however thresholds were increased overall.

3. Literature Review

This present study is motivated by Marks' (2009) examination of the impacts of HECS debt on socioeconomic inequality and transition to adulthood outcomes. This paper is the first to test two main criticisms of HECS debt: that HECS debt is implicated in increasing socioeconomic inequalities, and in postponing the transition to adulthood. While his paper is technically flawed, it provides an interesting way of testing the social impacts of HECS debt. From a methodological perspective, Marks' (2009) approach ignores many issues including selection and omitted variable bias. In addition, it is not clear whether Marks (2009) is testing the impact of HECS debt, or the impact of the 1997 HECS policy changes. Furthermore, Marks (2009) does not consider the 2004-05 HECS policy reforms, which may be a confounding factor since he is using wave 6 of the HILDA survey, taken in 2006.

Marks (2009) seeks to test whether socioeconomic inequalities have increased as a result of HECS debt, using a logit model with the dependent variable as university qualification. This is regressed on year of birth, the measure of socioeconomic status, and an interaction between year of birth and socioeconomic status. The two socioeconomic status measures used are the ANU4[5] and "a combination of occupational status with the average of father's and mother's highest level of education" (p.77). Marks (2009) does not present a rationale for these measures, nor why he ignored the readily available indicators of socioeconomic status in the HILDA survey. As HECS debt is not included as an explanatory variable and socioeconomic status is not the dependent variable, Marks (2009) does not test whether having a HECS debt impacts socioeconomic inequality (his intention). He merely tests whether his measures of socioeconomic status affect the likelihood of obtaining university qualifications. Furthermore, he does not test distributions, and thus cannot assess whether socioeconomic inequality is worsening for individuals with HECS debt. Thus he inaccurately concludes that HECS debt does not worsen socioeconomic inequality.

Birch and Miller (2006) recognise that there are only a few studies which quantify the impact of HECS debt on socioeconomic inequality. They find, using a comparison of means, that many students from high socioeconomic backgrounds pay their HECS upfront and thus do not incur a HECS debt. Meanwhile, students from low and middle socioeconomic backgrounds are less likely to pay upfront. Birch and Miller (2006) use students' postcodes obtained by Australian universities

[5] ANU4 is a social research scale which ranks occupational status scores.

to construct a socioeconomic background measure. They use a postcode hierarchy with low socioeconomic status defined in the Commonwealth Equity Scheme as the lowest 25% of postcodes. 20 groups are formed from the ranked postcodes and upfront contributions are controlled for.

Birch and Miller (2006) find that when comparing the mean HECS debt levels between rich and poor neighbourhoods, students from low socioeconomic backgrounds defer larger proportions of their HECS liability; hence they incur greater debts. Students from the two poorest neighbourhood groups deferred 84% of their HECS liability on average, compared with 60% for students from the two richest groups. These results show that the burden of HECS debt falls on students from low socioeconomic backgrounds, which contradicts Marks' (2009) conclusion. However, Birch and Miller (2006) are unclear about which year/s they include and do not examine how socioeconomic inequality has changed in response to HECS policy reforms. Birch and Miller (2006) only plot a single cross section from an unknown time period, and though informative, we do not know how this affects students' current or future decisions and outcomes. From a theoretical perspective, we can imagine that HECS debt may affect students' course selections, which in turn may affect earnings, and may also impact home buying and child bearing decisions. Marks (2009) terms these decisions 'transition to adulthood'.

Marks (2009) finds that HECS debt does not statistically impact marriage decisions, leaving the parental home or buying a home, but does reduce the likelihood of having a first child. However, omitted variables and self selection pervade this analysis. Important covariates are omitted. For instance, relationship status is not considered in the marriage regression, and the unobserved selection into degrees is ignored. As such, we should be cautious about inferring causal effects. Furthermore, his sample includes all individuals with university qualifications. So his analysis also captures both eligible and non eligible tertiary students who did not incur any HECS loans. This underestimates the mean level of HECS debt and may contaminate estimates.

Yu, Kippen and Chapman (2007) stress the importance of a valid counterfactual to measure the impact of HECS debt on fertility decisions but it is unclear why they focus on counterfactuals outside of an experimental setting. Since they do not disclose their estimation procedures, we assume they implemented ordinary least squares. Yu et al. (2007) acknowledge an identification problem of separating the effect of HECS debt on fertility from the effect of education on fertility. However, they do not attempt an instrumental variable approach. Using wave 2 (year 2002) of the HILDA survey, they assume that every student who completed a degree after 1989 had/has a HECS debt. This is problematic because the year of degree completion cannot be inferred from the HILDA survey. Yu et al. (2007) conclude that neither having a HECS debt nor the size of the HECS debt have an impact on expected lifetime fertility. This differs from Marks' (2009) analysis of the decision to become a parent. As in the case of Marks (2009), a causal effect cannot be inferred due to a possible simultaneity bias. HECS debt may affect fertility outcomes, though the decision to have a child may also influence individuals' decisions to take HECS loans or not.

The discussed literature does not consider important aspects of the HECS policy changes, which may have potential implications for selection into degree types. Problems of omitted variables, confused methodologies and selection issues imply that causal effects cannot be inferred. We can use more sophisticated techniques to infer causal effects and to capture important elements of the dataset.

4. Summative Analysis

4.1. Data

The Household, Income and Labour Dynamics in Australia (HILDA) is a longitudinal survey, representative of Australian households. The first wave commenced in 2001, and comprised a sample of 19,914 individuals interviewed from 7,682 households. Presently, there are nine waves of data, which have been collected through face-to-face interviews, and in some instances through telephone interviews (Watson, 2010).

The reference population consists of all household members within a private dwelling as of wave 1. The sample is selected at a household level, defined as 'a group of people who usually reside and eat together', according to the ABS (Watson, p.98, 2010). Shifts in the composition of households are tracked over time. Thus, all household members interviewed in wave 1 are subsequently followed in later waves, as are new children born or adopted in any given year. Since we are using cross sections in this study, we do not follow the same respondents over time; hence attrition is not a major concern.

The HILDA survey provides information on key variables for our research, including HECS debt, transition to adulthood outcomes, and a range of socioeconomic measures. Nevertheless, the lack of detailed information on education, such as degree type and starting/completion year of education, can bias estimates and makes it difficult to construct treatment and control groups. The International Social Survey Programme (ISSP) and the Australian Census are alternative sources of cross-sectional data, which contain specific information on education. The Australian Census provides key data on starting/completion year of education (ABS, 2010). Additionally, in the ISSP, respondents are queried on whether the government should spend more or less on education (GESIS, 2010). However, both surveys lack information on HECS debt, crucial to our analysis, hence the appropriateness of using the HILDA survey.

4.2. Sample

For precision, analyses were conducted separately for individuals of all age groups, and those aged 20-40 years. However, no sizeable differences were observed between each group. Thus, our sample focuses on individuals aged 20-40 years who have completed a bachelor's (Honours inclusive) degree[6].

4.3. Outcome variables

As discussed in *section 3,* Marks (2009) considered two types of outcomes: socioeconomic inequality and transition to adulthood. The socioeconomic measures used in Marks (2009) are difficult to interpret. However, the HILDA survey provides three readily available measures of socioeconomic outcomes: decile of index of relative socioeconomic advantage/disadvantage, decile of index of economic resources, decile of index of education and occupation (Appendix A.2). The robustness of key results may be gauged by comparing obtained estimates or means across these three socioeconomic measures.

[6] In the remainder of this paper, the completion of a 'degree' terminology refers to the completion of a bachelor's (Honours inclusive) degree.

The transition to adulthood encompasses four outcomes: leaving the parental home, getting married, parenthood (having a first child), and (first) home ownership. For ease of discussion, this study will solely focus on the possible effects of a HECS debt on the decile of index of relative socioeconomic advantage/disadvantage and home ownership outcomes.

4.4. Key explanatory variables

An extensive set of characteristics is available in the HILDA survey, including HECS debt information. A note of caution is warranted before pursuing any analysis. Firstly, data on HECS debt is only available for waves 2 and 6, which restricts our study to cross-sections. Yet, having data on HECS debt two years before and two years after the 2004-05 HECS changes provides clear-cut time periods in looking at the HECS debt implications on the selected outcomes. Secondly, respondents are asked whether they have 'any outstanding student loans or debts (including any HECS or FEE-HELP debts)', though excluding any housing or business loans (HILDA, p. 41, 2010). This measure may be problematic, as we may get a diluted effect, since after 2005 this measure would also include FEE-HELP debts. However as discussed in *section 2*, these individuals only accounted for approximately 10% of students in any year, and so we assume that they will not significantly impact our analysis.

In investigating socioeconomic and home ownership outcomes, a range of control variables is considered. Personal characteristics, such as age, ethnicity, number of dependent children and marital status are likely to affect these outcomes. Similarly, yearly individual disposable income and other debts are expected to explain a large portion of the variation in both socioeconomic and home ownership outcomes. Indeed, we expect wealthier individuals to be able to afford a house to a larger extent than their less wealthy counterparts, on average. Finally, it is intuitive to anticipate employment status to play a part in individuals' decision to buy a house, and to explain some of the variation in socioeconomic inequalities (Appendix A.2).

4.5. Descriptive analysis

Table 4.1 displays descriptive statistics on the amount of HECS debt for individuals aged 20-40 years, and who have completed degrees. In our sample, the mean HECS debt amounts to $10,487 and $13,777 per individual for waves 2 and 6, respectively. This disparity across the two waves becomes more apparent when looking at percentiles. Indeed, half of individuals in our sample have a HECS debt of $8,500 in wave 2 and $12,000 in wave 6. Our observed mean HECS debt figure in 2002 matches the amount reported by the Australian Bureau of Statistics (2005), which Marks (2009) cited but could not match in his analysis. In addition, 10% of individuals have a HECS debt of $20,000 in wave 2, relative to $28,000 in wave 6. Additionally, gender differences are observed, with higher HECS debts for males than females, on average.

Table 4.1: Individuals aged 20-40 years with a HECS debt - completed a bachelor's (Honours inclusive) degree						
	Wave 2			Wave 6		
	Pooled obs.	Males	Females	Pooled obs.	Males	Females
Number of obs.	319	118	201	323	112	211
Mean debt ($)	10,487	10,701	10,361	13,777	15,070	13,091
Median debt ($)	8,500	9,000	8,000	12,000	14,000	12,000
75th Percentile ($)	13,000	13,000	12,000	18,000	20,000	17,000
90th Percentile ($)	20,000	20,000	19,000	28,000	29,000	26,000
95 Percentile ($)	26,000	29,000	25,000	32,000	35,000	30,000

Figures 4.1 and 4.2 exhibit a scatter plot of HECS debt by individual age for waves 2 and 6, respectively. Again, the sample is restricted to individuals aged 20-40 years, who have completed degrees. Despite a relatively uneven HECS debt distribution across age, we observe a peak mean HECS debt of $18,804 per individual at age 22 in wave 2. This compares with a peak mean HECS debt of $19,025 at age 24 in wave 6. Isolating potentially influential observations does not affect these figures[7].

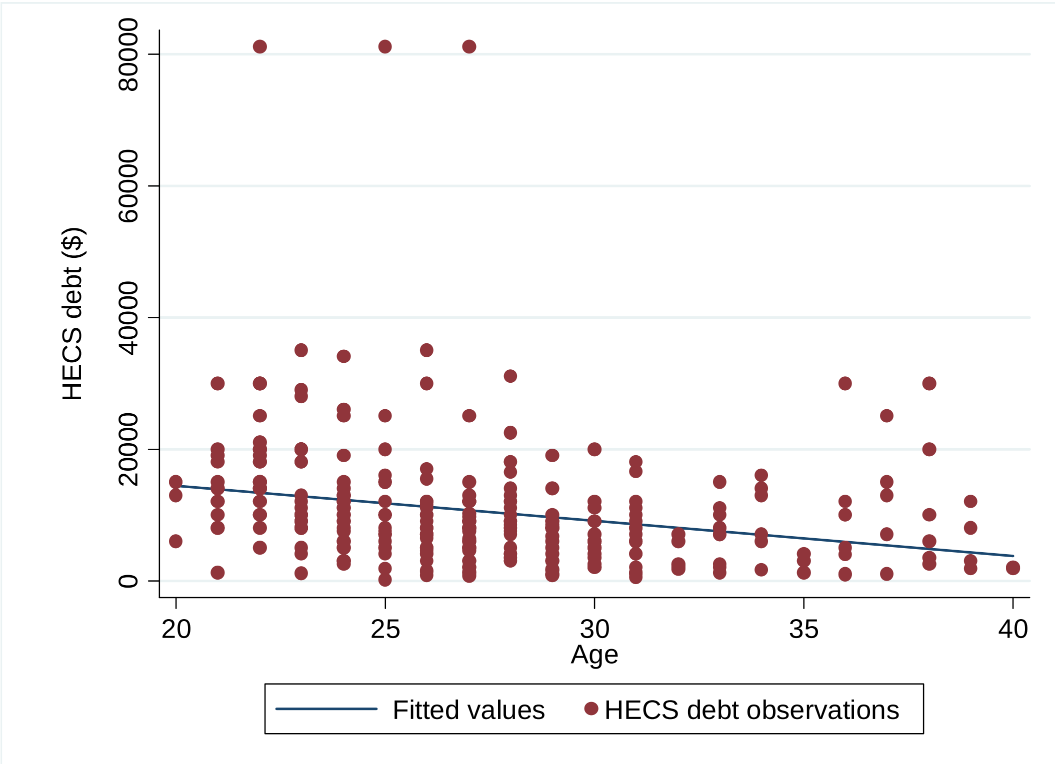

Figure 4.1: HECS debt distribution by age in 2002
Individuals aged 20-40 years

[7] Separate sets of descriptive statistics were run for both the complete HECS debt range, as well as observations with a HECS debt below $60,000.

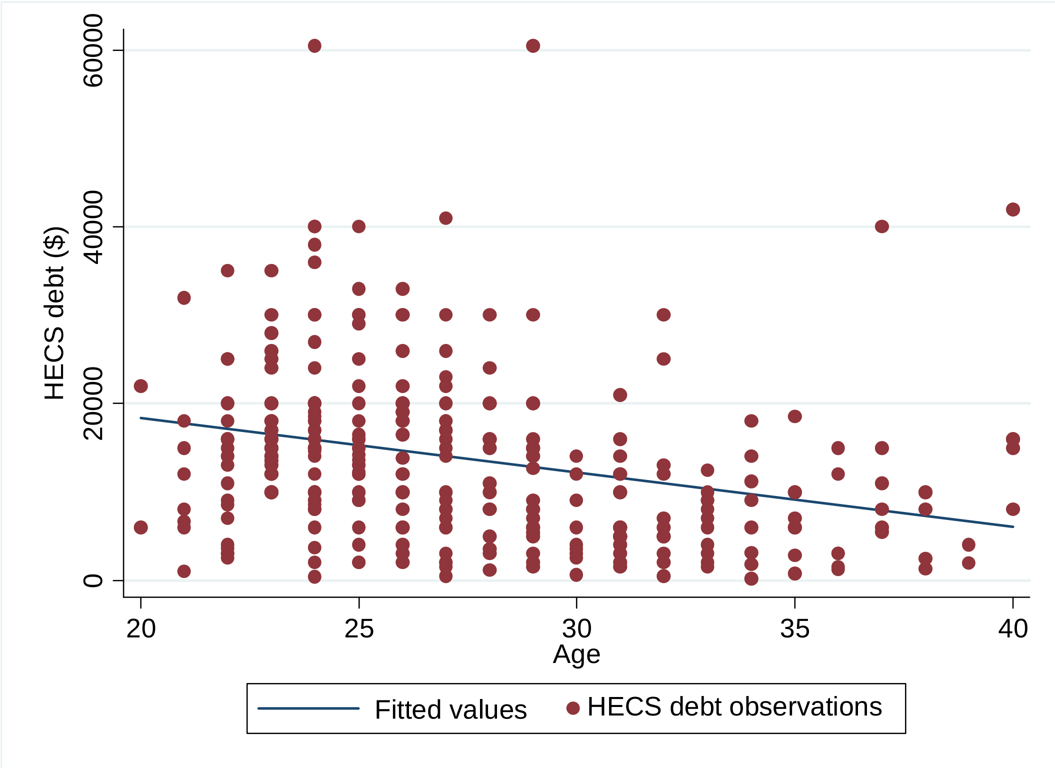

Figure 4.2: HECS debt distribution by age in 2006
Individuals aged 20-40 years

As displayed in *figures 4.3 and 4.4,* half of the individuals who own houses have an approximate HECS debt of $10,200 in wave 2, relative to a mean HECS debt of $12,000 in wave 6. One in two individuals who do not own a house have a mean HECS debt of $10,600 in wave 2, compared to roughly $15,300 in wave 6. These figures concur with our expectations, as having a HECS debt would disincentivise home ownership. Individuals owning a house appear to have a lower HECS debt than those who do not own a house. Notably, we observe greater disparity in mean HECS debts across those who own and do not own a house in 2006, when compared to the 2002 figures.

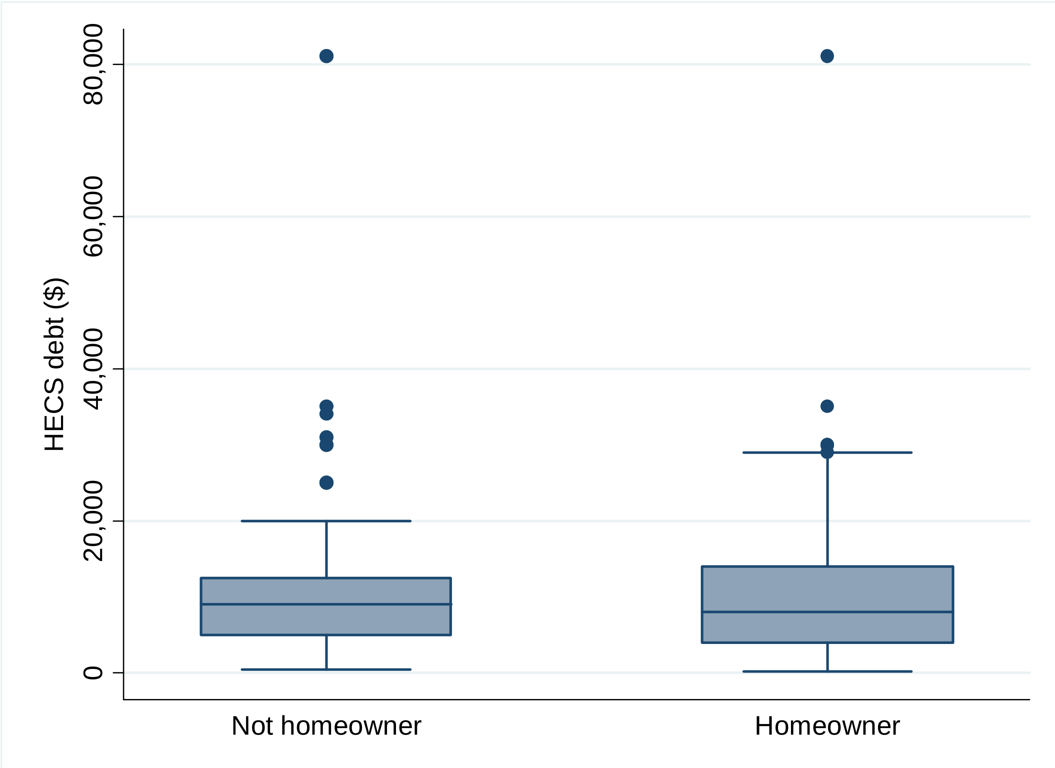

Figure 4.3: Box plot – HECS debt by home ownership status in 2002
Individuals aged 20-40 years

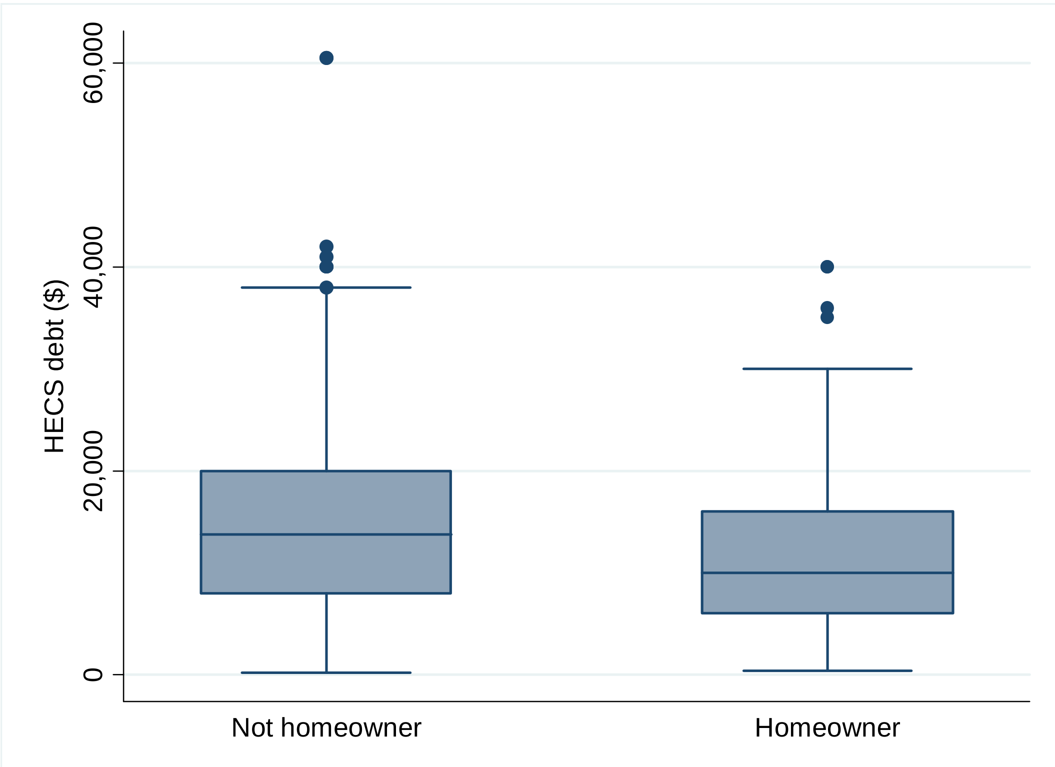

Figure 4.4: Box plot – HECS debt by home ownership status in 2006
Individuals aged 20-40 years

Figures 4.5 and 4.6 present box plots of individuals' HECS debts by an index of decile of socioeconomic advantage/disadvantage in 2002 and 2006, respectively. In regards to the socioeconomic outcome measure, [1] refers to the lowest socioeconomic outcome and [10] refers to the highest socioeconomic outcome. We would expect individuals from a lower socioeconomic background to have a higher HECS debt, as having a lower purchasing power than individuals from a higher socioeconomic background. While the descriptive statistics from 2002 (figure 4.5) concur with our expectations; those from 2006 (figure 4.6) do not reveal a clear relationship. As previously noted, we observe a greater level of disparity in HECS debt levels across the range of socioeconomic deciles in 2006.

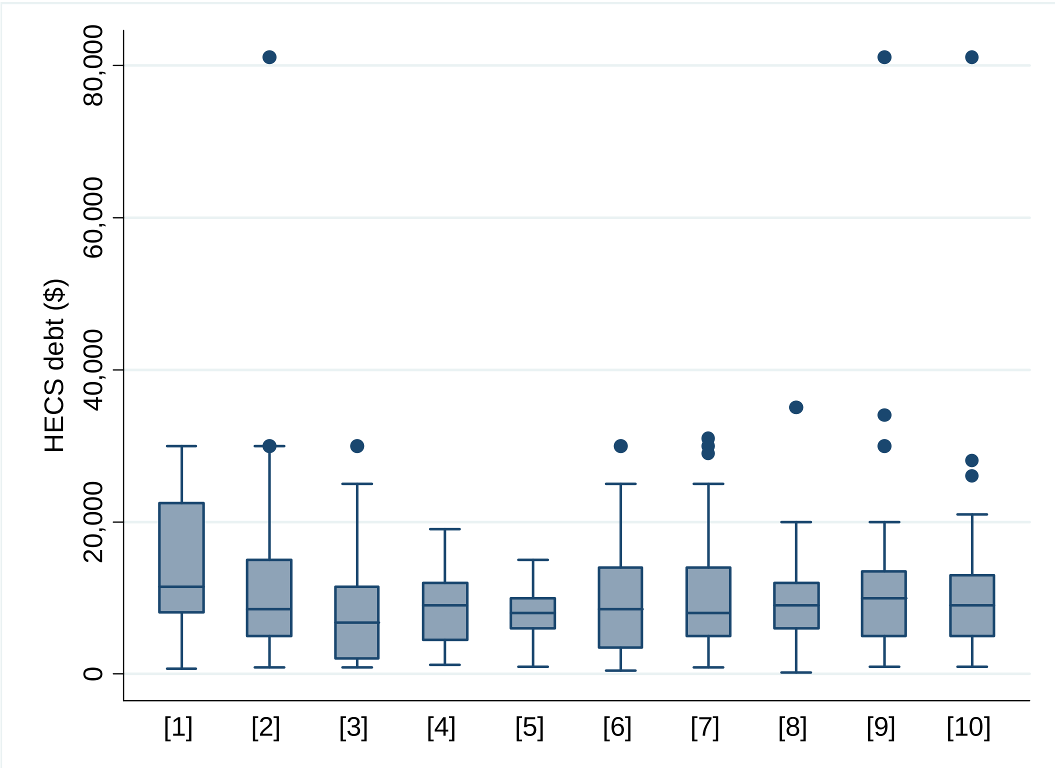

Figure 4.5: Box plot – HECS debt by index of relative socioeconomic advantage/disadvantage in 2002 - Individuals aged 20-40 years

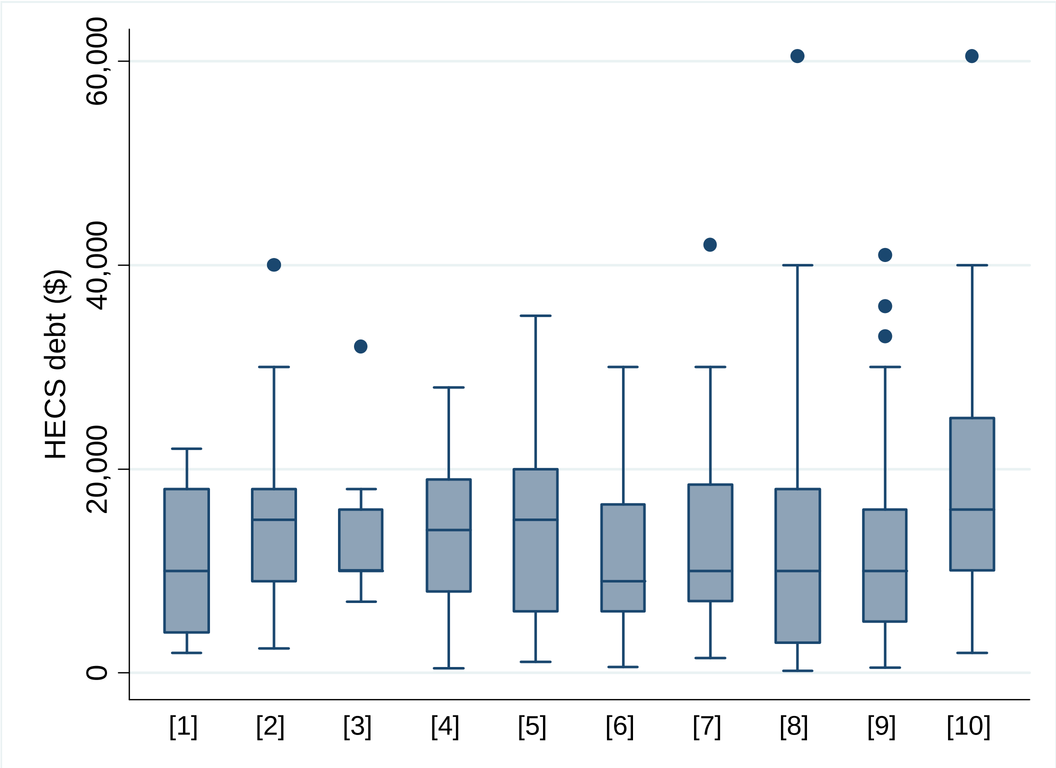

Figure 4.6: Box plot – HECS debt by index of relative socioeconomic advantage/disadvantage in 2006 - Individuals aged 20-40 years

5. Empirical approach

5.1. Primary analysis

Considering the 2004-05 HECS policy change as our treatment, we have both pre and post treatment information. This allows for a wealth of empirical techniques, provided we have valid treatment and control groups. The 2002 and 2006 HECS debt variables will be integral to the construction of treatment and control groups.

For the purpose of this study, a difference-in-differences (DID) approach is meaningful, since our regressor of interest (HECS debt) changes across government supported students[8]. A DID approach requires pre and post treatment information about the control and treatment groups. DID allows us to remove time-specific and individual fixed effects, which are present. *Table 5.1* summarises our treatment and control groups:

[8] While we are analysing the 2004-05 HECS policy changes, which constitute a discontinuity in treatment, we are unable to use a regression discontinuity design. Indeed, we cannot attribute the differences in treatment and control outcomes solely to the HECS policy changes. This is because of the existence of confounding factors, which are likely to impact outcomes. For instance, the Baby Bonus will influence child bearing decisions and the First Home Owner's Grant may affect the decision to buy a house and may increase socioeconomic inequality.

Table: 5.1: Constitution of treatment and control groups for 2002 and 2006

	Treatment	**Control**
2002 (pre)	Consists of individuals who have HECS debts and have completed their bachelor's degree, commencing between 1997 and 2004[9]. $E[Y_{oit} \vert H_i = 1, P(X_i)]$	Consists of individuals who have completed their bachelor's degree, do not have HECS debts, and who commenced anytime between 1997 and 2004. $E[Y_{oit} \vert H_i = 0, P(X_i)]$
2006 (post)	Consists of individuals who have HECS debts and have completed their bachelor's degree, but commenced from 2005 onwards. $E[Y_{1it+1} \vert H_i = 1, P(X_i)]$	Consists of individuals who have completed their bachelor's degree, do not have HECS debts, and who commenced from 2005 onwards. $E[Y_{0it+1} \vert H_i = 0, P(X_i)]$

As the 2004-05 HECS policy changes do not affect students enrolled before 2005, any students enrolled before provide valid pre-treatment control and treatment groups.

The construction of the pre-treatment groups is predicated on an assumption: due to the nature of the HECS policy, once a student commences their study, the cost of their degree remains in the same band regardless of changes to bands. For instance, nursing students who commenced their degree in 2003 would be charged according to band 1 fees, even though nursing was placed in the priorities tier in 2005 (Table A1.1). As we have no data on degree starting/completion dates, assuming the average student completes a three-year degree, the post-treatment group should consist of students that commenced post 2005 and finished from 2008 onwards. Information on HECS debt is only provided in waves 2 and 6, which prevents any immediate empirical analysis due to the lack of data on post-treatment treatment and control groups. Consequently, in this study we will conduct a hypothetical analysis.

In order to execute DID, we must be willing to assume that the pre-treatment control and treatment groups have similar trends across time in terms of their observed characteristics. For instance, a preliminary analysis of mean individual disposable income between pre-treatment control and treatment groups reveals sizeable differences. In wave 2, individuals without HECS debts have, on average, 42% higher disposable incomes than individuals with HECS debts. In wave 6, disposable income disparities between individuals with HECS debts and individuals without HECS debts amount to 39%, on average. In relation to potential outcome variables, we also find a greater number of married individuals without HECS debts than those married with HECS debts, on average. As such, we suggest a propensity score matching (PSM) approach before undertaking DID.

To estimate the policy effect, it is essential that the control and treatment groups are similar prior to the treatment. If this is the case, we can attribute differences in the mean outcomes for the treatment group post-treatment to the treatment, assuming that there are no other confounding factors which could affect differences in mean outcomes for post-treatment treatment and control groups. Matching enables us to form control groups which are similar to treatment groups in terms of observables. We have chosen PSM instead of matching by covariates because we have a high

[9] As discussed in *section 2*, before 1997 all domestic students had a status equivalent to HECS presently. Thus, there was no distinction between HECS and non-HECS students prior to 1997.

degree of dimensionality across the covariates. An additional benefit of matching is that it performs better in small samples when compared to regression, since we are able to check for common support more thoroughly. Furthermore, regression can be more cumbersome, as the model would have to be saturated in the covariates to ensure that we are not extrapolating across cells without both treatment and control units. Our sample size is relatively small, thus PSM is advantageous.

PSM is predicated on two key assumptions: unconfoundedness and overlap. Since we are using PSM for the purpose of minimizing possible selection bias in the DID estimates, we do not need to invoke the stronger 'levels' unconfoundedness assumption. We must invoke a weaker unconfoundedness assumption, which is based on changes in levels over time (Angrist and Pischke, 2009).

> **Unconfoundedness:** Conditional on propensity scores, the change in mean outcomes over time for the control group untreated equals the change in the mean outcomes over time in the treatment group 'had they not been treated'. This assumption follows naturally from the Strong Ignorability conditions. Rosembaum & Rubenstein (1983) show that we can condition on propensity scores *P(X$_i$)* instead of a vector of covariates X_i.

$$E[Y_{0it+1}|H_i = 1, P(X_i)] - E[Y_{0it}|H_i = 1, P(X_i)] \qquad [5.1]$$

$$= E[Y_{0it+1}|H_i = 0, P(X_i)] - E[Y_{0it}|H_i = 0, P(X_i)]$$

> Unconfoundedness asserts that conditional on observables, selection bias is removed (Angrist and Pischke, 2009). In this study the propensity score $\mathrm{P}(X_i)$ reflects an indidivual's probability of having a HECS debt given a set of observed characteristics. The dummy variable H_i equals one if an individual has a HECS debt and zero otherwise.

In *Appendix A2,* we present a list of covariates which are likely to explain some differences in the time trends between the untreated control and treatment groups. For instance, factors such as yearly individual disposable income, labour force status and health conditions are controlled for. However, important explanatory variables are not available in the HILDA survey. Parents' incomes can be correlated with one's HECS status (or FEE-HELP status), as well as outcomes such as socioeconomic status, and possibly transition to adulthood. The HILDA survey provides data on parents' occupations, which could potentially proxy for parents' incomes. Furthermore, the lack of information on university entry marks can be problematic as it may explain selection into HECS positions. [10] These unobserved marks are fixed across time and thus would be differenced out through the DID approach. Since we have potential solutions to these issues it seems satisfactory to invoke unconfoundedness.

> **Overlap:** For each value of the covariates, $X_i = x$, there are both treatment and control units.

$$0 < \Pr(H_i = 1|X_i = x) < 1 \quad \forall\, x \qquad [5.2]$$

[10] To obtain a FEE-HELP place, students usually require lower entry marks than for the corresponding HECS-HELP place.

For any particular value of the covariates, if the probability of treatment is zero, there are no treatment units at that value. If the probability of treatment is one, we only have treatment units at that value. This is important because we want to construct a valid counterfactual group. For each treatment observation, we ascribe a counterfactual which consists of a control observation with similar observed characteristics.

The interested reader may refer to *Appendix A.3* for a detailed discussion of the steps involved in PSM.

Once the matching conditions are met, we can assume that the trend over time in mean outcomes for the untreated control group is the same as the counterfactual time trend in mean outcomes for the treatment group. By counterfactual time trend we mean the time trend we would have observed had the treated group remained untreated. This can be demonstrated graphically:

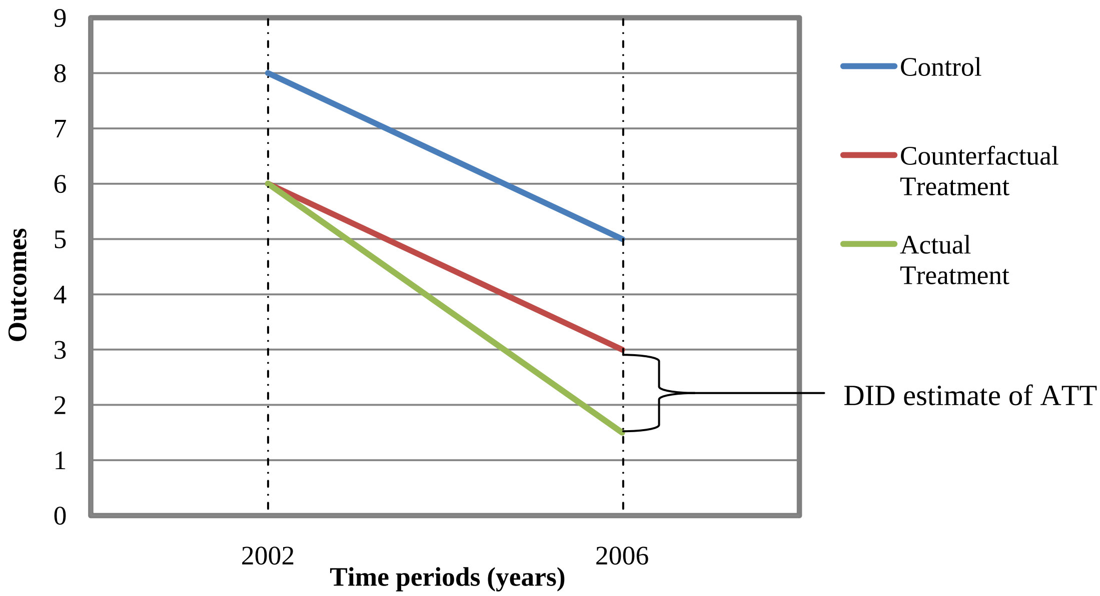

Figure 5.1: DID Approach

The direction of the trend depends on the outcome measure. We expect the treated treatment group to be less likely to have a first child, leave the parental home and less likely to be in a higher socioeconomic level. Thus, we expect a downward trend in mean outcomes over time.

$$\delta = \frac{1}{N}\sum_{i \in H_1}\left[\left(Y_i - \sum_{j \in H_0} w_{ij}^{H} Y_j\right) - \left(\sum_{j \in C_1} w_{ij}^{C} Y_j - \sum_{j \in C_0} w_{ij}^{C} Y_j\right)\right] \qquad [5.3]$$

Subscripts i and j refer to treatment and matched counterfactual observations, respectively. The weights are denoted by w, H_k refers to treatment group and C_k to the control group; where k=*0,1* indicating pre or post treatment period, respectively. Following Doiron (2004), the groups are weighted and matched using a propensity score estimator before using DID. We would account for observable differences between pre-treatment treatment and control groups through the use of PSM, and also remove time-invariant unobservables via DID. *Equation 5.3* expresses the difference in two differences. Firstly we would take the difference in the outcomes for the post-treatment treatment group (Y_i) and the matched pre-treatment treatment group ($w_{ij}^{H} Y_j$). Secondly, we would take the difference in the outcomes for the post-treatment control group and the matched pre-treatment control group. The DID estimator of the ATT (δ) is then obtained by taking a difference of those two differences, and dividing it by the number of observations *N*.

6. Conclusion

We have acknowledged the difficulties in isolating the impact of the 2004-05 HECS policy changes due to possible omitted variables and selection bias. We have been more careful than existing literature to consider the issues involved in this type of study. Even so, we have discussed how we could potentially perform a DID with weighted PSM counterfactuals, if more complete data, as well as a longer time span were available. Understanding the social impacts of HECS debt is critical since this can be a major debt burden for many students, which may colour their future decisions. This paper has demonstrated the caution required in inferring causal effects correctly from data in this area. Conducting meaningful empirical research is crucial for the design of more effective policies, going forward. Hopefully this research will help to guide future econometric endeavours.

Appendices

A.1. Institutional Framework Data

Table A1.1 presents information on the federal government's band number classifications for all subject areas in each year since the tiered system was introduced in 1997. These would determine the student contribution rate that HECS students would be subject to for their entire degree, if commencing in that year. Changes to the classification band for subject areas have been highlighted. Groupings did not change between 1997 and 2004, but following the 2004-05 HECS policy changes and the introduction of the four tiered system, some subject areas have changed bands.

Table A1.2 shows the trends in domestic student enrolments since the 2004-05 policy changes according to payment/loan type. The number of domestic students in each category is presented as a percentage of total domestic students in that year. The data includes both continuing and commencing domestic students in that year. The majority of domestic students have been those with HECS-HELP loans, and this trend has remained consistent over time.

Table A1.1: Student contribution bands over time

	2011	2010	2009	2008	2007	2006	2005	2004 - 1997
Accounting	3	3	3	**3**	**2**	2	2	2
Administration	3	3	3	**3**	**2**	2	2	2
Agriculture	2	2	2	2	2	2	2	2
Behavioural Science	1	1	1	1	1	1	1	1
Built Environment	2	2	2	2	2	2	2	2
Clinical Psychology	1	1						
Commerce	3	3	3	**3**	**2**	2	2	2
Computing	2	2	2	2	2	2	2	2
Dentistry	3	3	3	3	3	3	3	3
Economics	3	3	3	**3**	**2**	2	2	2
Education	1	**1**	**P**	P	P	P	**P**	**1**
Engineering	2	2	2	2	2	2	2	2
Health	2	2	2	2	2	2	2	2
Humanities	1	1	1	1	1	1	1	1
Languages	1	1	1	1	1	1	1	
Law	3	3	3	3	3	3	3	3
Maths	P	P	**P**	**2**	2	2	2	2
Medicine	3	3	3	3	3	3	3	3
Nursing	1	**1**	**P**	P	P	P	**P**	**1**
Science	P	P	**P**	**2**	2	2	2	2
Social Studies	1	1	1	1	1	1	1	1
Statistics	P	P	**P**	**2**	2	2	2	
Surveying	2	2	2	2	2	2	2	
Vet Science	3	3	3	3	3	3	3	3
Visual and Performing Arts	1	1	1	1	1	1	1	1

Sources: DEEWR (2004-2011), DSET (2008), Jackson (2003)

Key: 1 = Band 1, 2 = Band 2, 3 = Band 3, P = National Priority Band

Table A1.2: Domestic student enrolments by payment type

	2008	2007	2006	2005
HECS-HELP	73.36%	73.18%	73.56%	74.08%
FEE-HELP	10.60%	10.09%	9.37%	8.28%
Domestic Full Fee	10.64%	11.38%	11.28%	11.74%
Total domestic	770,814	755,349	732,214	716,589

Source: DEEWR (2004 - 2008)

Figure A1.1 plots the trend of commencing student enrolments each year since 2003, by broad subject area. The most popular area has been the relatively low cost humanities subjects, while the most expensive area, medicine, has been the least popular. However this may also be a reflection of the level of difficulty in being admitted into these courses. As discussed in *section 2*, there does not appear to be any particular correlation between the cost of courses and their popularity, and commencing student numbers have also remained relatively steady over time.

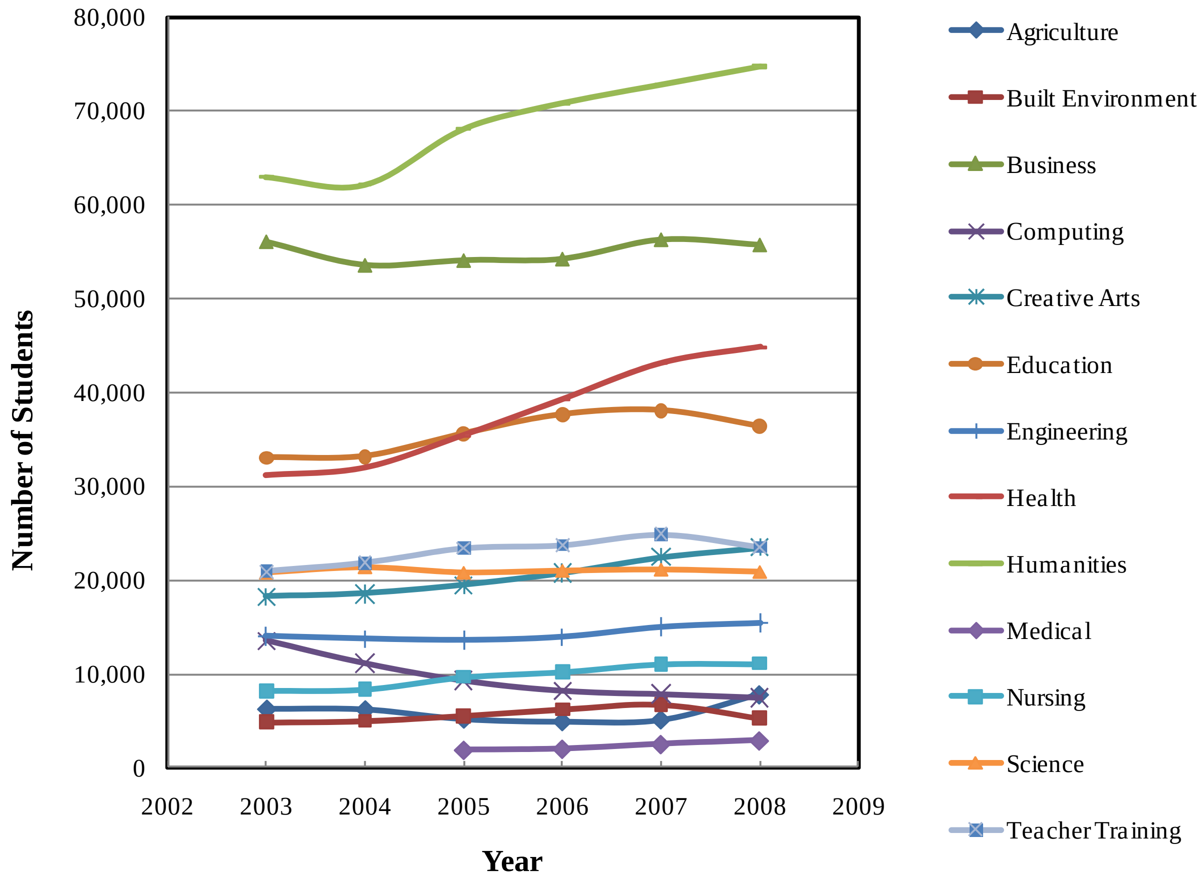

Figure A1.1: Commencing Students Over Time
Source: DEEWR (2004 - 2008)

Finally, *figure A1.2* shows how the repayment rates have changed over time. Each line represents the percentage of the debt that an individual was required to repay in each financial year, based on their income. Where the taxable income falls between two lines, that individual would be subject to the repayment percentage of the uppermost line. For any income level below the 0% line in a year, that individual was not required to make any compulsory repayments, though they could make voluntary repayments. All individuals whose income level was above the top line in a year would be subject to the highest percentage repayment.

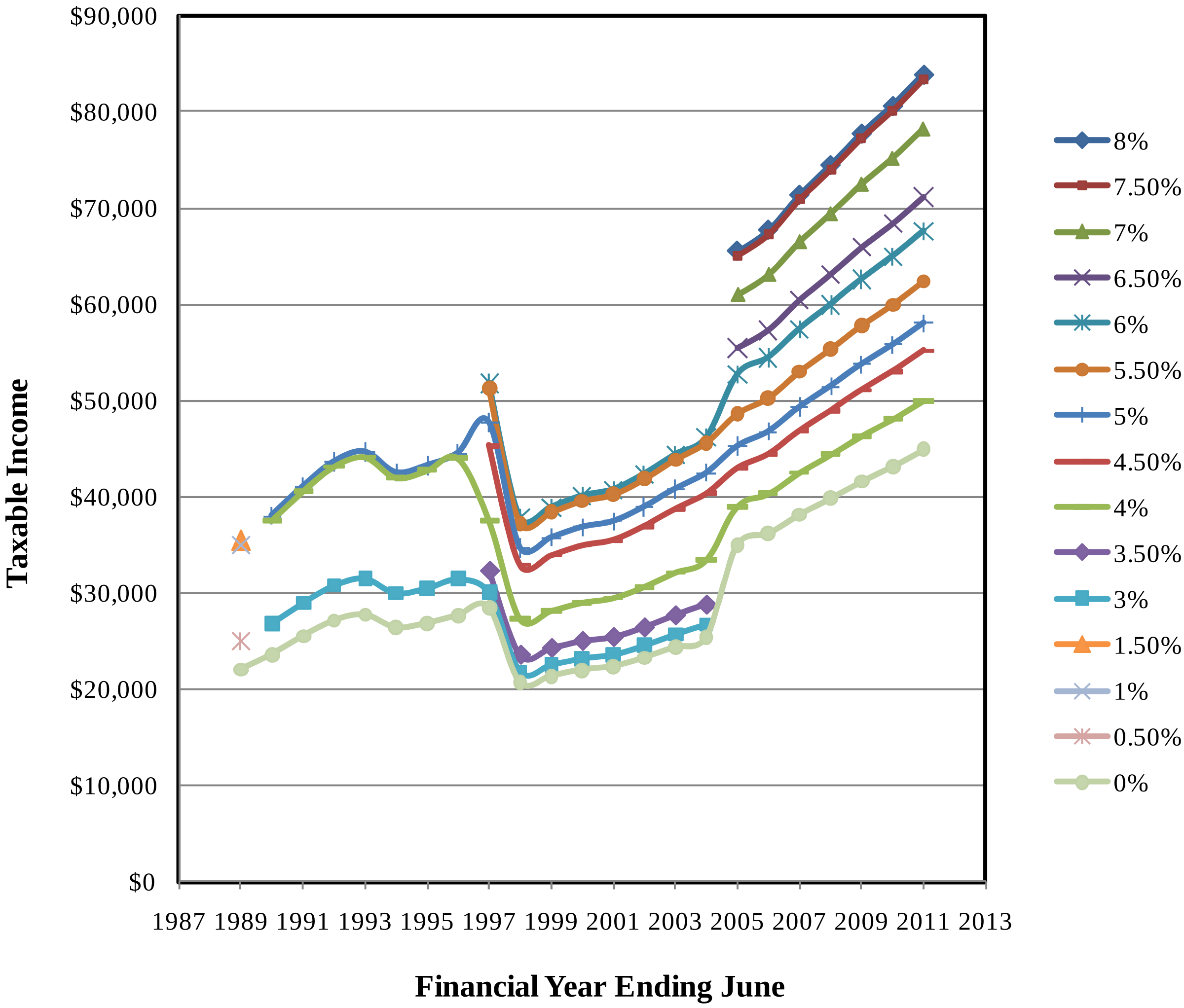

Figure A1.2: Repayment Rates (% of debt) Over Time
Source: ATO (2010)

A.2. Data

Table A2.1 displays a suggested list of control variables for each corresponding outcome variable. In regards to our outcome variables, the original variable names from the HILDA survey are provided in brackets. This list of variables is applicable to a study with waves 2 and 6 of the HILDA survey.

Table A2.1: Covariates

		Transition to Adulthood Outcomes				Socioeconomic Outcomes		
Variable	Description	Leaving the parental home (*fmageln*)	Getting married (*mrcurr*)	Having a (first) child (*tchad*, *icprob*)	Buying a (first) home (*hstenur, hstenr*)	Relative socio-economic advantage/disadvantage (*hhda10*)	Economic resources (*hhec10*)	Education and occupation (*hhed10*)
Aboriginal or Torres Strait Islander	Indicates if an individual is of Aboriginal or Torres Strait Islander descent	x	x	x	x	x	x	x
Age	Individual's age in years	x	x	x	x	x	x	x
Arrived in Australia	Indicates if an individual arrived in Australia within the past 10 years	x	x	x	x	x	x	x
Birth Order	Indicates whether the individual is the oldest child, or not	x			x			
Carer status	Whether the individual is currently a carer for someone	x	x	x	x	x	x	x
Health condition	Whether the individual has any serious/ongoing health issues	x	x	x	x	x	x	x
HECS debt	Indicates whether an individual has an outstanding student loan or not	x	x	x	x	x	x	x
Hours worked	Hours per week worked in main occupation	x	x	x	x	x	x	x
Income	Individual disposable income for the financial year ($)	x	x	x	x	x	x	x
Labour force status	Currently employed, unemployed, or not in labour force	x	x	x	x	x	x	x
Marital staus	Never married, married, defacto, separated, widowed or divorced	x	x	x	x	x	x	x
Number of children	Number of dependent children (including partners' children) aged between 0 and 24 years	x	x		x	x	x	x
Other debt	Car loans, investment loans, personal loans, hire purchase, overdue bills	x	x	x	x	x	x	x
Other sources of income	Superannuation, child support, workers compensation, redundancy, inheritance, allowance from parents	x	x	x	x	x	x	x
Size of HECS debt	Amount currently owed on student loans ($)	x	x	x	x	x	x	x
Where children live	Whether the individual's dependent children currenly reside in their house or elsewhere	x	x		x	x	x	x

Source: HILDA

A.3. Propensity score matching (PSM) procedure

The PSM procedure is flexible. Researchers have discretion in what probability models they choose and in the weighting method they use. Choice is influenced by the characteristics of the data. Since this is a hypothetical study we will outline various techniques which could be used.

1. Establish common support and remove observations which lie outside the common support in the covariates vector.

2. Estimate propensity scores using a probability model. Logit and probit models are most commonly used, and usually yield similar results. Choose covariates which reduce selection (variables which influence the participation decision and the outcome) and omitted variable bias.

3. Match units for the purpose of establishing a counterfactual observation for each treatment observation. The counterfactual used is a control observation with a similar propensity score to that of the treatment observation. A weighting function is required to match units. There are many weighting functions, including nearest neighbor, kernel, caliper and stratification. The function used depends on characteristics of the data and the number of ‘good’ matches that can be performed. For instance, nearest neighbor matches a treatment observation with a control observation which has the most similar propensity score. Control observations can be replaced or removed from the pool of potential matches once they have been matched. Deciding whether to replace or remove matched control observations involves a tradeoff between precision and bias. If we do not have many control observations with similar propensity scores to treatment observations, we may want to replace matched control observations. This is because we want close matches to minimize the bias and reduce the number of bad matches. This will reduce precision because a single control observation will be matched many times. An alternative approach is to use a few weighting functions to check the robustness of results. Even so, Dehejia & Wahba (2002) find that various weighting functions result in similar matches.

4. Estimate the Average Treatment Effect on the Treated (ATT).

A.4. Extensions

A possible extension relates to correcting for an unobserved selection bias. Self-selection into degree type is correlated with HECS debt because of the tiered system. Also, degree type may impact both socioeconomic and transition to adulthood outcomes since there may be systematic differences between individuals who choose different degrees. For instance, individuals from high socioeconomic backgrounds may choose degrees which entail high incomes to maintain their status. Individuals who are likely to have children may choose degrees which allow them flexibility in the workplace. We have a problem of omitted variable bias, since degree type is unobserved. However, we have a few means at our disposal to treat, or bound the true value.

The 2004-05 HECS policy changes can possibly be used as an instrumental variable, as it provides a transparent, exogenous source of variation in HECS debt (key explanatory variable). To be a valid instrument, the 2004-05 change in HECS must satisfy two assumptions:

> **Instrument relevance**: The policy change needs to be highly correlated with individuals' HECS debt. $Corr(Z_i, H_i) \neq 0$ where Z_i is the 2004-05 HECS policy change and H_i is the assignment to treatment dummy:
>
> $$H_i = \begin{cases} 1 & \text{if individual has a HECS debt} \\ 0 & \text{if individual does not have a HECS debt} \end{cases}$$
>
> As indicated in *section 2,* there does not appear to be a shift in student enrolment in reaction to the 2004-05 HECS policy changes. This supports the use of the policy change as an instrumental variable, as it would be highly correlated with our variable of interest (HECS debt), but it appears to be uncorrelated with the omitted variable, degree type. Instrument relevance can be established via a regression of HECS debt status (our endogenous variable) on the 2004-05 HECS policy change (our instrument) and an exogenous set of covariates.
>
> **Instrument exogeneity**: To satisfy this condition the instrument must not be correlated with the omitted variable. The policy change must not be correlated with degree type. The discussion provided in *section 2* provides evidence that there has not been a systematic change in student enrolments within and between set types of degrees as a result of the 2004-05 changes in HECS fees. In addition, existing literature suggests that HECS does not seem to substantially affect students' decision in their choice of degree (Chapman and Ryan, 2005).

Alternatively, we can use a proxy variable instead of an instrument. A proxy must be correlated with the omitted variable (degree type), uncorrelated with the error term and preferably not correlated (or at most only weakly correlated) with the variable of interest (in this case the HECS debt status). An individual's occupation acts as a valid proxy because it is potentially correlated with degree type. Intuitively, individuals choose particular degrees to be qualified for particular careers, or because they are intrinsically interested in the subject area and wish to pursue it as a career.

It would be ideal for occupation not to be correlated with HECS debt, since this would yield an estimate of HECS debt which is unconfounded by degree type and occupation. If this cannot be guaranteed, it is preferred that the proxy is weakly correlated with the variable of interest. The issue is if the proxy is correlated with the variable of interest we would not be able to gather an unbiased estimate of the impact of HECS debt on our outcome variables. However, we may be able to bound the true value.

As a thought experiment, consider the true relationship between socioeconomic status, HECS debt and degree type.

$$Y_i = \alpha_0 + \delta H_i + \gamma DT_i + e_i$$

where

Y_i	Socioeconomic outcome
α_0	Constant term
H_i	HECS debt dummy
DT_i	Degree type
e_i	Error term

and $corr(e, DT) = corr(e, H) = 0$

Degree type is not observed in the HILDA. As such we estimate:

$$Y_i = \alpha'_0 + \delta' H_i + u_i$$

$$\delta' = \frac{cov(Y, H)}{var\ (H)}$$

$$\delta' = \frac{cov(\alpha_0 + \delta H_i + \gamma DT_i + e_i, H)}{var(H)}$$

$$\delta' = \frac{cov(\alpha_0, H) + \delta var(H) + \gamma cov(DT, H) + cov(e, H)}{var(H)}$$

$$\delta' = \delta + \frac{\gamma cov(DT, H)}{var(H)} = \delta + \gamma\theta$$

The estimated effect of HECS debt on socioeconomic status δ', captures the effect of the omitted variable bias $\gamma\theta$, where θ is the coefficient of a regression of degree type on HECS debt. We would expect a positive coefficient on θ, as a higher level degree would entail a larger HECS debt. Using a measure of relative advantage/disadvantage of socioeconomic status we expect $\gamma > 0$. Since socioeconomic status is an ordinal variable ranging from disadvantage to advantage, an ordered probit model would seem appropriate. We expect that obtaining a degree would increase the likelihood of moving into a higher level of socioeconomic advantage. This indicates that δ' would overestimate the effect of HECS debt on socioeconomic status.

If we use occupation as a proxy, even if it is correlated with HECS debt, we would be able to bound the true value. Assuming occupation to be correlated with degree type:

$$O_i = \pi_0 + \pi_1 H_i + \pi_2 DT_i$$

where O_i represents an individual's occupation, which is a cardinal variable. We assume $corr(O, DT) \neq 0$, $\pi_2 > 0$ and $\pi_1 > 0$. Using occupation as a proxy for degree type and rearranging in terms of DT, we get:

$$DT_i = \frac{O_i - (\pi_0 + \pi_1 H_i)}{\pi_2}$$

Substituting for degree type we estimate:

$$Y_i = \left(\alpha_0 - \gamma \frac{\pi_0}{\pi_2}\right) + \left(\delta - \gamma \frac{\pi_1}{\pi_2}\right) H_i + \frac{\gamma}{\pi_2} O_i + e_i$$

We may still have a biased estimate of HECS debt (H_i), as occupation may be correlated with HECS debt. However we are able to bound the true value of HECS debt using the two biased estimates, because we have made assumptions on the expected sign of these effects.

Based on the previous discussion, the true value of HECS debt would lie between:

$$\left(\delta - \gamma \frac{\pi_1}{\pi_2}\right) < True\ value\ H_i < \delta + \gamma\theta.$$

References

Angrist, J., D., Pischke, J-S. 2009, *Mostly Harmless Econometrics: An Empiricist's Companion,* Princeton University Press.

Australian Bureau of Statistics (ABS). 2010, *Census Data,* accessed 7 November 2010, <http://www.abs.gov.au/websitedbs/d3310114.nsf/home/census+data>.

Australian Bureau of Statistics (ABS). 2005, *Paying for university education,* accessed 7 November 2010, <http://www.abs.gov.au/ausstats/abs@.nsf/Previousproducts/1301.0Feature%20Article192005?opendocument&tabname=Summary&prodno=1301.0&issue=2005&num=&view=>.

Australian Taxation Office (ATO). 2010, *HELP and HECS repayment thresholds and rates*, accessed 6 November 2010, < http://www.ato.gov.au/print.asp?doc=/content/8356.htm>.

Beer, G. & Chapman, B. 2004, 'HECS System Changes: Impact on Students', *Agenda*, vol. 11, no. 2, pp 157 – 174.

Birch, E., A. & Miller, P., W. 2006, 'HECS and HECS-HELP: Equity issues', *Journal of Higher Education Policy and Management,* vol. 28, no. 2, pp. 97-119.

Chapman, B., & Ryan, C. 2005, 'The access implications of income-contingent charges for higher education: Lessons from Australia', *Economics of Education Review*, vol 24, no. 5, pp. 491–512.

Dehejia, R.H. & Wahba, S. 2002, 'Propensity score-matching methods for non-experimental causal studies', *The Review of Economics and Statistics,* vol. 84, no. 1, pp. 151-161.

Department of Education, Employment and Workplace Relations (DEEWR). 2004 – 2011, *Information for Commonwealth Supported Students: HECS-HELP,* accessed 6 November 2010,

(2011)<http://www.goingtouni.gov.au/NR/rdonlyres/06523655-0868-4623-B3E9-BE53220E9AA5/0/2011_HECSHELP_booklet.pdf>.

(2010)<http://www.google.com.au/url?sa=t&source=web&cd=1&sqi=2&ved=0CBYQFjAA&url=http%3A%2F%2Fwww.goingtouni.gov.au%2FNR%2Frdonlyres%2F05FB27A6-1F1B-43CD-9244-D92B6388BDC1%2F0%2FHECSHELP2010PDF.pdf&rct=j&q=information%20for%20commonwealth%20supported%20students%202010&ei=ubDXTJr9EM_CcdiQ8K4L&usg=AFQjCNE4YwApM0Y5KR_0PzgA4fZKP06r4Q&cad=rja>.

(2009)<http://www.google.com.au/url?sa=t&source=web&cd=1&ved=0CBkQFjAA&url=http%3A%2F%2Fwww.goingtouni.gov.au%2FNR%2Frdonlyres%2FD605333E-C121-490E-8810-6DD50B0C1AEC%2F0%2FICSS2009bookletforGTU.pdf&rct=j&q=information%20for%20commonwealth%20supported%20students%202009&ei=x7DXTOW0AYS4vgP55vDsCQ&usg=AFQjCNG0RIi2jD-iRczbvkJeZ0NesR8iAw&cad=rja>.

(2008)<http://www.google.com.au/url?sa=t&source=web&cd=1&ved=0CCsQFjAA&url=http%3A%2F%2Fwww.goingtouni.gov.au%2FNR%2Frdonlyres%2F35A9BDC0-C476-4FC2-9BF6-1A55C510E07B%2F0%2F07_068_HECSHELPiNTERNET_004.pdf&rct=j&q=information%20for%20commonwealth%20supported%20students%202008%20site%3Ahttp%3A%2F%2Fwww.goingtouni.gov.au&ei=JbHXTPP2IJDmvQPupb29CQ&usg=AFQjCNF6n788r1Q_KdhBsExD5rosUiUF2g&cad=rja>.

(2007)<http://www.google.com.au/url?sa=t&source=web&cd=1&ved=0CCoQFjAA&url=http%3A%2F%2Fwww.goingtouni.gov.au%2FNR%2Frdonlyres%2FB18D0E80-6242-4242-973D-2364DCC1ABEC%2F0%2Ficss_booklet_2007.pdf&rct=j&q=information%20for%20commonwealth%20supported%20students%202007%20site%3Ahttp%3A%2F%2Fwww.goingtouni.gov.au&ei=MrHXTNefKY_CvQOq-a21CQ&usg=AFQjCNEzq3HD0BhbISoWPV074rd5z2QJfg&cad=rja>.

(2006)<http://www.google.com.au/url?sa=t&source=web&cd=1&ved=0CCoQFjAA&url=http%3A%2F%2Fwww.goingtouni.gov.au%2FNR%2Frdonlyres%2F9A024399-A4F6-411D-9012-5FA4C3003DD5%2F0%2Ficss_booklet_2006.pdf&rct=j&q=information%20for%20commonwealth%20supported%20students%202006%20site%3Ahttp%3A%2F%2Fwww.goingtouni.gov.au&ei=OrHXTJylCoSuuQOKhuyoCQ&usg=AFQjCNFyGK_ZO-2Ry4Y48AGZdYvjMubBRQ&cad=rja>.

Department of Science Education and Training (DSET). 2008, *Higher Education Loan Programme – HECS-HELP – March 2004*, accessed 6 November 2010, <http://pandora.nla.gov.au/pan/35488/20090212-0001/www.backingaustraliasfuture.gov.au/fact_sheets/pdf/fs5.pdf>.

Department of Education, Employment and Workplace Relations (DEEWR). 2004 – 2008, *Higher Education Statistics, Full Year - Tables, Commencing Students,* accessed 6 November 2010, <http://www.deewr.gov.au/HigherEducation/Publications/HEStatistics/Publications/Pages/Students.aspx>.

Department of Education, Employment and Workplace Relations (DEEWR). 2004 – 2008, *Higher Education Statistics, Full Year - Tables, Liabilities*, accessed 6 November 2010, <http://www.deewr.gov.au/HigherEducation/Publications/HEStatistics/Publications/Pages/Students.aspx>.

GESIS – Leibniz Institute for the Social Sciences, 2010, *ISSP 2006 source questionnaire,* accessed 7 November 2010, <http://www.gesis.org/en/services/data/survey-data/issp/modules-study-overview/role-of-government/2006/>.

Household, Income and Labour Dynamics in Australia (HILDA). 2010, *Wave 6 New Person Questionnaire,* accessed 7 November 2010, <http://www.melbourneinstitute.com/downloads/hilda/Questionnaires/NewPersonQuestionnaireW6.pdf>.

Jackson, K. 2003, *The Higher Education Contribution Scheme*, *Parliamentary Library, Analysis and Policy, Social Policy Group*, accessed 7 November 2010 , <http://www.aph.gov.au/library/intguide/SP/hecs.htm>.

Marks, G., N. 2009, 'The Social effects of the Australian Higher Education Contribution Scheme (HECS)', *Higher Education*, vol. 57, pp 71 – 84.

Rosenbaum, P. & Rubin, D., B. 1983, 'The Central Role of the Propensity Score in Observational Studies for Causal Effects', *Biometrika,* vol. 70, pp. 41–55.

Watson, N. 2010, 'HILDA User Manual – Release 8', Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Yu, P., Kippen, R. & Chapman, B. 2007, 'Births, debts and mirages: the impact of the higher education contribution scheme (HECS) and other factors on Australian fertility expectations', *Journal of Population Research,* vol. 24, no. 1, pp. 73-90.