MINISTRY OF EDUCATION

Te Tāhuhu o te Mātauranga

What factors impact on graduates' earnings three years post-study?

Report
What factors impact on graduates' earnings three years post-study?

Author
Bhaskaran Nair, Senior Research Analyst
Tertiary Sector Performance Analysis & Reporting, Ministry of Education
Email bhaskaran.nair@minedu.govt.nz
Telephone 64-4-463-7797
Fax 64-4-463-8526

Acknowledgements
The author gratefully acknowledges the support of Sid Durbin, David Paterson, Roger Smyth, Warren Smart, David Scott and Jamie Hyatt in their role as peer reviewers.

Disclaimer
Access to the data used in this study was provided by Statistics New Zealand under conditions designed to give effect to the security and confidentiality provisions of the Statistics Act 1975. The results presented in this study are the work of the author, not Statistics New Zealand.

Summary Statistics New Zealand Security Statement: Integrated Data Disclaimer
The integrated dataset on Student Loan Scheme Borrowers is based on the integration of data from the Ministry of Social Development, the Inland Revenue Department and the Ministry of Education. This project has been approved by Statistics New Zealand as a data integration project with data access provided through the Data Laboratory under relevant legislation and policy. Only approved researchers who have signed Statistics New Zealand's declaration of secrecy can access the integrated data in the Data Laboratory. For further information about confidentiality matters in regard to this study please contact Statistics New Zealand.

Published by:

MINISTRY OF EDUCATION

December 2006

This report and an associated set of statistical tables are available
on the Ministry of Education's Education Counts web site at www.educationcounts.edcentre.govt.nz

ISBN (Web) 0-478-13556-4
ISBN (Print) 0-478-13555-6

Contents

1. Executive summary

This study aims to deliver a partial answer to the policy question "What are the post-study earnings of people with various educational and demographic characteristics?" This report forms part of a wider project on the assessment of investing in tertiary education. The main focus of this study is to assess the impact of demographic and study-related variables on student earnings three years post-study, using student information from the Integrated Dataset on Student Loan Scheme borrowers (IDSLS) maintained and managed by Statistics New Zealand.

A statistical model, generalised logistic regression analysis was used to model the relationship between the nominal outcome variable - student earnings three years post study - and the explanatory variables - demographic and study- or learning-related factors. This model isolates the independent relationship of each variable to earnings after accounting for the relationship of other relevant variables. The analysis drew on information gathered from about 98,000 students who attended tertiary education institutions, used the Student Loan Scheme and left study in the years 1997 to 1999. It both reinforces some of the findings of earlier studies and highlights the impact of some variables - like employer industry - and the interaction effects of various study-linked variables that were not studied in the previous works.

An assessment of overall fit of the generalised logistic regression model showed a pseudo R-square of 82 percent indicating the goodness of fit. The key findings from this study are that factors such as industry of employment, qualification level, the tertiary education provider, and the field of study influenced graduate earnings significantly. The interaction effects of the variables like: qualification level by industry, qualification level by provider, field of study by industry and age-group by qualification level, exerted a greater influence on graduates' post-study earnings than the factors considered individually. The demographic variables were generally not as strongly related to earnings, except age-group which showed a moderate relationship to student earnings. The individual effect of each explanatory variable is estimated by keeping all other factors constant.

The predicted average earnings of individuals attaining various qualification levels with different fields of study are estimated from the logistic regression analysis and compared. Students who chose Health, Engineering, Management and Commerce, and Education as their fields of study are likely to earn higher incomes than the individuals from other fields, controlling for all other relevant factors. This study also corroborates the earlier findings that the higher the qualification levels, the higher will be the earnings. The interaction effect of qualification level by field of study showed that people holding postgraduate qualifications in Management and Commerce or Health studies are likely to earn the highest. Among bachelors degree holders, the Engineering and Related Technologies, Health, and Education fields fetched higher predicted earnings. Individuals holding sub-degree qualifications in the fields of Engineering and Related Technologies, Health, and Education are also likely to earn high incomes relative to other fields of study.

The earnings differential for a sub-degree qualification is about 20 percent lower than the earnings for a bachelors level student. Postgraduate qualifications earned about 58 percent more than the bachelors level when adjusted for other factors.

Among industries of employment, Engineering, Mining, Telecommunication Services, Finance and Insurance, Property and Business Services, and Health and Community Services (in that order) provide high returns to tertiary study. Although Engineering, Mining, and Telecommunication Services provide high returns to tertiary qualifications, they have significantly lower employee counts. Individuals attaining qualifications in occupation-focused fields like Education, Health, Management and Commerce, Agriculture, Engineering, etc are most likely to be employed in industries that match their fields of study. The Government Administration and Defence, Property and Business, Health and Community Services, and Education sectors are the most common employment destinations.

The variables completion status (whether someone has completed a qualification or not), nature of study (full-time or part-time) and prior activity (main activity before taking up study) demonstrated statistically significant but weak relationships to student earnings. The premium on completing a qualification assumes significance only when combined with qualification level. For any given level of study, students who complete a qualification earn more than those that don't complete. However, when considered independently of level of study, the premium for completing a qualification is only marginal.

The premium on earnings is higher if the student studied part-time. Students, who attained their tertiary qualification from a university, have a high likelihood of earning higher income three years post-study. In general, highly qualified university graduates with prior work experience have a very high probability of earning a high income.

2. Introduction

Overview

Most individuals who invest in tertiary education do so on the assumption that this investment will yield future benefits such as greater employment opportunities, higher earnings and higher living standards. Similarly, government policy encourages New Zealanders to increase their educational attainment and enhance their skills and knowledge. Increased educational attainment and skills are valued because they are believed to result in better labour market and social outcomes. Better-educated workers earn higher wages; have greater earnings growth over their lifetimes, experience less unemployment and work longer. Higher education is also associated with longer life expectancy, better health, and reduced participation in crime. Much empirical evidence supports this view and show that attainment of tertiary qualifications is one of the best predictors of a successful future.

Tertiary education has expanded in New Zealand over the last 15 years and education providers have welcomed students from a broader cross section of the population. The labour market for tertiary qualifications has contributed to the rise, despite the large increase in the number of qualified people in the labour force.[1,2] In addition, the premium paid by the labour market for tertiary-educated workers is increasing as the economy moves towards being knowledge-based.[3] As a result, the nature of graduates' work is undergoing a transition from the application of traditional skills to more diversified and knowledge-based skills. Many students already work part-time throughout their courses, and are graduating into a wider range of jobs and career patterns. More firms are interacting with tertiary education providers not only to source new recruits, but also for up-skilling their existing employees.

Formal tertiary education in recent years has witnessed steady growth in terms of both student enrolment and completion. Between 1998 and 2005, the number of domestic students attaining a tertiary qualification grew by 54 percent, while the proportion of the population aged 15 years and above participating in tertiary education has grown from 10.0 to 14.5 percent in the same period, partly as a result of a change in the government's removal of the rolling cap on the number of student places it funded.[4]

In this context, this study aims to deliver a partial answer to the following policy questions, as part of a much wider project on the assessment of socio-economic impacts of investing in tertiary education:

- How do earnings differ among different qualification levels and fields of study?
- How do earnings differ among different employers?
- How do age, gender and ethnicity influence earning levels?
- How does the activity prior to taking up tertiary education influence income levels in addition to the influence of provider sub-sectors and nature of study?

The major focus here is on assessing the impact of demographic and education-linked variables on the graduates' post-study income profile. The probability distribution of a

[1] Maani, S. A. (1999). Private and Public Returns to Investments in Secondary and Higher Education in New Zealand over Time: 1981-1996. Research report commissioned by the New Zealand Treasury, www.treasury.govt.nz/workingpapers/default.htm, accessed 10 June 2006.
[2] Penny, N. (2005). The Approach to Measuring the Returns to Secondary and Tertiary Qualifications in New Zealand: An Investigation and Update Using Data from the 2001 Census. Unpublished thesis.
[3] Dillingham, S. (2003), New Zealand's Workforce: Qualifications and Evidence of Up-skilling. Labour Market Policy Group. Wellington: Department of Labour.
[4] Ministry of Education (2005). Profile and Trends: New Zealand's Tertiary Education Sector 2004, page 164.

student loan borrower assigned to various pre-defined income bands is estimated using a multivariate logistic model with multiple responses.[5] Patterns and trends observed in the estimated probability distribution are analysed and discussed in the results and discussion section.

Prior research on the relationship between education levels and earnings

This report is based on the assumption that returns to education are an important indicator of the tertiary outcomes. There is a substantial body of evidence that shows that those with higher levels of education are more likely to participate in the labour market, face lower risks of unemployment, have greater access to further training and receive higher earnings on average. Interest in these measures arises because they are an indicator of the private benefits individuals obtain as a result of their education. However, their importance goes beyond the private benefit. The earnings employees can command in the labour market reflect the value that those employees create for the firms that employ them. A firm will continue to employ someone only as long as the value that person creates for the firm exceeds the cost of their employment. In other words, differences in earnings among different groups are a proxy measure for differences in the human capital they possess. Implicit in this is an income-based model of human capital.[6]

Most research on the contribution of human capital to economic growth and its role in the distribution of income use only relatively crude indicators such as educational attainment and years of labour market experience. Educational attainment is generally measured by years of schooling or highest level of education attained. Labour market experience is unobserved in most tertiary datasets and is often proxied by factors like age and prior activity before taking up tertiary studies. However, individuals with the same level of education and experience may have substantially different skills, which depend on their innate ability, their family environment, their fields of study, their work experience and on-the-job training, and other factors. More generally, education and work experience are the input variables in human capital theory instead of direct measures of outcomes like skills and competencies. There are very few studies that have examined the roles of these direct measures of ability factors.[7]

In spite of hundreds of studies carried out on the relationships between inputs such as education and experience and outcomes such as employment and earnings, relatively little is known about the relationship between direct measures of skills and labour market outcomes. Conventional estimates of the returns to education and to labour market experience confound two effects. The first is the impact of education and experience on skill development (the relationship between human capital inputs such as education and experience) and outputs such as problem-solving skills. The second is how these skills are valued in the labour market (the relationship between problem-solving skills and market earnings). In the absence of direct measures of these skills, it is not possible to separate these two influences on labour market outcomes, resulting in a confounding effect leading to upward bias in the estimation, thus overstating the role of education.

Most studies of this nature are centred on a theory of human capital that emphasises the influence of educational attainment in raising earnings through its enhancement of workers'

---

[5] See Appendix 1 for details.
[6] Le, T, .J. Gibson and L. Oxley (2005). Measures of Human Capital: A Review of the Literature. http://www.treasury.govt.nz/workingpapers/2005/wp05-10.asp
[7] For example, Boissiere, Knight and Sabot (1985), Rivera-Batiz (1990, 1992), Murnane, Willett and Levy (1995) and Charette and Meng (1998).

skills, thus making employees more productive and more valuable to employers. However, according to signalling/screening theory (Stiglitz, 1975), education has no effect on human productivity, but it acts as a signal of the productive capacity of an individual. A basic assumption of this model is that education is less costly to acquire for individuals who are innately able. If this assumption holds, higher ability individuals will invest more in education than lower ability individuals. Both high and low ability individuals face the same potential benefits from investing in schooling, but low ability workers face higher costs and therefore will acquire less education. As a consequence, education is seen as privately productive but as having little effect on public productivity. This also suggests that the relationship between education and earnings exists through a third factor, namely individual ability. As ability is unobservable, we are not sure of the extent to which individual productivity is influenced by education or the person's innate ability.

An information-based theory of human capital emphasises that educational attainment may help people learn about their comparative advantages, in addition to directly enhancing skills and knowledge. Further, the model suggests that mobility between jobs should decrease with time in the labour market as workers learn about their own abilities and are more likely, as a result of moving from job to job, to find a good match.

This approach has led to a review of the relative importance of generic versus specific skills in human capital. Much empirical work states that workers who have been working with a firm a long time have higher wages than similar workers with less tenure. This was previously interpreted to mean that firm-specific skill was very important and its accumulation was associated with increasing wages. However, the information-based model suggests that causality may also go in the other direction. Good firm-worker matches have high wages because they benefit both parties, and are more likely to endure.[8]

Relatively few studies have analysed the trends and patterns in post-study earnings in New Zealand. A brief review of some of the work carried out using Student Loan Scheme data and Census, Household Labour Force Survey and Income Survey information is presented below.

Evidence on the longer-term outcomes of tertiary education and the evidence for the benefits people gain from tertiary education is summarised in the Ministry of Education (2005)[9]. Here, the analysis of the New Zealand Income Survey dataset clearly demonstrates that income level increases with the higher level of qualification. The report also gives a detailed analysis of income levels among various ethnic groups.

The Maani and Maloney (2004) study examines the returns to post-school qualifications in New Zealand using Household Labour Force Survey (HLFS) income supplements from 1997 to 2002. In this study they estimate returns to education using four models across the years 1997 to 2002 on income, earnings, hourly wage and weekly wage. They compare the returns of post-school to school qualifications in the presence of demographic characteristics for the New Zealand working-age population. The results show positive returns to schooling and how the returns change for different education levels.

---

[8] See Riddell and Arthur (2006) for more information about these models.

[9] Section 3 of the Profiles and Trends 2004, (Ministry of Education) focuses on the benefits of undertaking tertiary education for finding an answer to the question "Is the opportunity cost of spending in tertiary education adequately compensated by the post study earnings"?

Hyatt et al (2005) analyse the post-study total income in 2000 for student loan borrowers who last studied in 1997. This report focuses on the medians and percentiles of total income ratios as influenced by the field of study, level of study, age, gender, completion status and ethnicity.[10] This study doesn't use a model approach to study the influence of educational factors on total income. The analyses mainly focus on classifying the borrowing students according to the level of study they undertake and whether or not they complete that qualification. This approach provides a more realistic assessment of the value added by tertiary qualifications compared with the population-based comparison of outcomes between these two groups. The distinction on the basis of completion status provides a proxy for differences in prior qualification or the innate ability of the individual.

A recent study by Penny (2005) repeating Maani's (1999) approach using data from Census 1996 and 2001 shows that the rate of returns increased for all tertiary qualifications, despite the fact that the supply of qualifications rose over the period under study.

Marè and Liang (2006) examine the labour market outcome in the first five to 10 years after graduation with reference to the variability in the outcomes among individuals with the same qualification levels using 1996 and 2001 Census data. The 1996 and 2001 Census datasets are used to compare labour market outcomes of 18 to 30 year olds and 18 to 65 year olds who had attained post-school qualifications. The outcomes are analysed by fields of study to highlight the variations within the post-school graduate population. They also investigate the demand and supply position of the post-school graduate population in different fields and compare the results between the two Census periods.

Hyatt and Smyth (2006) analysed and reported the changes in earnings between three and five years following study. The result indicated that successful completion of study paid a lower premium at sub-degree levels of study. By studying the changes in earnings between the third and fifth years post-study, this analysis attempts to distinguish the effect of labour market values experience from the influence of attainment of a qualification. The unobserved ability factor in this analysis is tackled as explained under Hyatt *et al* (2005). The results showed that tertiary level qualification is associated with a higher starting salary but the advantage will be weaker over time if the qualification is below degree level.

To some extent the Integrated Dataset on Student Loan Scheme borrowers (IDSLS) used in the analysis provide an opportunity to analyse the longitudinal dimension of the income, which was otherwise not possible with the existing data sources like the Census or other income based surveys. The IDSLS is still evolving and new longitudinal income data are added to the dataset every year, enabling us to understand the temporal changes in tertiary education and its socio-economic implications. In this context, this report has a special significance in understanding the relationship between education and earnings where a more realistic data are used for analysis.

Scope

An assessment of the impact of some specific study-related factors on the earned income of those students who have finished study – whether complete or incomplete - in 1997, 1998 and 1999 is included in this report. The details of the study-linked variables and demographic variables used in this analysis are given in Section 4. The income of a former student, three years after study, is taken as the target variable. This study excludes assessment of the

[10] Hyatt, J., P. Gini and R. Smyth (2005). Income of Student Loan Scheme Borrowers. Wellington: Ministry of Education.

effects of post-study earnings on student loan borrowing and repayment ability. The assumptions underlying this analysis are given in Section 3.

Report structure

This report is divided into five sections. The first section gives the executive summary and the second section gives an introduction to the report. Section 3 explains the methods and data used to analyse the impact. It also gives an overview of the methods used to model the intrinsic association of study-related variables and demographic variables on earned income. Section 4 looks at the results and discusses the relationship between study-related variables and income profile. The conclusion and key findings are given in Section 5. This also gives an overview of the implications of the results on the socio-economic factors and looks into future courses of action to deal with the evolving ISLDS longitudinal dataset.

Appendix 1 describes the generalised multinomial logistic model used in the analyses and Appendix 2 gives the theoretical descriptions on the correspondence analysis. In Appendix 3 the Type III analysis report from the logistic analysis output given. ANZSIC96 industry two-digit codes and descriptions and NZSCED fields of study codes are given in Appendix 4. Tables and graphs related to results and discussion part are given in the remaining Appendices.

3. Methods

Data - scope and coverage

This report is based on the findings from the analysis of the Integrated Dataset on Student Loan Scheme borrowers managed and maintained by Statistics New Zealand. This dataset has integrated information on student profiles collected from tertiary education providers, student borrowing information collected from Study-Link and student income profile information from Inland Revenue, namely tax, loan balance and repayments information.

The dataset covers students who have completed their study, whether successful or not, for the period from 1997 to 1999 and their earned income status three years post-study in the years 2000, 2001 and 2002, respectively. Three demographics variables, seven study-related factors and one industry variable (used as a proxy to occupation) are included in the analysis. The earned income from different employers is derived from the IRD income profile integrated into the dataset. The three leaving cohort years covered approximately 98,000 students.

The integrated dataset is reported to be covering more than 50 percent of the total number of students and 70 percent of the full time students enrolled in New Zealand. However, results from this study reflect only the students in the loan scheme and hence inferences about the general population may be drawn only with caution. Where there are multiple sources of income, only earned income from the single employer who contributes the maximum to the annual income is included in the dataset. Students who have declared their residence as overseas are excluded. Students' earnings from self-employment showing negative income also are excluded from the data as they constitute a negligible proportion, to avoid any outlier issues affecting the statistical model.

The ISLDS currently contains matched records for years 1992 to 2004. Although Inland Revenue data are available since 1992, complete information on the student profiles from the tertiary education provider is available only from 1997 onwards. Therefore matched data are available for this study from 1997 onwards. A small number of records that did not match the information from various data providers were excluded. The numbers on enrolments and completion status, etc reported in this study may vary from the numbers available in the Ministry of Education's statistics database due to some data mismatches in the integrated dataset.

The addition of variables like prior activity and age group is expected to capture the effect of work experience. However, there is an absence of data on the cost of education, which is a prerequisite for estimating the rates of return, as it is not within the scope of this analysis. We also have some apprehensions about how far a short period of three-year post-study earnings is able to capture the effect of firm-specific skills, mobility between jobs and the longer-term effect of types of education, which are other important unobservable factors influencing the earnings.

Underlying assumptions

Human capital theory assumes that the lifetime earnings profile of more qualified employees is above the equivalent earnings profile of less-educated workers. Further, it also accounts for the fact that earnings rise with work experience, although at a diminishing rate. The increase in earnings with experience is especially pronounced during the first five to 10 years after entering the workforce. Education and work experience enhance the individual's skills, thereby raising their market value to employers.

Many experts in this field interpret the correlation between education and earnings as evidence that education exerts a causal effect on earnings. The estimates of the returns to education are biased upwards because of unobserved factors, estimated average rates of return to education may substantially over predict the economic benefits that a less-educated person would receive if they acquired higher levels of tertiary qualification. The unobserved ability factor bias issue is important not only for the question of how we should interpret the positive relationship between earnings and education, but also for the emphasis that should be placed on education in public policies. Therefore care needs to be exercised in the interpretation of the results in the given context as this report analysis doesn't include data that directly control for the unobserved ability and motivation factors.

The distinction between generic and specific skill levels in the borrowing student population is also an important factor that influences the post-study earnings. Generic skill refers to skills and knowledge that are useful to many employers, while specific skill is specific to one employer but not to others. This may also include skills that are industry-specific and occupation-specific. This distinction is useful in understanding the premium paid for individual workers for gaining specific skills vis-à-vis generic skills. We assume that industry-specific skill effects are accounted for, to some extent, by the interaction effects of variables industry by fields of study. Some basic assumptions we hold in drawing conclusions from this analysis are given below.

- Student income level is a reflection of skills attained through the tertiary education.
- The industry in which the student is employed is treated as a proxy to the student's occupation.
- Earned income is treated as an ordinal variable and follows a multinomial distribution.
- The Labour Cost Index is used to standardise the annual earnings.
- Industries are classified according to ANZSIC96 code (Appendix 4).

Variable selection - dependent variable

Individual annual earned income is treated as the dependent variable and is measured by 10 equal width income bands as shown in Table 3.1. Only one income source that contributes the maximum to the total annual income is considered for those students having multiple sources of income. The frequency distribution of the individuals within each leaving cohort and the income band codes are shown in Table 3.1. Self-employed individuals showing negative income are excluded from the analysis because of their very small occurrence frequency.

Table 3.1 Response profile

Earned Income ($)	Code	Frequency 1997	1998	1999	Combined*
0	O	3,276	4,869	5,520	13,665
1-10,000	A	5,217	5,694	6,054	16,767
10,001-20,000	B	4,074	4,377	4,818	13,083
20,001-30,000	C	4,854	5,379	5,886	15,816
30,001-40,000	D	5,361	6,534	7,557	19,077
40,001-50,000	E	2,577	3,300	4,263	10,794
50,001-60,000	F	984	1,356	1,701	4,266
60,001-70,000	G	360	504	654	1,614
70,001-80,000	H	180	237	276	720
>80,000	High	231	315	351	954
Total		27,114	32,565	37,080	96,756

Note: * The combined total may not add up to individual cohort year totals as the mismatched, missing entries are excluded from the analysis.

Variable selection – independent variables

The independent variables are selected from two major groups, namely study-linked and demographic variables. The details of the independent variables and their category levels are given in Table 3.2. Eight factors, namely field of study, level of study, completion status, activity before taking up the programme, provider sub-sector, EFTS (equivalent full-time student) programme years, and nature of study (part-time or full-time) are selected as the study-related factors. EFTS in years is the only continuous variable included in the study-linked factors. Age as on 1 July of the year of completion (expressed in five-year age group), gender and ethnicity form the demographic variables. Other relevant variables like income from benefit, and number of years of schooling were not included in the analysis due to a large number of missing cells.

Table 3.2 Descriptions of explanatory variables

Explanatory variables	Category description
Age- (at 1 July of leaving study year in 5-year age groups)	1=15-20, 2=21-25, 3=26-30, 4=31-35, 5=36-40,6=41-45, 7=46-50, 8=51-55, 9=56-60 and 10=> 60
Gender	0=Female, 1=Male
Ethnicity	1=European, 2=Māori, 3=Pasifika, 4=Asian 5=Others, 6=Unknown
Qualification Level	01=Certificate Level 1-3; 02= Certificate Level 4; 03= Certificate and Diploma Level 5-7; 04=Bachelors; 05=Postgraduate Certificate /Diploma/Honours; 06=Masters; 07=Doctorate
Field of Study	01=Natural and Physical Sciences; 02=Information Technology; 03=Engineering and Related Technologies; 04=Architecture; 05=Agriculture and Environmental Studies; 06=Health; 07=Education; 08=Management and Commerce; 09=Society and Culture; 10=Creative Arts; 11=Food, Hospitality and Personal Services; 11=Mixed Field Programme
EFTS Programme Years	Continuous variable, Describes the theoretical equivalent full-time study of a qualification in years.
Nature of Study	P=Part-time; F=Full-time; N=Not studying
Completion Status	0=Incomplete; 1=Complete
Prior Activity	01=Secondary school student; 02=Non-employed/Beneficiary; 03=Wage or salary worker; 04=Self-employed; 05=University student; 06=Polytechnic student; 07=College of Education student; 08=House-person or retired; 09=Overseas (irrespective of occupation); 10=TOPS student; 11=Private training establishment;12= Wānanga student
Provider Sub-sector	Polytechnics; Colleges of Education; Universities; Wānanga; Others TEP
Industry	See Appendix 2 for details

The purpose of this study is to understand and describe the relationship between study-linked factors and employment-based earned income. Therefore, in building a model, the decision to add a variable will be based not on its statistical significance, but rather on whether the presence of that variable in the model significantly changes the relationship between study variables and the earnings.

Analytical methods

In the literature, one important approach used to analyse the relationship between education attainments and earnings is standard multivariate methods such as ordinary least squares (OLS) estimation. But this method suffers from the limitation that it may estimate the correlation between earnings and education, after controlling for other observed influences on earnings, rather than isolating the causal impact of education on earnings.

This study is based on the human capital theory model. This model uses a multivariate method called "generalised multinomial logistic regression model" with multiple response classes to analyse the association between study-linked factors such as field of study, qualification levels, etc, in which the outcomes can assume one of many income bands. This model is best suited to situations similar to this data, where explanatory variables are categorical and the response has more than two levels. Multinomial logistic regression, sometimes referred to as "polychotomous logistic regression", is the extension of the logistic regression model when the outcome is recorded at more than two levels (Hosmer and Lemeshow, 2000).[11]

The known outcomes (earnings) from the data serve as the basis for estimating income probabilities. The income band in the range of $1 to $10,000 served as a reference category for the calculation of the odds ratio. This analysis gives a set of coefficients or weights. The logistic regression model chooses the set of coefficients that would produce prediction probabilities that match very closely to the observed probabilities. The sign of the coefficients indicates whether the presence of the characteristics increases or decreases the likelihood of the income probability and the size of the coefficients reflects the strength of the relationship between the characteristics and the outcome. A statistically significant coefficient denotes that the related variable has statistically significant influence on the outcome (earned income). Unfortunately, the interpretation of logistic regression coefficients is not straightforward and in order to understand their meaning it is necessary to convert them to another form. When the sample size is too large, as in our case, a statistical significance by itself means little. The most popular form is odds ratio, or alternatively we may use predicted probability or delta-p. All comparison of effects in the model is made on the basis of the predicted probabilities in this report.[12] Note that R^2-like measures shown in Table A3.2 (Appendix 3) are not goodness-of-fit tests but rather an attempt to measure strength of association.

The statistical significance and the magnitude of Wald chi-square statistics from the Type III Analysis of effects table are used as a crude method to determine whether a particular independent variable has statistically significant effects on the dependent variable or not[13] . The results indicate that all the independent variables influenced the earnings outcome significantly. It is possible that when the sample size is too large, the coefficients are likely to show significance. In such a situation it is better to depend on the magnitude of the Wald chi-square value to establish the importance of the variable.

In this study the relative importance of the variables is identified on the basis of the Wald chi-square value. However, it may be noted that this statistic may sometimes overestimate the importance of a variable when the coefficient is too large. In such a situation it is better to compare the relative importance using the -2*LL* (log likelihood ratio) values for models with and without the variables concerned.

Comparison of log likelihood ratio was not performed here due to the computational difficulties and convergence problems faced in analysing large datasets like this one. Fitting logistic analysis using cross-validation technique was also could not be performed due to the limitations in the analytical environment. An outline of the logistic model employed in the analysis is given in Appendix 1.

All the interaction effects above two levels are hard to interpret, and most importantly these effects are generally not statistically significant. So, the interaction effects that are really important are considered at only two levels. There are instances where higher levels of

[11] Several reports and studies have effectively used the logistic regression methodology to measure the impact assessment in higher education, e.g. Steiner and Teszler (2005), who have reported the use of logistic regression analysis in the analysis of student loan defaulters in Texas. Adelman (2006) reported on the analysis of paths to degree completion from high school through college, using logistic regression analysis.

[12] See Appendix 1 for the definition of predicted probability.

[13] A detailed statistical diagnostics and analytical outputs can be obtained from the author on request.

interaction may be important for discussion but, keeping in view the difficulty in interpreting, the higher order interactions are not considered in this study.

Correspondence analysis for two-way tables

Two-way tables with the incidence (frequency) of qualification levels by field of study, qualification level by ethnicity, qualification level by prior activity, qualification level by provider sub-sector, qualification level by age groups, etc are interpreted through graphical visualisation produced through correspondence analysis. The correspondence analysis is a statistical visualisation method for picturing the associations between the levels of a two-way contingency table. In a two-way contingency table, the observed association of two traits is summarised by the cell frequencies, and a typical inferential aspect is the study of whether certain levels of one characteristic are associated with some levels of another.[14]

This method is a geometric technique for displaying the rows and columns of a two-way contingency table as points in a two-dimensional space, such that the positions of the row and column points are consistent with their associations in the table. Graphical display of these interaction tables and their interpretations is shown in the respective sections dealing with the interaction effects, to give a general view of the data that is useful for interpretation. The details of the correspondence analysis are given in Appendix 2.

Tables and graphs

Relevant frequency distribution tables are given at the beginning of each section of this report under *Results and discussion*. The line graphs are given for predicted probabilities of income distribution for several sub-groups under different scenarios. For the sake of convenience, the income distribution for interaction effects, namely level of study by industry, level of study by prior activities, etc, is spread across more than one graph to get a clear view of the patterns and trends in the income distribution within a sub-group.

The bi-plot graphs depict the nature of similarities and dispersion within the (row and column) clouds as well as the correspondence between the clouds. Distances between points of different clouds have little meaning! What is important is the relative closeness of points in one cloud to points in the other cloud (correspondence).

[14] Correspondence analysis is related to the Principal Component Analysis, method used for categorical variables. See Appendix -2 for details. (Van der Heijden and de Leevw, 1985).

4. Results and discussion

The type III analysis of effects produced by the generalised logistic regression is listed in Table 3.1. This table lists explanatory variables, its reference category, the parametric estimates, chi-square value, and significance in the order of their occurrence in the model.

The results of generalised logistic regression analysis are summarised in a number of tables. The analysis of maximum likelihood estimates result outputs are available on request from the author. This table includes parameter name, variable level and a response variable column to identify the corresponding logit by displaying the non-reference level of the logit. It also contains a maximum likelihood estimate of the parameter, estimated standard error of the parameter, Wald chi-square statistic (similar to t-statistics), computed by squaring the ratio of the parameter estimate divided by its standard error estimate, and *p*-value of the Wald chi-square statistic with respect to a chi-square distribution with one degree of freedom. There are nine intercepts, one for each generalised logit regression equation, and nine regression slopes for each variable category. All the variables except *EFTS years* are categorical with varying category classes.

All variables included in the model are tested for their significance using the Wald chi-square value shown against each effect in the analysis of effects shown in Appendix 3 (Table A3.1). The results of the goodness-of-fit test for the model based on the difference between fitted model and the intercept only model are given in the Model fit statistics table in Appendix 3 (Table A3.2). The model statistics and the deviance goodness-of-fit statistics indicate that the model fits the data adequately. All the variables used in the model show statistically significant chi-square value indicating the significant influence of each variable to the earned income. A very high R^2 (82 percent) measure, shows that a high degree of association exists between the independent variables and earned income.

Model output results

Since it is difficult to interpret the logistic coefficients in their original form, the predicted probabilities estimated using these coefficients are used to identify the impact of the factors. The predicted probabilities, plotted for each effect against the corresponding response levels, are shown in the relevant section. What the logistic regression analysis does here is to estimate the occurrence probability of earnings in each income band and observe the distribution pattern by plotting them against income bands. Average earnings due to the impact of each explanatory variable are obtained by taking the sum of products of the event probability and the mid-point of each income band. The impact is assessed by taking the percentage difference in the predicted earnings observed among categories of a specific variable. The results of the analysis are discussed in detail in the following sections. The contribution of each explanatory variable to the variation in outcomes among individuals is judged using the magnitude of Wald's chi-square statistics given in Appendix 3 (Table A3.1).

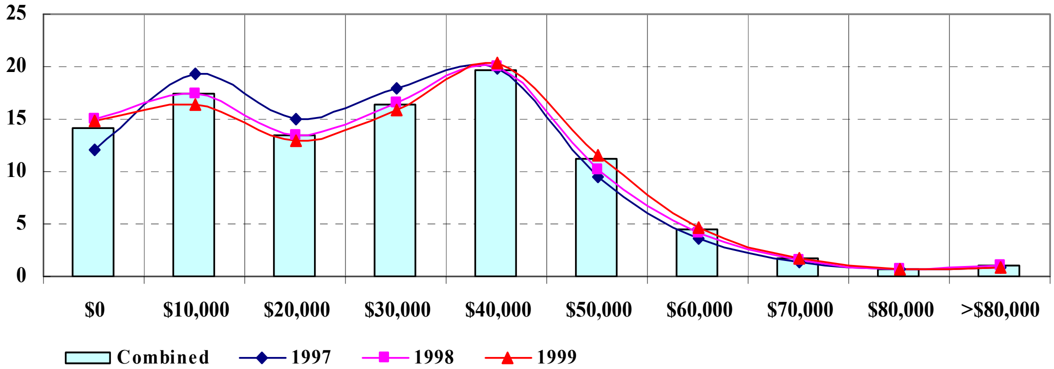
Income distribution (dependent variable)

Income frequencies observed in cohort years 1997, 1998 and 1999 and the combined cohorts three years post-study for different income bands[15] are shown in Figure 4.1. The income series from 2000 to 2002 are adjusted for movements in the Labour Cost Index. The deflated

[15] The income band groups and other details are given in the response profile table (Table 3.1).

income series are then used here to express the earnings in real terms (constant 2002 dollars). About 12 percent of the individuals recorded zero income during 1997, which slightly increased to about 15 percent in the subsequent years. The proportion decreased steadily from income band $40,000 onwards. The proportion of people in income bands $10,000, $20,000 and $30,000 are higher for the 1997 cohort than for the 1998 and 1999 cohorts.

Figure 4.1 Three-year post-study income distributions (%) in each cohort year and the combined cohort



However, the proportion drops subsequently in the higher income bands compared with the 1998 and 1999 cohorts. The income distribution shows a bimodal distribution with two peaks with values clustering around $10,000 and $40,000, as seen in Figure 4.1. The proportion falls steeply beyond $40,000. The summary statistics like mean, median, standard deviation, etc may distort the picture if the distribution is bimodal. This is typically indicative of two groups of people included in the sample. The people who have low income form one group and the other group consists of those who have higher income. It is also interesting to note that more than 45 percent of individuals remain in the first three income bands ($10,000, $20,000 and $30,000). One natural way of analysing bimodal distribution data is to group them and apply logistic regression to each group.[16] In this report the earned income is grouped into 10 income bands and the generalised multinomial logistic regression model is applied. This method fits the logistic model to each group independently.[17]

Explanatory variables

The student characteristics that relate solely to demographic, study-related factors and industry variable (as a proxy to occupation) are included in the model as explanatory variables. Additional details of the categories for all the variables are given in Table 3.2. The Analysis of effects table showing the details of the factors, their level of significance and their contribution to the total variability appears in Appendix 3 (Table A3.1). The results primarily focus on the study-linked variables and their interaction effects that appear to be important intuitively. Variables are selected on the basis of their importance and availability.[18] The student characteristics used in the analysis are discussed in the following sections.

---

[16] Logistic regression is an ideal method to analyse this kind of data which is capable of addressing the bimodal distribution effect. http://www.fon.hum.uva.nl/paul/papers/MeasuringCueWeighting.pdf

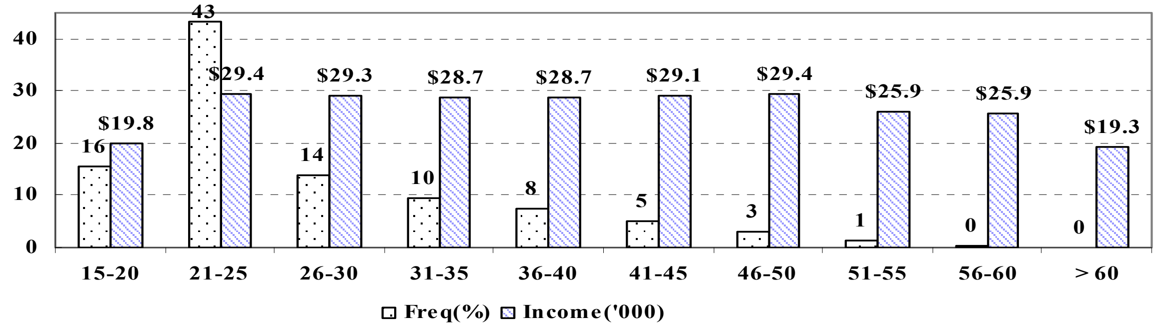
[17] Wermuth, N. and D. R. Cox (2001). Separation and Joint Response Graphs induced by Triangular systems. Research report. Australian National University. http://www.maths.anu.edu.au/research.reports/01srr.html

[18] The importance of the variables to be included in the model is ascertained after consultation with the experts in the field.

Age

Age as on 1 July of the year of leaving study is included in the model as a categorical variable with 10 classes of five-year age groups. Its contribution to the variation in outcome is seen to be negligible relative to other factors like industry, qualification levels and fields of study, by comparing the magnitude of the Wald statistics. Age-wise distribution of students and their observed median income in the combined cohorts are shown in Figure 4.2. About 43 percent of the students are in the 21-25 year age group and the frequency decreases steadily in the subsequent age groups. The median earnings of students in each age group are also shown in Figure 4.2. The median earnings remain above $25,000 and generally steady across the 21-25 to 46-50 year age groups. Thereafter the earnings decline as the age advances.

Figure 4.2: Frequency (%) and median earnings in different age groups



The analysis indicates that age factor has a positive and statistically significant, but weak, relationship with earnings (Appendix 3, Table A3.1). The impact of different levels of education plays a significant role in determining the income in the higher age groups as evidenced by the statistically significant interaction effect for qualification level by age group. The predicted average earnings, estimated from the probability distribution, for three qualification levels in different age groups are shown in Table 4.1. The earnings in different age groups show a quadratic relationship, especially if the qualification is postgraduate.

Table 4.1: Predicted earnings ($) of different age groups by qualification levels

Age Group	Sub-degree	Bachelors	Postgraduate
15-20	20,847	21,984	33,692
21-25	23,016	28,590	39,137
26-30	25,050	29,590	45,651
31-35	21,801	31,053	46,550
36-40	20,808	32,592	53,188
41-45	22,179	31,113	57,451
46-50	19,293	32,206	58,069
51-55	21,278	25,731	49,698
56-60	23,155	21,497	34,585
> 60	12,194	12,859	20,415

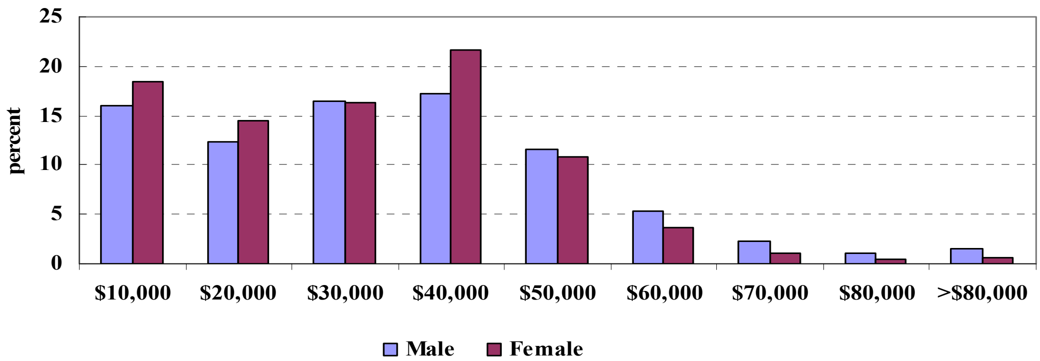
It may be noted that the observed earnings in the age groups 21-25 to 46-50 years do not reflect the usual parabolic curve structure of the earnings (Figure 4.2). However, after controlling for other intervening factors in the model, the earnings structure shows the normal parabolic curve, if the earnings shown in Table 4.1 are plotted for bachelors and postgraduate earnings. The earnings are relatively flat for sub-degrees, indicating the predominant influence of the sub-degrees qualifications in the data.

Further results on the probability distribution of earnings with reference to age groups and qualification levels are plotted in Appendix 5 (Figures A5.1a to A5.1f). The distribution curve shows certain patterns specific to qualification and age. For example, the pattern of the curve for age group 21-45 years holding a sub-degree qualification has a probability curve skewed to the right with a maximum density occurring at the lower income range (Appendix 5, Figure A5.2a). Likewise, bachelors degree holders in the same age group have a very symmetric earning distribution showing maximum density in the mid income range (Appendix 5, Figure A5.2b). A typical probability curve for postgraduate holders has a probability curve skewed to the left owing to maximum density in the higher income range (Appendix 5, Figure A5.2c). Income probabilities of these three qualification groups at the top (>56 years) and bottom (15-25 years) of the age groups are shown in Figures A5.2d to A5.2f (Appendix 5). The income distribution trend for sub-degrees and bachelors levels shows similar patterns, whereas in the postgraduate qualification groups the income distribution is rising and falling without any pattern (Appendix 5, Figure A5.2f).

Gender

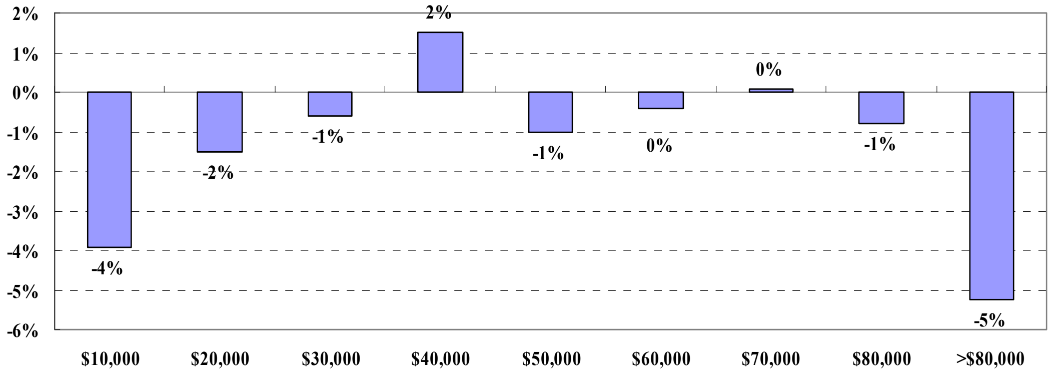
This is another demographic variable with two categories included in the model. In general, the female enrolment proportion was higher than the male as is evident from the graph shown below. The female proportion was higher than the male in all cohort years. In all the cohort years studied, the median income from earnings is greater for men than for women. The gender-wise share of income in 10 income bands is shown in Figure 4.3. The female proportion is generally high in the lower income range below $40,000 but fell vis-à-vis the male proportion beyond the $40,000 range. The gender influenced the earnings significantly but its overall effect is not as strong as industry, qualification level, fields of study and provider sub-sector, as indicated by its Wald chi-square value (Appendix 3, Table A3.1). Its interaction effect with other variables also showed negligible contribution.

Figure 4.3: Frequency distribution (%) of gender in 10 income bands – aggregated over cohort years



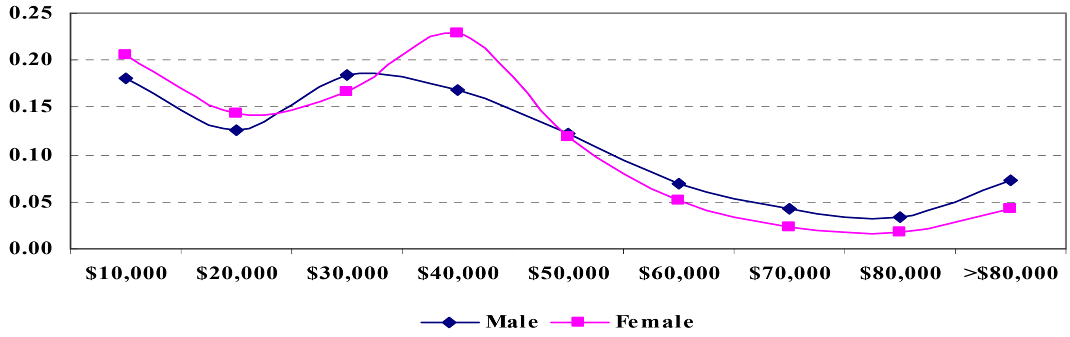
Earning differentials (the percentage changes in earnings with reference to female) between men and women in 10 income ranges are shown in Figure 4.4. In the graph, a positive percent change refers to earning differential in favour of female and a negative percent change is in favour of male. The earning differential was in favour of men in all income ranges except in the $40,000 range. However, statistically significant earning differences are observed only at the start and end points of the income range.

Figure 4.4: Earning differentials due to gender – aggregated over cohort years



Predicted income probabilities of earnings for men and women estimated from the analysis are depicted in Figure 4.5. It may be noted that the more women were in the lower income bands than men, and their earning probability decreases in the higher income bands. The average earnings estimated from the predicted probability areare $34,047 for men and $30,052 for women. An overall income differential due to gender effect is observed to be around 12 percent after controlling for other factors.

Figure 4.5: Predicted earning probabilities in men and women - aggregated over cohort years



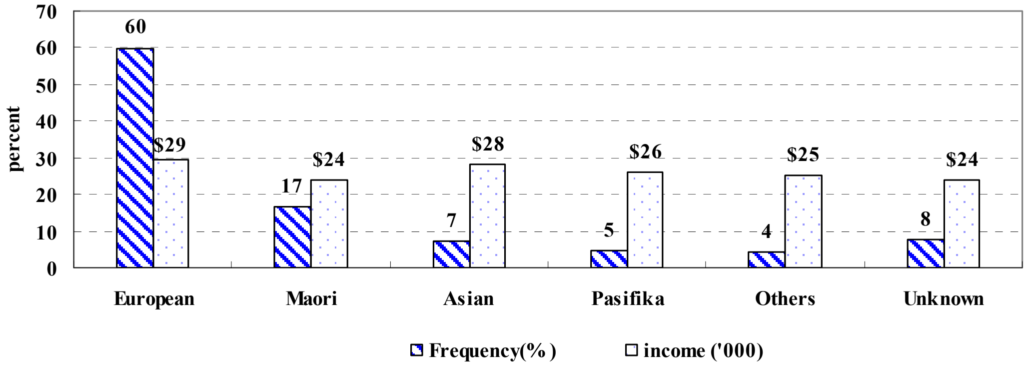
It is well documented in the literature (Heckman 1978; Smith 1980) that women's work participation decisions are quite different from those of men. Because of the common attitude toward traditional gender roles, most working women have to balance family and work demands (Vella 1994; Boden 1999). Therefore, withdrawal from the labour market on either a temporary or permanent basis is commonly observed for working women, resulting in loss of earnings.

Ethnicity

Prioritised ethnicity[19] is the third demographic variable included in the model. It has six categories, of which two non-descript categories are labelled "*Others*" and *Unknown*. The *Others* category includes all those who are excluded in the first four categories and unspecified ethnicity is grouped under *Unknown*. A summary of ethnic group frequencies and their corresponding median earnings is shown in the Figure 4.6. This summary reflects the returns to tertiary education by ethnic groups across three cohort samples.

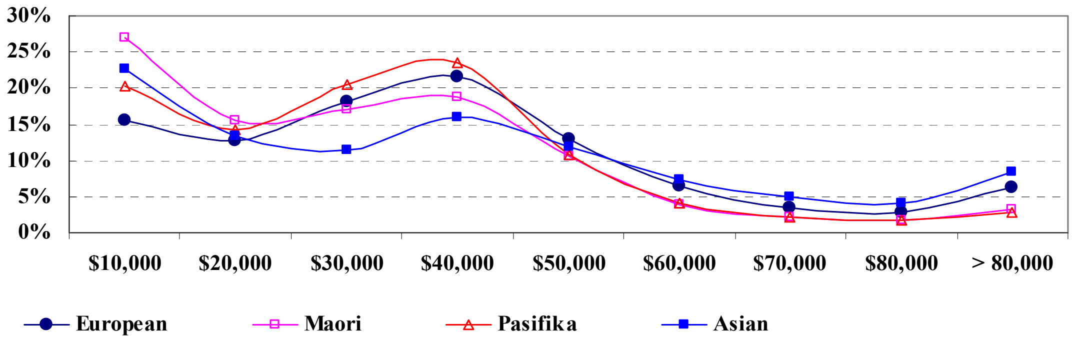
[19] Prioritised ethnicity refers the method used to prioritise ethnicity data to one response per person where multiple responses have been provided.

Figure 4.6 Ethnicity distribution and median earned income-aggregated over cohort years



Predicted income probabilities estimated for each ethnic group are depicted in Figure 4.7. Ethnicity influences the student earnings is statistically significant, but its overall contribution in explaining the earnings is only very small as indicated by the low Wald chi-square value (Appendix 3, Table A3.1). Māori and Pasifika peoples have higher likelihood of earnings in the lower income bands than European and Asian peoples. Their earning probabilities in the higher income bands were lower than the European and Asian groups.

Figure 4.7: Predicted probability of earnings by ethnicity - aggregated over cohort years



The correspondence analysis of ethnicity by qualification levels using the frequency tables aggregated over three cohort years is shown in Appendix 6 (Figure A6.1). The correspondence plot depicts that Europeans and Asian ethnic groups are clustered in the left side of the plot, while Māori, Pasifika, Others and Unknown are grouped in the right-hand side. Asian ethnic group is clustered around masters, postgraduate diploma and doctorate qualifications, whereas the Europeans are closely clustered around bachelors degree. The rest of the ethnic groups are clustered around the sub-degree qualifications area.

Field of study

This study-linked variable has 12 categories covering a wide range of subject fields.[20] Although its direct contribution towards the outcome is small relative to industry, qualification level and provider sub-sector, the indirect contribution through qualification level and industry is very substantial (Appendix 3, Table A3.1) as seen by its Wald chi-square value. The reference category for comparison of odds ratio is *Agriculture, environment and related studies*.

[20] The fields of study are classified as per the NZSCED codes shown in Appendix 6.

The overall results from the analysis indicate that *Health, Management and commerce*, *Engineering and related studies*, *Health* and *Information technology* have higher likelihood of earnings in the higher income range relative to the reference category *Agriculture, environment and related studies*. The fields of study distribution in sub-degree qualification levels are more concentrated in fewer areas than in bachelors and postgraduate levels. However, the pattern varies considerably according to qualification type. Actual median earnings, the frequency distribution and the coefficient of variability in income for each of the 12 fields of study are given in Table 4.2.

Table 4.2: Actual median income and frequencies in different fields of study

Fields of Study	Median Income($)*	Frequency (%)	Coefficient of Variation (%)
Society and Culture	25,484	24	86
Management and Commerce	31,653	17	91
Education	31,051	12	55
Creative Arts	21,935	11	82
Natural and Physical Sciences	28,192	8	88
Health	35,687	6	76
Food, Hospitality and Personal Services	20,143	6	72
Engineering and Related Technologies	35,411	4	74
Information Technology	26,750	4	85
Mixed Field Programme	19,973	3	117
Agriculture, Environment and Related Studies	25,722	3	79
Architecture and Building	29,909	2	73

Note: *Average of the median income of three cohort years; Zero income cases are excluded.
Source: Statistics New Zealand, Integrated Dataset on Student Loan Scheme borrowers

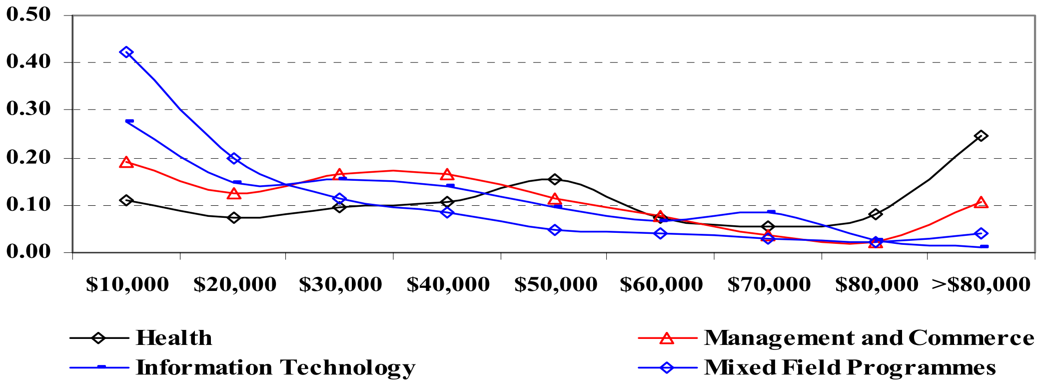
Society and culture is the most popular field with 25 percent, followed by *Management and commerce* with 19 percent. *Architecture and building* recorded the lowest presence of 2 percent. Median earnings more than $30,000 are found in *Health* and *Engineering and related technologies, Management and commerce,* and *Education*. Least median income is recorded for *Mixed Field programme*. This result indicates that the earning potentials are not the only criteria for choosing the fields of study.

Professional fields like *Health*, *Management and commerce*, *Engineering and related subjects,* etc have shown high variability in earnings due to highly varying skill levels within each field. The *Mixed programme studies* has shown the highest coefficient of variation (CV) (117 percent) indicating a high variability in the income distribution. Perusal of the data shows that all doctorate students are categorised under *Mixed programme* by data design. This has resulted in upward bias in the statistics like mean or median resulting in high variability.

Predicted income probability plots with reference to these three groups are depicted in Figures 4.7a to 4.7c. The predicted probability of earnings in 10 income bands is classified into three groups on the basis of their probability distribution pattern for the sake of convenience. The first group consists of fields like *Health, Management and commerce, Information technology* and *Mixed programme*; *Architecture and building, Engineering and related studies* and *Education* formed the second group; the third group comprises *Creative arts, Natural and physical sciences, Agriculture, Society and culture* and *Food, hospitality and personal services*.

In Figure 4.7a, the earning probability level varies considerably in each of the fields in the lower bands, but starts decreasing and flattens in the mid income region. The earnings probability in the lower income region is higher for *Mixed programme* and *Information technology* at 43 percent and 27 percent, which gradually drops after $30,000. The drop is very conspicuous for *Mixed field programme*. At the terminal points the probability becomes negligible for *Mixed field programme* and *Information technology.* The earning probability peaks between $30,000 and $40,000 for *Information technology* and drops off until $60,000 and again peaks at $70,000. It may be noted that *Mixed field programme* are a mix of basic life skills and their likelihood of being in the lower income bands is very high, as evidenced from Figure 4.7a. There is a more than 70 percent chance that a student with *Mixed field programme* is earning an income in the lower income range of $10,000 to $30,000.

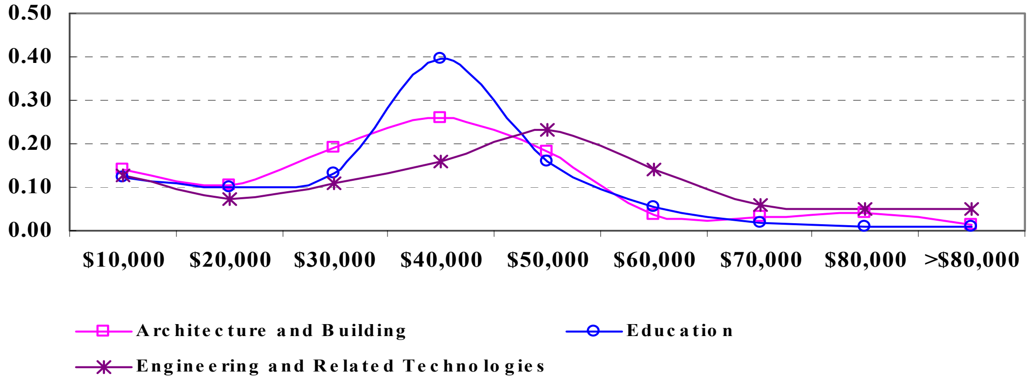
Figure 4.7a: Predicted probability with similar patterns - in selected fields of study: Group 1



The income probability for *Health* studies starts with 10 percent in the lower income bands ($10,000) and gradually increases and peaks at $50,000. Further, the probability decreases until reaching another peak at >$80,000 with 25 percent probability. It is obvious that *Health* studies includes skill levels such as general medicine, dentistry and veterinary sciences that attract higher earnings, as well studies like nursing, pharmacy, radiography, etc which fall in the lower income bands. A similar trend follows for *Management and commerce*. *Management and commerce* studies start off with a lower likelihood of falling in low income bands and end with a higher probability at higher income bands. *Management and commerce* studies peak at around $30,000 to $40,000 and drop more slowly than *Information technology* until $60,000 and again peak at > $80,000. The varying skill levels available within each Group-1 field of study may have caused multiple modes in different income regions as observed in Figure 4.7a.

Figure 4.7b depicts the predicted income probability distribution for Group-2, consisting of three fields of studies covering 17 percent of the students under study. All the three fields start with the same level of probability in the lower region and gradually increase at varying rates until dropping off after $70,000.

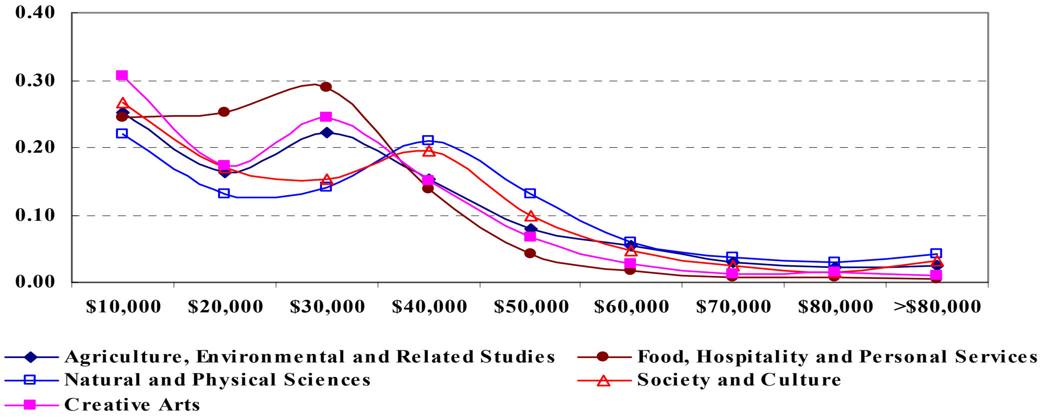
Figure 4.7b: Predicted probability with similar patterns - in selected fields of study: Group 2



Education speaks at $40,000 with 40 percent likelihood, while *Architecture and building* also peak at the same band but with a low of 25 percent. A typical bell shaped curve for *Education* studies indicates that the average income levels are more likely to be in the $30,000 to $50,000 income range (about 70 percent) with a low variability (Table 4.2; coefficient of variation, 55 percent).

Engineering and related technologies picks up more slowly, peaks at $50,000 and then drops steadily but more slowly than the other two. There is a 70 percent chance that an individual with *Engineering and related technologies* would be in the income range greater than $30,000, keeping all other factors constant. *Architecture and building* has a higher probability of earnings in the lower income bands ($10,000 to $50,000), and the earning probability dropped considerably after $50,000 in relation to the other two fields of study. Figure 4.7c depicts the predicted probability of income a student is likely to earn for having taken up any of the five studies listed in Group-3.

Figure 4.7c: Predicted probability with similar patterns - in selected fields of study: Group 3



This group covers about 53 percent of the students under study. The *Food, hospitality and personal services* field shows a very high likelihood (more than 90 percent) of being in the income band $10,000 to $40,000 in relation to other fields of study in this group, with a modal value $30,000. The other two fields, namely *Creative arts* and *Agriculture and related studies,* also show high likelihood of being in the lower income band, peaking at $30,000, but with a lower probability than *Food and Hospitality*. But their low likelihood of being in the higher income bands above $40,000 drops off considerably in relation to *Natural and physical sciences* and *Society and culture* studies.

The *Natural and physical sciences* and *Society and culture* fields have a lower likelihood of being in the lower income region and both show a modal value at $40,000. Their likelihood of earning higher income beyond $50,000 is larger than for the other three fields. *Natural and physical sciences* has a higher likelihood of being in the higher income bands than *Society and culture,* keeping all other factors constant.

The results indicate that the selection of field of study has a statistically significant influence on the earnings, keeping all other factors constant. Higher earning potential is observed in fields of study like *Health*, *Management and commerce*, *Engineering and related stu*dies than others. *Education* has shown consistent earnings with low variation. Areas like *Food, hospitality and personal services*, *Creative arts*, *Agriculture, environment and related studies* have higher probabilities of being in the income band $10,000 to $40,000. It is also observed that lower skilled studies are more likely to be very active in the left side of the graph. Predicted earnings estimated from the predicted probabilities, keeping all other relevant factors constant, are shown in Table 4.3.

Table 4.3: Predicted average earnings in different fields of study

Field of study	Predicted Income ($)
Society and Culture	23,672
Management and Commerce	30,456
Education	28,899
Creative Arts	20,590
Natural and Physical Sciences	25,266
Health	41,852
Food, Hospitality and Personal Services	18,814
Engineering and Related Technologies	33,472
Information Technology	25,166
Mixed Field Programme	18,314
Agriculture, Environment and Related Studies	24,691
Architecture and Building	26,424

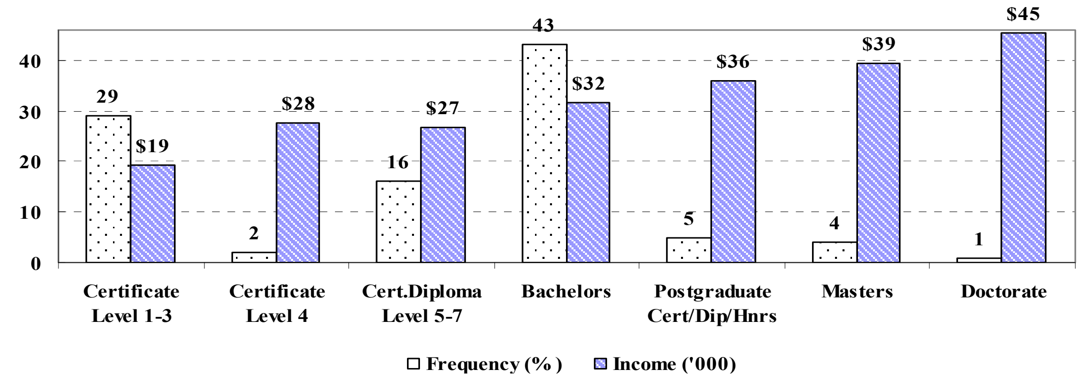
Note: Estimated over three cohort years.
Source: Statistics New Zealand, Integrated Dataset on Student Loan Scheme borrowers

Health, followed by Engineering and Management and commerce, records the highest predicted average earnings. Mixed field programme and Food, hospitality and personal services are the lower earning fields of study. The Mixed field programme consists of studying basic-level skills whereas the other fields of study are supposed to offer specialised skills in addition to the basic skills.

Level of study (qualification level)

This variable is one of the most important study-linked factors influencing the variation in earnings. It has seven qualification levels as shown in Table 3.2 above. The interaction effect of "qualification level by industry" and "qualification level by field of study contributes the most towards the variation in outcome (Appendix 3, Table A3.1) with a very high Wald chi-square value. Higher variability in the data occurs due to the clustering of doctorates in either *Society and culture* or *Mixed field programme,* which may sometimes cause bias in the estimates.

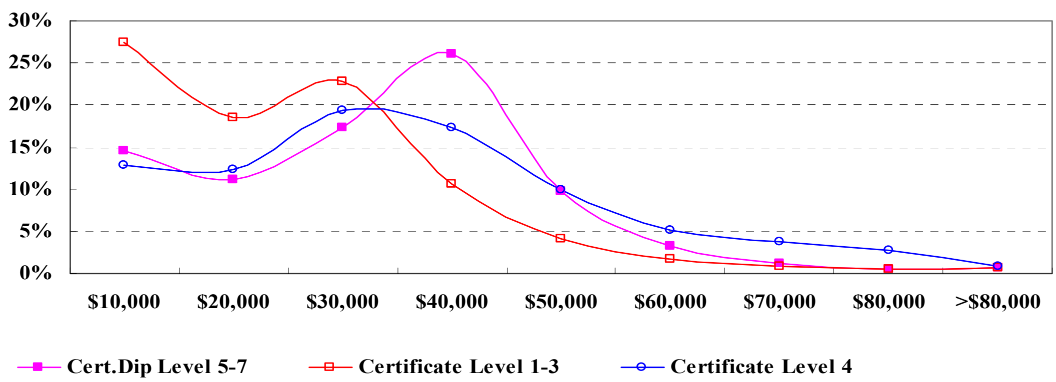
Figure 4.8: Distribution percentage and median earnings due to qualification levels



Note: Students with zero income are excluded
.Source: Statistics New Zealand, Integrated Dataset on Student Loan Scheme borrowers

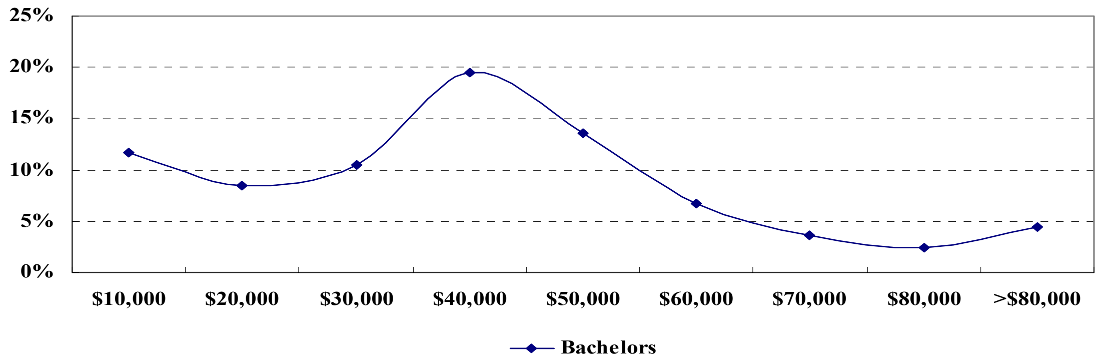
Figure 4.8 shows that the frequency of students in each level of study and their median (average) earnings are influenced by their qualification levels. The higher the qualification level higher will be their income. Approximately 43 percent of the students are enrolled in bachelors degree and about 10 percent are enrolled in postgraduate studies. The graph indicates that students acquiring bachelors or higher qualification have earnings much higher than their sub-degree counterparts. Predicted income probabilities estimated through the generalised logistic regression coefficients for seven levels of study are shown in Figures 4.9a to 4.9g.

Figure 4.9a Predicted probability of earnings for sub-degree qualification levels in different income bands



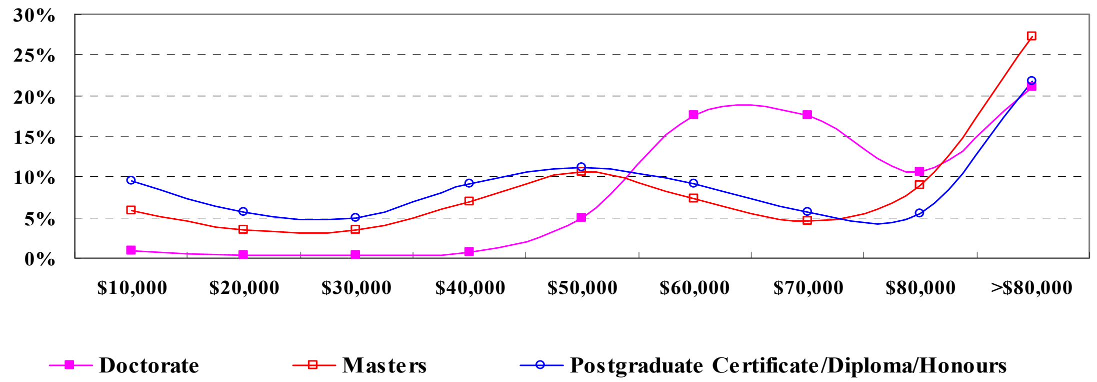
Distribution of earnings probability for sub-degree qualification levels shows peaks and bumps at the left side of the graph (Figure 4.9a). The probability for bachelors degree shows the modal value in the mid income range with lower likelihoods in both the left and right side of the graph (Figure 4.9b).

Figure 4.9b: Predicted probability of earnings for bachelors degree qualification levels in different income bands



The peak at the right side and lower likelihood at the left side of the probability graph is typical for postgraduate qualification levels (Figure 4.9c).

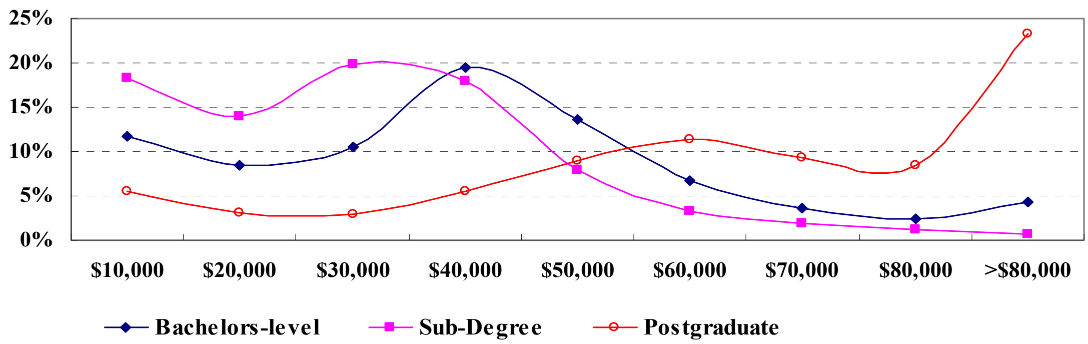
Figure 4.9c: Predicted probability of earnings for postgraduate qualification levels in different income bands



The sub-degree level consists of all certificates and diplomas below degree level, while bachelors level consists of bachelors degree, graduate certificate and diplomas qualifications, and postgraduate level covers all postgraduate qualifications, such as the postgraduate certificate, diplomas, honours, masters and doctorate qualifications. This grouping was necessary to prevent any confidentiality issues, as very small cell counts are encountered for masters and doctorate qualifications in certain sub-groups.

For the sake of comparison, the qualification levels are grouped into three classes. Class-1 comprises all sub-degree qualifications, Class-2 covers bachelors-level and Class-3 has all postgraduate levels. The predicted probability plots of these three groups are shown in Figure 4.10.

Figure 4.10: Predicted probability of earnings for sub-degree, bachelors and postgraduate qualification levels



The overall effect of qualification levels on the earnings indicates that sub-degree qualifications dominate the lower income ranges, while the bachelors degree level occupies the mid income range and the postgraduate qualification lies in the higher income range. Predicted average earnings for sub-degree, bachelors degree and postgraduate qualifications are computed as $22,470, $29,000 and $45,800, respectively (Table 4.4). Earning differential due to qualification levels estimated from the predicted earnings is also shown in the table. The drop in the earnings for certificate holders ranged between 11 and 39 percent over the bachelors degree. Doctorates have earned a premium of 71 percent over bachelors degree while masters degree and postgraduate diploma/honours degree holders show a premium of 58 and 45 percent, respectively.

Table 4.4: Predicted earnings differential due to qualification level (compared with earnings from bachelors degree holders)

Qualification Levels	Average Earnings ($)	Change (percent)
Certificate Level 1-3	17,831	-39
Certificate Level 4	25,857	-11
Cert/ Dip Level 5-7	23,721	-18
Bachelors	29,012	0
PG /Cert/Dip/Honours	42,152	45
Masters	45,826	58
Doctorate	49,502	71

Source: Statistics New Zealand, Integrated Dataset on Student Loan Scheme borrowers

Correspondence analysis carried out to study the propensity of qualification levels to earned income and fields of study is shown in Appendix 8 (Figure A8.1). The correspondence analysis for qualification level by earnings indicates that the higher the qualification the greater the propensity to earn higher income. We can see from the graph that postgraduate qualifications are bunched together in the upper right quadrant along with the high earning ranges. In the right hand lower quadrant bachelors degree is associated with the earning ranges $30,000 to $40,000 and $40,000 to $50,000. Sub-degree qualifications are clustered together around the lower income ranges.

Field of study versus level of study – interaction effects

The interaction effects of only two factors are discussed here. Our focus is to highlight the importance of the interaction effects that are found to be statistically significant in the model.

The odds ratio estimated for qualification levels versus fields of study through the logistic model is compared with the reference category *Bachelors/Advanced diploma* in the field of *Agriculture, environment and related studies*. The nature of income probability distributions for fields of study are grouped into three sub-groups that have similar or near similar trends in interacting with three qualification level sub-group, resulting in nine sub-groups. This was necessary for interpreting results more easily. It may be noted that the group within a qualification level may not match the group elements from the other qualification level. For example, the members of Group-1 in the sub-degree level are not same as Group-1 in the postgraduate levels.

The predicted earnings for three qualification levels under different fields of study are given in Table 4.5. It may be observed that the income increased in each field of study as the qualification level increased.

Table 4.5: Estimated student earnings for qualification level by field of study

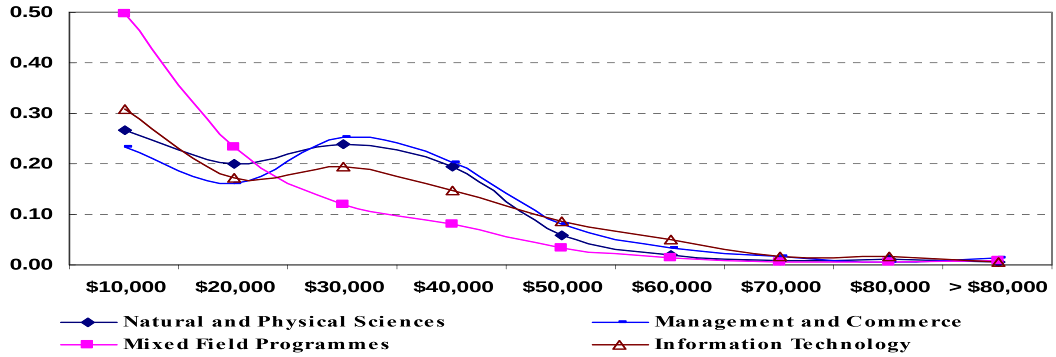
Field of Study	Sub-degree	Bachelors	Postgraduate
Agriculture, Environment and Related Studies	23,053	40,422	46,102
Architecture and Building	28,520	32,661	49,787
Creative Arts	21,439	28,263	44,933
Education	29,581	33,503	50,644
Engineering and Related Technologies	36,275	43,988	50,262
Food, Hospitality and Personal Services	21,518	30,886	71,616
Health	29,571	51,087	74,231
Information Technology	23,693	45,963	60,620
Management and Commerce	25,016	42,474	76,561
Mixed Field Programme	15,541	33,766	53,281
Natural and Physical Sciences	22,468	30,880	41,153
Society and Culture	20,012	29,163	49,463

Predicted income probabilities estimated through the regression coefficients for interaction between fields of study and levels of study are shown in Figures 4.12a to 4.12i.

Sub-degree qualification level versus field of study

Predicted probability plots of qualification levels in different fields of study shown in Figures 4.12a to 4.12i are discussed here. In Figure 4.11a, there are four fields of study included in the sub-degree qualification Sub-group 1, which has a similar earnings distribution pattern. It may be noted that an individual who has acquired a sub-degree qualification in a *Mixed field programme* has a higher likelihood (~ 70 percent) of earning income in the lower income bands below the $40,000 band, relative to other fields of study.

Figure 4.11a: Predicted probability of earnings for sub-degree qualifications in selected fields of study – Group-1



The other three fields of study have shown similar trends in the early income bands, with a slight peak at the $30,000 band and falling at a slow rate until $50,000. *Information technology* has a lower probability of earning in the income range $20,000 to $40,000, but chances are greater in the higher income region than for *Management and commerce* and *Natural and physical science*s.

Figure 4.11b: Predicted probability of earnings for sub-degree qualifications in selected fields of study – Group-2

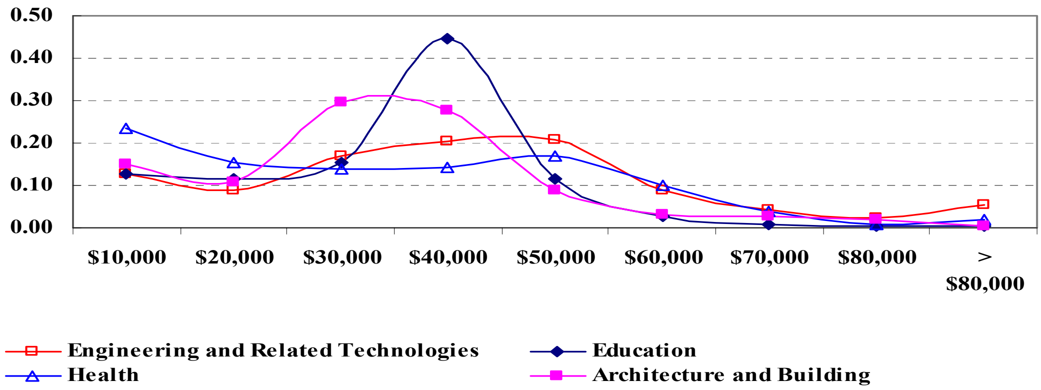
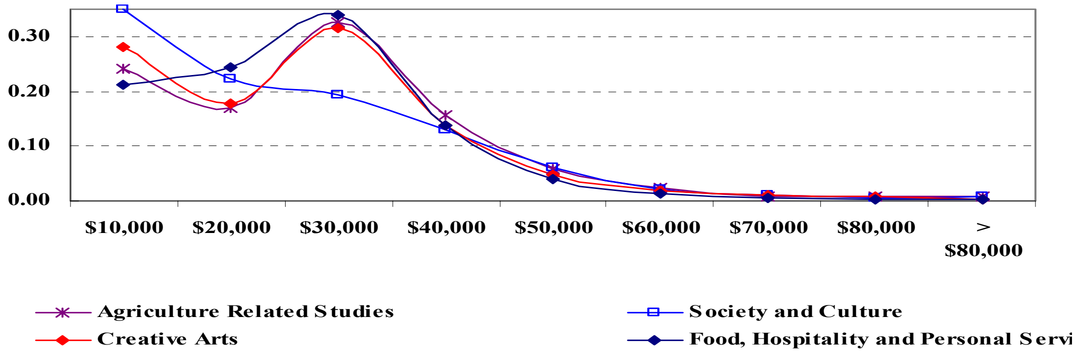


Figure 4.11b has four fields of study in the sub-degree qualification Sub-group 2. Flattened likelihood probability curve is seen for *Engineering* and *Health* studies with a small hump at $50,000. *Health* studies starts with a slightly higher probability than the rest.

Figure 4.11c depicts the probability distribution of students included in Sub-group 3, which covers the remaining four fields of study. These fields have a very high tendency to be in the lower income bands. *Society and culture* shows a steady drop in probability after a high start at $10,000. *Agriculture, Creative arts* and *Food, hospitality and personal services* have shown a similar distribution pattern with a peak prominent at $30,000 and steadily dropping thereafter. All these fields of study have shown higher likelihood of (approximately 70 percent) earnings in the lower income bands below $40,000.

Figure 4.11c: Predicted probability of earnings for sub-degree qualifications in selected fields of study – Group-3



Education (59 percent)*, Engineering and related technologies* (50 percent)*, Health* (41 percent) and *Architecture and building* (40 percent) have a higher likelihood (shown in parentheses) of earnings in the mid range of $30 to $60,000. With the exception of *Health studies* these fields of study also emerge as the high earning fields in the >$60,000 range. There is a very high probability that Mixed field programme (85 percent), Food and hospitality (80 percent), Creative arts (78 percent), Society and culture (77 percent) and Agriculture, environment and related studies (74 percent) are likely to be in the lower income bands < $30,000.

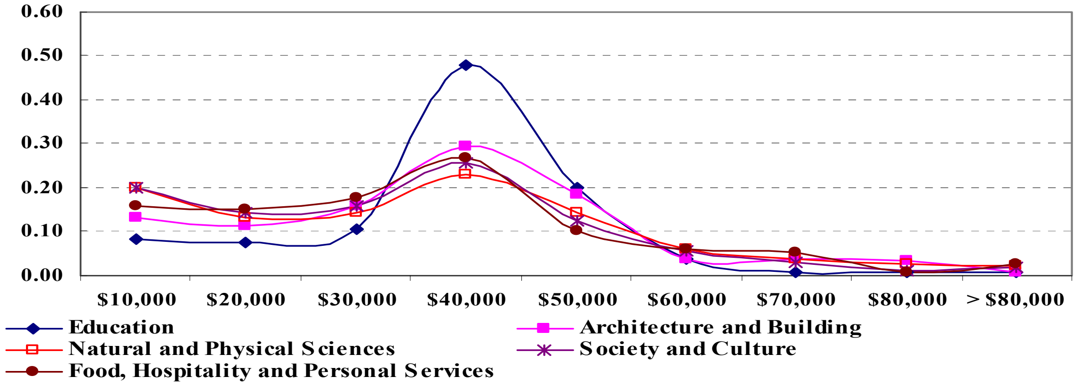
It is clear from the results that sub-degree qualifications in the fields of *Education*, *Management and commerce, Architecture and building*, *Engineering*, *Information technology* and *Natural and physical sciences* have a higher probability that the students will be earning in the mid income range of $30,000 to $60,000. *Agriculture, environment and related studies*, *Creative arts* and *Food and hospitality* have a higher propensity to be in the low income group, the <$30,000 range, for sub-degree qualifications.

Bachelors degree qualification level versus field of study

Figure 4.11d depicts the probability distribution of earnings in different income bands involving five fields of study with a similar distribution pattern that form the first sub-group under bachelors degree. *Society and culture* and *Natural and physical sciences* have a higher probability of earnings in the lower income ranges that drop off at higher bands after raising a peak at $40,000. *Education,* followed by *Architecture and building, Natural and physical sciences* and *Food, hospitality and personal service* start off with different probability levels, increase steadily to peak at $40,000 and then gradually taper off in the higher income ranges.

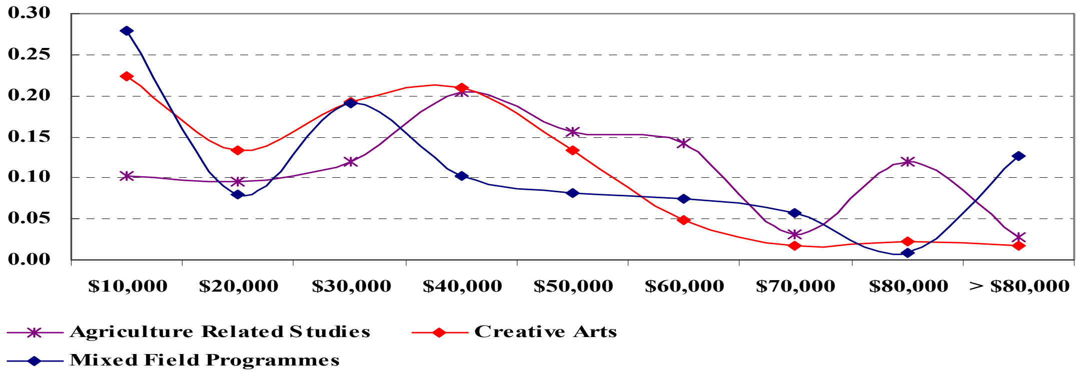
There is a high probability (68 percent) that *Education* studies are likely to be in the income range $30,000 to $50,000, supported by a strong peak indicative of lower variability in the income distribution.

Figure 4.11d: Predicted probability of earnings for bachelors degree qualification in selected fields of study – Group-1



Architecture and building starts with 13 percent at $10,000 and reaches a peak at $40,000, showing better chances of earnings in the higher income ranges than *Education* studies. *Natural and physical sciences* and *Society and culture* also show similar trends but with a higher starting probability than the other fields, also peaking at $40,000 and dropping off steadily. *Society and culture* (50 percent), followed by *Natural and physical sciences* (48 percent) and *Food and hospitality* (49 percent), has a higher likelihood (shown in parentheses) of being in the lower income range ((< $30,000) than the other top two. In the $30,000 to $60,000 income range *Education* studies followed by *Architecture and building, Society and culture, Food and hospitality* and *Natural and physical sciences* have shown a higher likelihood of 72 percent, 53 percent, 44 percent, 43 percent and 43 percent, respectively.

Figure 4.11e: Predicted probability of earnings for bachelors degree qualification in selected fields of study – Group-2



Group-2 under bachelors degree consists of three fields of study (Figure 4.11e). *Mixed field programme* and *Creative arts* studies start with high probability and drop off steeply. Although the general trend shows a drop in the probability, all the fields have multimodal values occurring in the start and end of the graph. In general, *Agriculture, environment and related studies* shows a higher likelihood of earning in the high income region relative to *Creative arts* and *Mixed field programme*.

Figure 4.11f: Predicted probability of earnings for bachelors degree qualification in selected fields of study – Group-3

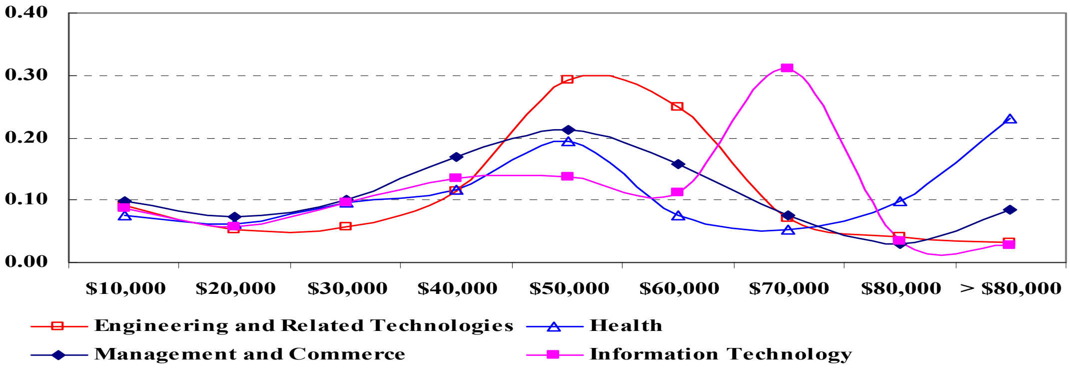


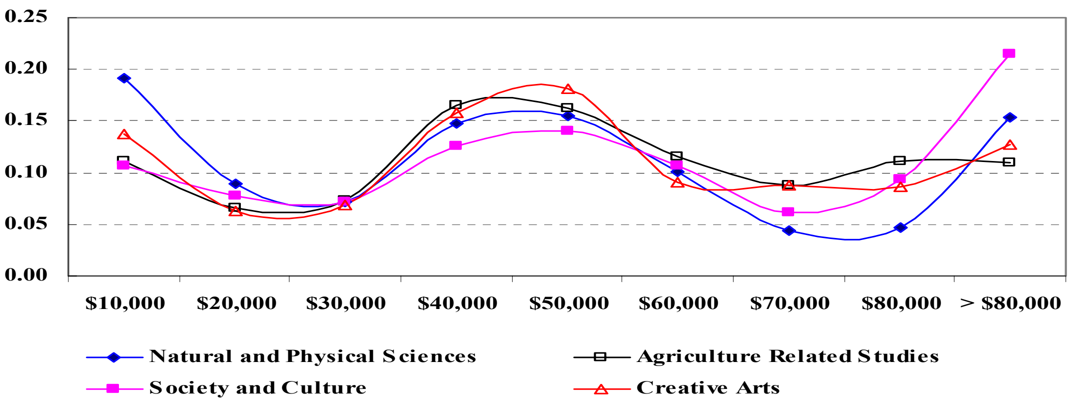
Figure 4.11f above shows Sub-group 3 with the rest of the fields of study. Bachelors degree holders in *Engineering and related technologies* and *Information technology* have a higher likelihood (25 percent) each of having earnings in the high income range of $50,000 and $70,000, respectively. *Health* and *Management and commerce* studies increase steadily, showing a peak at $50,000 with a likelihood of about 15 percent each.

However, *Health* records a higher probability after the $70,000 range, indicating the presence of heterogeneity in the group. The *Management and commerce* field shows a likelihood of about 55 percent in the range $40,000 to $60,000, while the *Engineering* field has 65 percent probability in the same income range. *Health* studies indicate 33 percent probability in the high income band greater than $70,000.

Bachelors degrees in the fields of *Health* studies*, Engineering and related technologies, Information technology, Education, Management and commerce* and *Creative arts* have a greater likelihood of higher earnings than the other fields of study.

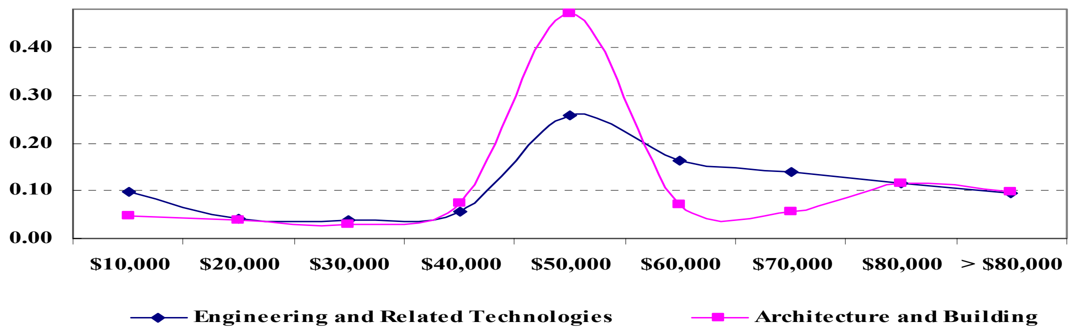
Postgraduate degree qualification level versus field of study
Four fields of study are bunched together in the first group under postgraduate qualification level in Figure 4.11g. *Natural and physical sciences*, *Agriculture studies*, *Creative arts* and *Society and culture* exhibit similar trends with a hump at around $40,000 to $50,000. *Agriculture, environment and related studies* show a higher probability of incidence until $80,000 than the other fields.

Figure 4.11g: Predicted probability of earnings for postgraduate qualification in selected fields of study – Group-1



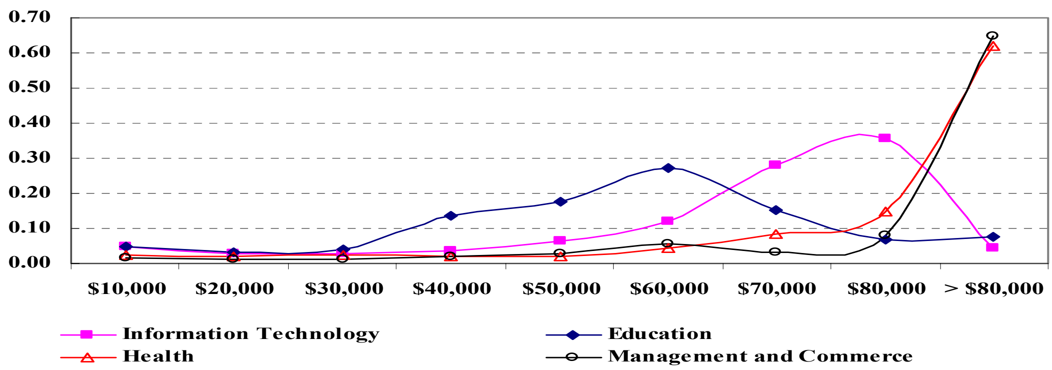
Natural and physical sciences and *Creative arts* start with a cumulative probability of 35 percent and 27 percent, respectively, in the income range below \$30,000, and in the mid income range \$30,000 to \$60,000 they have a cumulative proportion of 43 percent and 40 percent, respectively. However in the high income range of > \$60,000, the proportion is slightly in favour of *Creative arts*. *Agriculture* studies and *Society and culture* start off with the same probability levels in the low income range (25 percent), increasing to the level of 45 percent and 37 percent, respectively, in the mid income range. *Society and culture* overtakes *Agriculture* studies in the high income range. All these fields have shown equitable distribution of earnings in the low and high income ranges with higher probability of earnings in the mid income range.

Figure 4.11h: Predicted probability of earnings for postgraduate qualifications in selected fields of study – Group-2



Only two fields, namely *Engineering and related technologies* and *Architecture and building,* are covered in the second group under the postgraduate qualification level sub-group shown in Figure 4.11h. Postgraduates with *Architecture and building* as their field of study is more likely (about 60 percent) to have their earnings in the mid income range \$40,000 to \$60,000, whereas *Engineering and related technologies* have only 48 percent probability of being in the same income range. However, in the high income range, *Engineering* studies take the lead over *Architecture and building,* with 8 percent probability margin.

Figure 4.11i: Predicted probability of earnings for postgraduate qualifications in selected fields of study – Group-3



In the third group of the postgraduate sub-group we have *Health, Education, Management and commerce* and *Information technology* as the fields of study (Figure 4.11i). *Health*, *Management and commerce* and *Information technology* have a high likelihood of earnings in the high income range of >\$60,000, with 85 percent, 85 percent and 68 percent probability, while *Education* has a 70 percent likelihood of being in the $40,000 to $70,000 income range.

It is clear from the graph that postgraduate qualifications in the fields of *Health*, *Management and commerce* and *Information technology* have a very high probability of having higher earning potential than any other fields.

Correspondence analysis between fields of study and qualification levels, which is shown in Appendix 8 (Figure 8.2), indicates that bachelors degree and postgraduate levels are clustered around the fields *Society and culture, Natural and physical sciences* and *Health studies*. Certificate levels 5 to 7 are closely associated with *Education* studies, while Certificate level 4 has more affinity towards *Management and commerce, Engineering and related technologies* and *Architecture and building.* Certificate levels 1 to 3 are more closely linked to *Information technology*, *Mixed field programme* and *Food, hospitality and personal services*. This trend closely matches the actual trends.

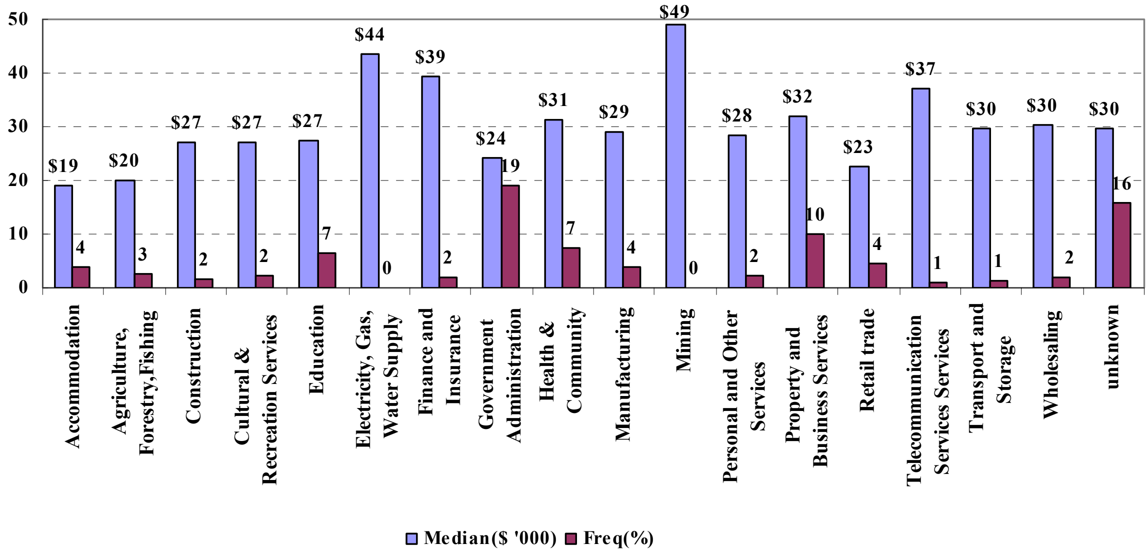
Industry
Earning levels are greatly influenced by the industry in which the individual is working. This variable contributes significantly and is considered another very important variable in determining the earnings of an individual after qualification level and provider sub-sector (Appendix 3, Table A3.1). The importance of this factor is measured from the Wald chi-square value. The international literature has attributed changes in the distribution of earnings to demographic side factors (e.g. education, age, gender), demand side factors (e.g. technological change, international trade), and institutional factors (e.g. industry, labour market regulation).[21] Earning inequalities can be attributable to the industry factor. Varying trends in employment in high wage versus low wage industries, as well as wage rate changes, influence the earnings averages.

Another important factor that influences the earnings is generic and specific skills associated with human capital. Generic skills refers to skills and knowledge that are useful to many employers, while a specific skill is useful to one employer but not to others. This may also include skills that are industry-specific and occupation-specific. This distinction is useful in understanding the incentives for individual workers and employers to pay for education or training. Inclusion of industry as one of the input variables would control for some of the generic and specific skills present in the individuals.

Also the industry-specific skill effects are accounted for by the interaction effects of industry by fields of study, to some extent. The employers associated with the earnings of each student are classified as per ANZSIC96 standard codes (Appendix 4).

[21] Borland, J. (2000). Economic Explanations of Earnings Distribution Trends in the International Literature and Application to New Zealand. Treasury Working Paper 00/16. This paper states that "generally, the extent to which labour market adjustment occurs through the distribution of earnings or through effects on employment appears to depend critically on the employers or institutions present in a country".

Figure 4.12: Frequency and actual median earnings in various industry sectors-average over three cohorts



There are 17 standard industry classifications as per the ANZSIC table. Earning differentials due to industry effects are controlled in the generalised logistic regression model by including industry as one of the explanatory variables. Earning differentials due to industry effects are discussed in this section. Figure 4.12 depicts the frequency of students employed in various industry sectors and their observed median (average) earnings industry-wise. Industry details or the employer details are not available for about 16 percent of the students and hence are classified as *Unknown.* The *Government administration and defence* sector employs the largest number, with 19 percent, followed distantly by the *Property and business services* (10 percent), *Health and community services* (7.5 percent) and *Education* (7 percent) sectors. The smallest number of individuals is employed in the *Mining, Electricity Telecommunication services* sectors; all have a less than 1 percent employment rate. These smallest employers give rise to the highest median earnings. *Mining, Electricity, Telecommunication services* record the highest median income of $49,000, $44,000 and $37,000, respectively. The *Accommodation*, *Agriculture Retail trade* and *Government administration* sectors record the lowest median earnings, below $25,000.

Industry versus field of study

The relative importance of the interaction effect of industry and field of study is next only to the qualification level by provider interaction effect (Appendix 3, Table A3.1), besides the main effect contribution of qualification level. Predicted average earnings are computed from the estimated probabilities for industry by field of study and are given in Table A7.1 (Appendix 7). If we look at the predicted average earnings between industry and field of study, there is no evidence to suggest that the earnings are better in in-field jobs under the assumption that industry is a proxy to the occupation.

The correspondence analysis graph between these two factors is shown in Figure A8.1 (Appendix 8). The correspondence analysis suggests that the industry-focused fields of study are clustered around the corresponding industries referred to as in-field jobs.[22] For example, we could see that Education studies and Education industry (N), Health studies and Health

---

22 Mare, D. and Y. Liang (2006).), Labour Market Outcomes for Young Graduates. Department of Labour. This study reported the outcomes in "in-field" and "out-field" jobs, referring to the employment of an individual in a field that matches the skills attained through education.

and community services (O), Agriculture and related studies and Agriculture industry (A), Food and hospitality study and Accommodation (H), technical skill-related studies (Engineering, Management, Architecture, etc) Electricity and Mining (B, D), Manufacturing (C) are clustered together. Similarly the Management studies are surrounded by related industries like Finance and Insurance (K), and Property and business services (L).

Considerable variation has been observed in trends in the distribution of earnings between specific fields of study and the employer industry. The evidence suggests that the earning differential between industries and different fields of study can be best explained by the range of skills attainable, different levels of educational attainments, demand and supply of skills, out-field jobs and the institutional factors. Among all industries *Electricity, gas and water supply*, *Mining*, and *Telecommunication services* show high predicted average earnings. It may be noted that these industries have small sample sizes and some of the predicted earnings are based on very small numbers of individuals within a field of study.

The overall interaction effect reveals that specific industries prefer some specific fields of study. However, the result did not show any strong evidence to state that industry-specific and occupation-specific skill sets are paid more. Skill shortage may have forced industries to employ people with non-specific skills. A definition of fields of study is too broad and sometimes a wide spectrum of skill sets is included in one group. For example, *Society and culture* has diverging skill sets, such as economics, psychology, law, linguistics, etc. Hence it is hard to find a compatible industry versus fields of study relationship. An examination of the sub-groups of fields of study might yield a more significant relationship between these two factors.

Industry versus qualification level

This interaction effect has the highest contribution to the variability in the outcome as evidenced by the magnitude of the Wald chi-square value. For convenience, the industries are classified into homogenous groups based upon the shape of the probability distribution. Effects of industry in different qualification levels are discussed in detail, whereas the interaction effect of industry and other study-linked factors like fields of study is discussed in brief. It may be noted that the interaction effect of industry by qualification levels and industry by fields of study explains more on the income variability than the industry by level of study.

Logistic regression analysis results indicate that Mining, Finance and insurance, Electricity, gas and water supply, Wholesaling, Health and community services, Property and business services, Manufacturing, Telecommunication services and Transport and storage have a statistically significant and positive influence on income with reference to the base industry category Agriculture and related industries.

For a better interpretation of the regression coefficients or the odds ratio, we have used the predicted probability estimates using graph plots for each industry within each qualification level. The qualification levels are classified into three groups, namely *Sub-degree*, *Degree,* and *Postgraduate*. The *Sub-degree* group consists of three levels of study, namely *Certificate levels 1 to 3*, *Certificate level 4* and *Certificate/Diploma levels 5 to 7*. The bachelors degree qualification alone is grouped into *Degree* and *Postgraduate,* representing all postgraduate degree qualifications.

Predicted probability graphs plotting industry against 10 income bands under each qualification group are further clustered into three to four groups depending upon the distribution style. Sub-degree and degree groups have three graphs each while postgraduate has four graphs. These graphs are depicted in Appendix 5 (Figures A5.2a to A5.2j). The results indicate that individuals with sub-degree qualifications employed in Accommodation, Agriculture, Cultural and Recreation, and Retail trade industries have higher likelihood of earnings in the lower range of income (Appendix 5, Figures A5.2a to A5.2c). Those who are in the Construction industry have better chances of earning more than the rest of the industries identified in this group. Sub-degree holders employed in *Government administration*, *Health and community, Education*, *Property and business*, *Finance and insurance* and *Transportation and storage* have a higher probability of earning income in the mid to high income range. Finance and insurance has the most astute distribution pattern indicative of less variability in the earning pattern, with a 60 percent probability of earnings in the range of $30,000-$50,000. Those acquiring sub-degree qualifications and employed in the *Mining, Electricity, gas and water supply* and *Telecommunication* sectors are likely to earn higher incomes than their counterparts in other industries. The probability curves all have multimodal distribution indicating the variations in earnings depending upon skill levels and experience.

Figures A5.2d to A5.2f (Appendix 5) depicts the predicted earning probabilities of individuals with bachelors degrees in different industries in 10 income bands. The distribution pattern is similar to the one described above for sub-degree qualifications. High wage industries are *Mining*, *Electricity*, *Telecommunication* and *Health and community services.* Bachelors degree holders earning a low income are employed in the *Accommodation, Agriculture, Cultural and recreation services, Retail trade* and *Construction* industries. The rest of the industries are mid wage earning industries.

Under the postgraduate group, *Mining, Electricity, Telecommunication services* have been dropped from the graph (Appendix 5, Figures A5.2g to A5.2j) as these have a very small sample size. Postgraduate holders working in *Finance and insurance, Property and business services, Health and community services, Transportation and storage* and *Manufacturing* are likely to a earn higher income than their counterparts in other industries. The graphs for other industries are highly undulating with multimodal shapes, indicating the variability in the earnings due to skill and experience factors. Average earnings for industry by levels of study are computed using the predicted probability and the mid-values of the income ranges are shown in Table 4.6.

Table 4.6: Predicted average earnings ($) in different industries by levels of study

Industry	Sub-degree	Bachelors	Postgraduate
Accommodation	19,311	22,605	36,428
Agriculture, Forestry, Fishing and Hunting	19,486	24,272	41,989
Construction	27,088	30,400	61,657
Cultural and Recreation Services	22,442	37,417	52,984
Education	24,554	30,115	43,264
Electricity, Gas and Water Supply	58,255	67,442	74,958
Finance and Insurance	36,204	50,753	78,759
Government Administration	19,871	27,511	42,743
Health and Community Services	24,380	47,154	64,248
Manufacturing	26,298	35,975	56,953
Mining	52,930	74,394	74,958
Personal and Other Services	23,632	34,557	56,090
Property and Business Services	25,434	40,532	66,854
Retail Trade	21,862	30,721	45,587
Telecommunication Services	38,189	55,945	81,521
Transport and Storage	27,725	35,711	55,784
Unknown	24,722	38,770	58,393
Wholesaling	26,468	37,601	66,999

Source: Statistics New Zealand, Integrated Dataset on Student Loan Scheme borrowers

Mining, Electricity and power and *Finance and Insurance* topped the industry list for higher earnings with sub-degree and bachelors degree qualifications, while for Postgraduate qualification *Telecommunication industry* topped the list, however closely followed by the industries that are listed as top earning industry in the lower qualification levels. It is important to note that the sample size of sub-groups within higher qualifications in the above industries is also relatively small compared to other industries (Appendix7, Table A7.2)

Completion Status

Completion status refers to whether the person has completed the requirement for a qualification or not. This is a study-related factor, which takes value 0 or 1 indicating whether the qualification was attained or not. The general effect of this variable on income showed a positive effect in the higher income bands and negative effect in the lower income bands. This indicates that overall effect of completion status influences the earnings positively (Appendix -1, Table A1.1), but has a negligible contribution towards the outcome relative to industry, qualification levels, fields of study and provider sub-sectors.

The observed median earnings for the completed and uncompleted groups were $26,534 and $26,247 respectively, aggregated over the three cohorts, controlling for all other factors. With a completion status of about 70 percent in the sample, the earning differential between the completed and uncompleted is very marginal as seen by their median earnings.

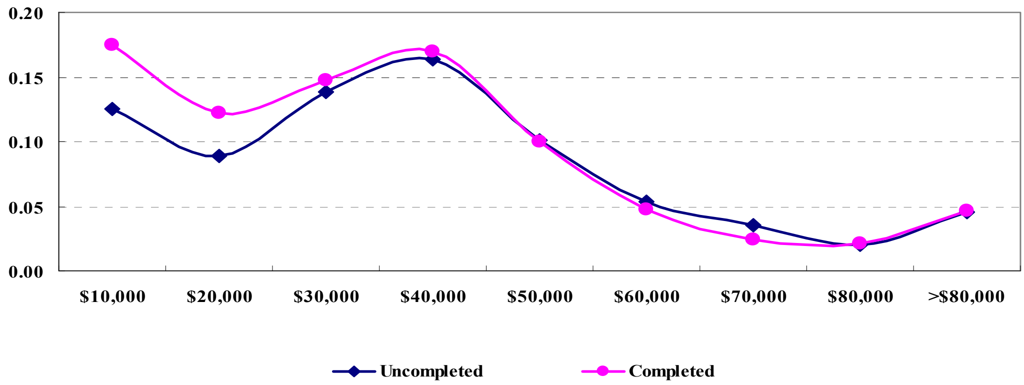
The predicted probability graph shown in fig 4.13 is in agreement with the observations made by Hyatt and Smyth (2005, 2006)[23] [24]. In the graph, the estimated probability is higher for those who have completed the qualification over the uncompleted group in the lower income range. However, the earnings between these two groups narrowed down as the earnings level increased. This is indicative of the fact that the premium for completion is higher while entering the workforce, and obviously in the lower earnings category. However, the premium became negligible in the higher earning regions, as the labour market value

[23] Hyatt, J and R Smyth (2006) 'How do graduates' earnings change over time', Ministry of Education.

[24] Hyatt, J. Gini, P. and R Smyth (2005) Income of Student Loan Scheme borrowers, Ministry of Education, Wellington

experience over the completion, especially at the higher levels of tertiary education. This may also be interpreted as a consequence of non-availability of skilled labour in the labour market.

Fig 4.13: Predicted probability of earnings for completed and uncompleted students



It is well established in human capital theory that earnings rise with work experience, although at a diminishing rate. The synergistic effects of innate ability of an individual and work experience enhances the individual's skills, thereby raising their market value to employers, and thereby marginalising the impact of completion of a qualification in the higher earnings region. However, the premium for completion of a qualification becomes statistically significant when interaction with qualification level is considered. This result corroborates the earlier findings by Roger and Jamie (2005, 2006).

The predicted probability graph shown in Figure 4.13 is in agreement with the observations made by Hyatt et al (2005) and Hyatt and Smyth (2006).[25] [26] In the graph, the estimated probability is higher for those who have completed the qualification than for the uncompleted group in the lower income range. However, the earnings between these two groups narrowed down as the earnings level increase. This is indicative of the fact that the premium for completion is higher while entering the workforce, and obviously in the lower earnings category. However, the premium became negligible in the higher earning regions, as the labour market values experience over completion, especially at the higher levels of tertiary education. This may also be interpreted as a consequence of non-availability of skilled labour in the labour market.

It is well established in human capital theory that earnings rise with work experience, although at a diminishing rate. The increase in earnings with experience is especially pronounced during the first five to 10 years after entering the workforce. The synergistic effects of innate ability of an individual and work experience enhances the individual's skills, thereby raising their market value to employers, and thereby marginalising the impact of education in the higher earnings region.

---

[25] Hyatt, J., P. Gini and R. Smyth (2005). Income of Student Loan Scheme Borrowers. Wellington: Ministry of Education, Wellington.
[26] Hyatt, J. and R. Smyth (2006). How Do Graduates' Earnings Change over Time? Wellington: Ministry of Education..

Prior activity

Prior activity refers to activities engaged in by the student at 1 October in the year before first studying at the most recent provider. Care is needed in interpreting the results of this bit of the analysis as the student's activity need not represent their main activity prior to 1 October. Hence, this variable is less reliable than the other study-related factors, due to this fact. There are 12 categories of prior activity. Although the contribution of this factor is negligible when considered independently, its contribution towards the variation in the earnings was significantly higher in combination with the qualification level as seen from the Wald chi-square value. The frequency distribution and observed median earnings for different activities are shown in Table 4.7.

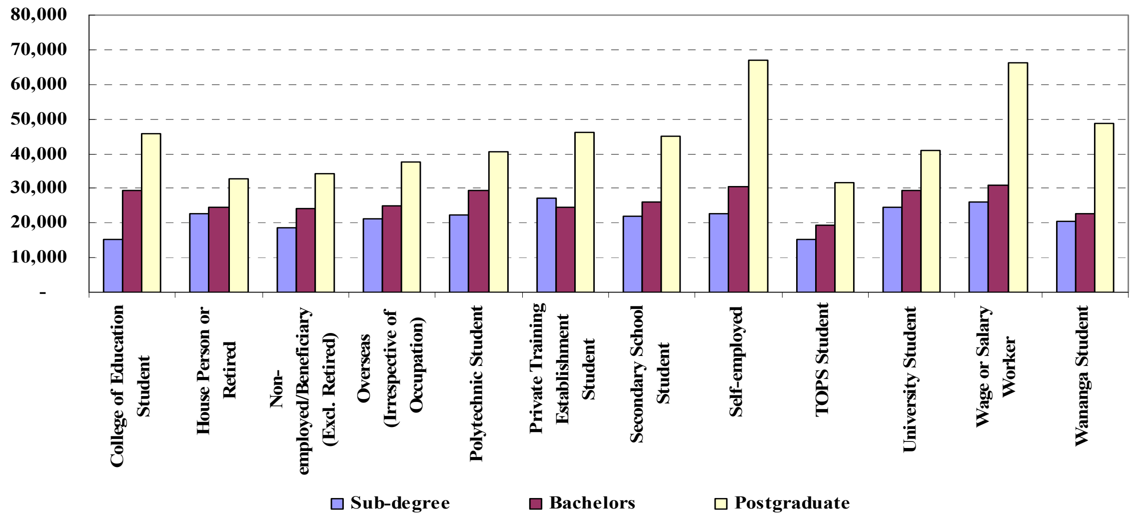
Table 4.7: Earned median income ($) and frequency distribution of prior activities

Prior Activity	Median (average) income ($)	CV (%)	Frequency (%)
College of Education Student	31,530	15	3
House-person or Retired	20,926	30	2
Non-employed/Beneficiary (excl. Retired)	20,530	32	13
Overseas (irrespective of occupation)	27,050	22	3
Polytechnic Student	26,743	26	17
Private Training Establishment Student	22,152	25	1
Secondary School Student	23,868	21	19
Self-employed	31,274	33	2
TOPS Student	17,271	17	1
University Student	32,717	19	19
Wage or Salary Worker	29,871	30	21
Wānanga Student	22,051	21	1

Source: Statistics New Zealand, Integrated Dataset on Student Loan Scheme borrowers

Of the total enrolments aggregated over three the cohorts, 42 percent were tertiary students while 19 percent were studying at a secondary school. About 23 percent were employed either part-time or full-time or were self-employed, while 13 percent were unemployed. Those with prior activity as a university student were earning higher than other groups, closely followed by college of education students and self-employed groups. Higher earning variability is seen in the self-employed group.

Figure 4.14; Predicted average earnings ($) by qualification levels and prior activity



Among those who studied at the bachelors degree level, prior activities of *Self-employed, Wage or salary workers, University students, Private training students* and *College of education students* are likely to earn more than others within the group. Those with *private training establishment student*, *Wage or salary workers* and study at the sub-degree level earned more than others within sub-degree qualification levels.

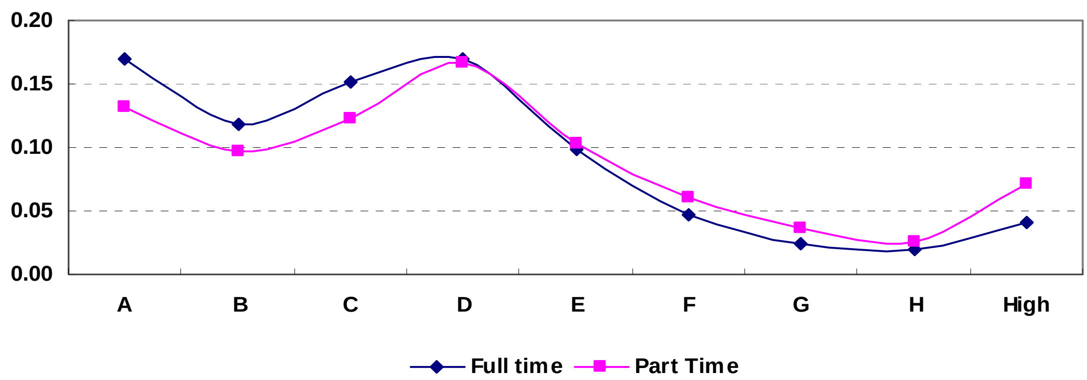
Predicted probability estimates of earnings by qualification levels and prior activities are shown in Appendix 6 (Figures A6.1a and A6.1b). Probability estimates indicate that those with a prior activity of *University student*, *College of education student*, or *self-employed* have a higher probability of being in the upper income bands. The qualifications are grouped as explained in the section *Field of study*.

The overall results reflect the fact that attaining higher qualifications with prior work experience gained prior to or during study certainly boosts earning potential.

Nature of study

Nature of study is another study-linked variable with two categories indicating whether the student is studying full-time or part-time. Part-time is used as the reference category for comparison. This factor has influenced outcomes significantly but with relative importance of being negligible. The interaction effect between qualification levels and nature of study was also statistically significant but of very low importance. The unadjusted median income is higher for part-time ($25,590) than for their full-time ($23,340) equivalents. The average earnings estimated from the predicted probabilities also showed similar trends with $25,730 and $29,560 for full-time and part-time students, respectively, controlling for all other factors in the model.

Figure 4.14; Predicted probability of earnings of full-time and part-time students



The predicted probability distribution of earnings in different income bands is depicted in Figure 4.14. At the left side of the graph high probability is observed for full-time students showing some advantage over part-time students; however this advantage is offset by the higher probability observed for part-time students at the right side of the graph.

The part-timers are 16 percent more likely to earn higher earnings than their full-time counterparts, due to the fact that about 60 percent of the part-timers are either wage or salary workers or self-employed. Many will have entered the workforce early or are older than full-time students. All these factors contribute to their higher earning capacity. Many of those who studied full-time entered tertiary education directly from school and were younger, resulting in lower probability of higher income.

Provider sub-sector

Registered tertiary education organisations are classified into five major provider sub-sectors, representing universities, polytechnics, colleges of education, private training establishments and wānanga. The sub-sector a person studied at influenced the earnings statistically significant with a high relative importance. However, its indirect effect through qualification level is more important, as indicated by the Wald chi-square value.

Unadjusted median (average) earnings, their distribution and variability percentages are presented in Table 4.8. Polytechnics and universities are the largest tertiary education providers, between them covering 93 percent of the total enrolments aggregated over three cohort years. Observed median earnings were highest among those who had been university students, followed by those who studied at a college of education. Those who studied at a college of education had the lowest variability in earnings. The logistic regression coefficients are compared against the reference group *Private training establishments.*

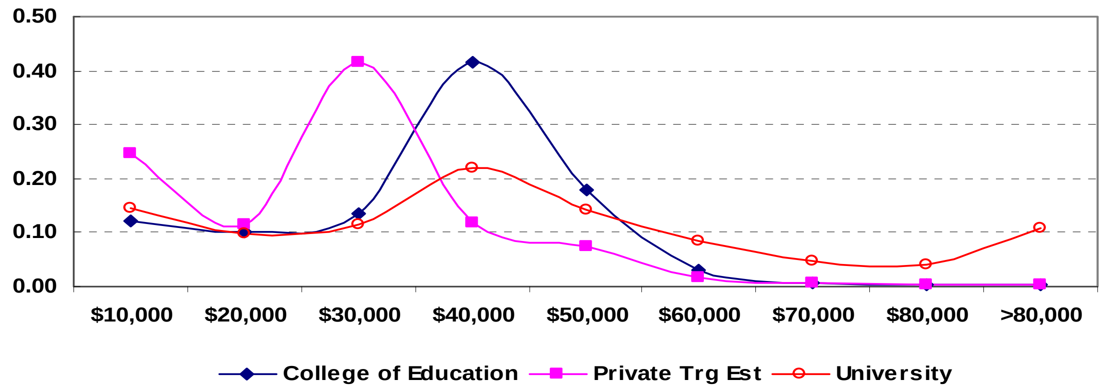
Table 4.8: Median earnings, distribution and variability for tertiary education provider sub-sector

Sub-sector	Median Income ($)	Distribution (%)	CV (%)
College of Education	30,151	6	15
Private Training Est.	23,655	<1	16
Polytechnic	23,168	48	22
University	32,548	45	16
Wānanga	20,124	1	30

The estimated probability for each provider sub-sector is shown in Figures 4.18a and 4.18b; the average earnings estimated from the predicted probabilities are also shown in Table 4.8. The probability plots for interaction effects of provider by qualification level are shown in Appendix 9 (Figures A9.1 to A9.3).

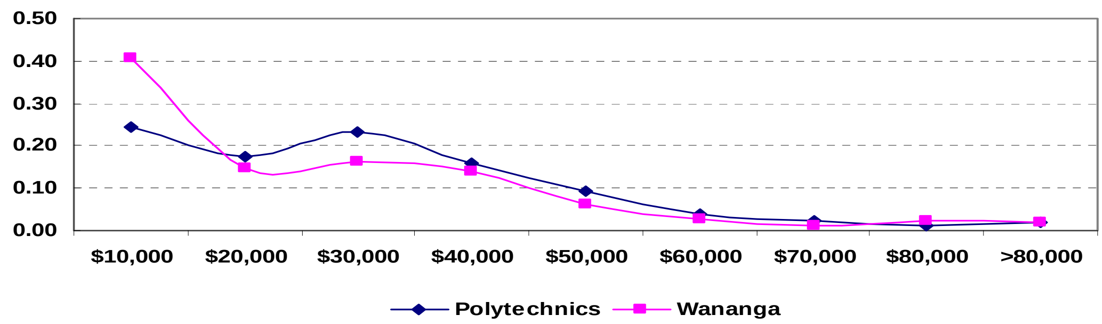
The probability distribution plots for three provider sub-groups (colleges of education, private training establishments, and universities) that have similar distribution patterns are shown in Figure 4.16a. The *College of education* sub-sector graph has a prominent peak at 40 percent in the mid income range. The *Private training establishment* graph has a similar peak at around 40 percent, but at the left side of the graph. The prominence of these modes in the graph suggests the low variability in the earnings. The *University* graph has a bump in the mid region of the graph, but the descent of the graph is slower in the higher income ranges than the other providers. This result indicates that universities have a higher likelihood of earning in the upper income ranges. Those who attended a college of education have a higher probability of earnings in the range of $30,000 - $50,000, while those from a private training establishment have a higher earning probability of earnings in the range $20,000 - $40,000. People from universities showed a flat curve, indicating that probability is distributed across income ranges especially at the right side of the graph.

Figure 4.16a: Predicted probability distribution of earnings due to provider sector



The probability plots for provider sub-sectors *Polytechnics* and *Wānanga* are depicted in Figure 4.16b. The probability plots show a typical low earning distribution pattern with high bumps at the left side of the graph compared with the bumps shown in Figure 4.16a. The result indicates that students from polytechnics are likely to earn more than wānanga students, who have a higher probability at the far left of the range. Predicted probability plots for interaction effects of qualification levels by provider sub-sectors indicate that the distribution pattern is similar to sub-degree, bachelors degree and postgraduate patterns seen elsewhere.

Figure 4.16b; Predicted probability distribution of earnings due to provider sector



The overall distribution pattern is similar to the main effects of provider sub-sectors. The probability plots for postgraduate qualification levels are not shown in these graphs due to the very small cell counts.

EFTS

The EFTS describes the theoretical equivalent full-time study of a qualification in years. This is the only study-linked continuous variable included in the logistic regression analysis. EFTS years is a measure of duration of the person's study in tertiary education. This factor is a statistically significant and has a positive influence on the student earnings, a relative importance of about 2 percent. The analysis indicates that the regression coefficients are positive and statistically significant; explaining that, as the EFTS year increases, the earning level rises in a linear fashion. The regression coefficient for this variable is to be interpreted differently from the other categorical variables. Here, the regression coefficient denotes the rate of change in earnings per unit change in EFTS year. For the sake of convenience, the variable is classified into six groups: less than 0.5 year, between 0.5 and 1 year, 1 to 2 years, 2 to 3 years, 3 to 4 years, and greater than 4 years. This classification is broadly in line with the different qualification levels, as the EFTS years for most levels of study have well-defined EFTS years.[27]

Table 4.9 shows the distribution percentage of students studied in this report, the median earnings and the variability index. The observed median earnings increase with the increase in EFTS years. However, earnings decrease slightly when the EFTS year is greater than 4.

Table 4.9; Median earnings, distribution and variability for tertiary education provider sub-sector

EFTS Years	Median Income ($)	Frequency (%)	CV (%)
< 0.5	26,507	5	19
0.5 to 1	26,358	30	21
1.001 to 2	31,804	14	15
2.001 to 3	32,987	38	13
3.001 to 4	42,679	11	10
> 4	38,219	1	11

The highest frequency was observed for EFTS years between 2 and 3. The next highest frequency is seen in the 0.5 to 1 EFTS year category, while the least participation is in the category of greater than 4 EFTS years. The highest variability in earnings was observed among those who had studied between 0.5 to 1 EFTS years. The high variability in this group could be attributable to the fact that some honours degree holders are classified under this group, as explained earlier; these people are likely to earn a higher income than the certificate holders. Most of the sub-degree-graduate-level qualifications have EFTS duration of less than 2 EFTS years. The estimated probability plots for different EFTS categories are shown in Figures 4.17a and 4.17b.

In Figure 4.17a, we can see that the pattern and trends for the first three categories (less than 0.5, between 0.5 and 1, and between 1 and 2 EFTS years denoted as Group-1) has probability distributions that are inclined on the left side of the graph.

Figure 4.17a: Predicted probability of earnings in different EFTS year groups -Group 1

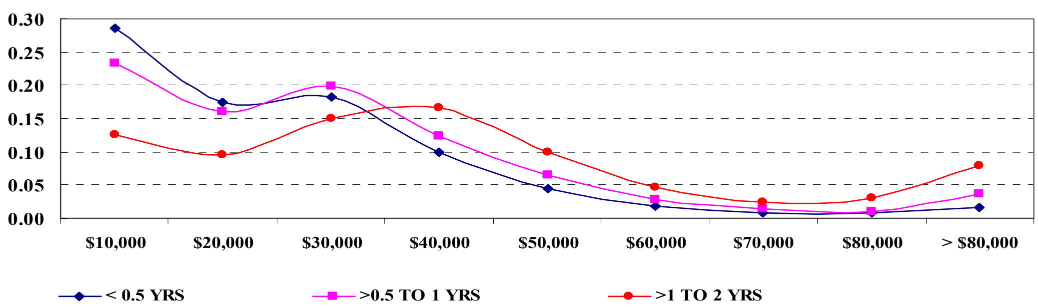
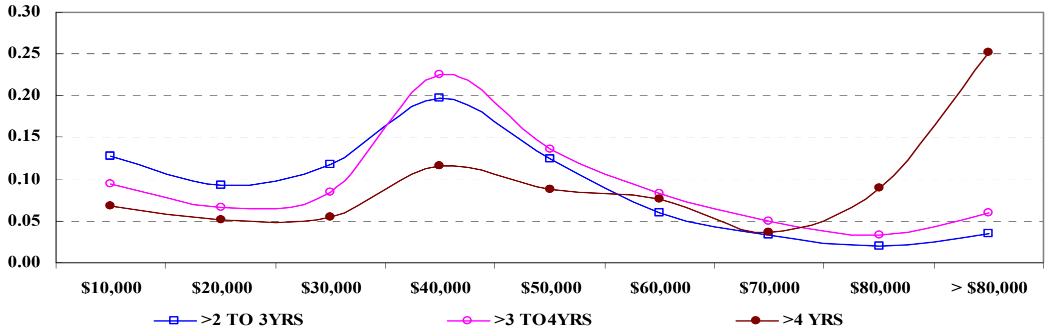


Figure 4.17b shows the estimated probability plots for the last three EFTS year categories, showing peaks at the centre of the plot denoted as Group-2. The greater than 4 EFTS year curve has a lower peak but with higher probability at the right side of the graph. This indicates that higher earnings are more likely when the EFTS years are greater than 4 years.

[27] There are some instances where the length of the qualification is defined differently by different provider institutions, which affects the estimates of earnings. For example, the honours degree is treated as a one-year programme in some universities, while some providers included this qualification with the bachelors degree of four years' duration.

The probability distribution pattern observed here can be equated to the distribution pattern observed in qualification levels. We can equate the low EFTS year students with sub-degree qualifications, the EFTS year ranges with the bachelors degrees and the higher EFTS students with postgraduate qualifications.

Figure 4.17b: Predicted probability of earnings in different EFTS year groups -Group 2



The predicted earnings estimated from the predicted probability distribution for each EFTS year categories are given in Table 4.10. Contrary to the pattern found in observed earnings, the earnings increase monotonically as the EFTS year increases.

Table 4.10: Predicted earnings in EFTS years

EFTS Years	Predicted Earnings ($)
< 0.5	17,534
0.5 to 1	22,158
1.001 to 2	29,045
2.001 to 3	27,429
3.001 to 4	32,859
> 4	45,345

The estimated earnings in different EFTS classes indicate that income is not linearly related to the duration. However, the fact that some qualifications (for instance the honours degree) are treated as a one-one-year programme in some cases and a four-four-year programme in others may have influenced the earnings in the groups that had studied between 1.001 and 2 EFTS, which has higher earnings than the following class.

5. Conclusion

This study forms part of a much wider project on the assessment of investing in tertiary education. This study aims to deliver a partial answer to some of the following policy questions:

- What are the post-study earnings of people with various educational and demographic characteristics?
- Which fields of study lead to the highest incomes over time?
- How much variation exists between the incomes from different levels of study?
- How does tertiary education affect the income gap between men and women?

The main focus of this study is to assess the impact of demographic and study-related variables on student earnings after three years of tertiary learning, using the student information from the Integrated Dataset on Student Loan Scheme borrowers maintained and managed by Statistics New Zealand. Analysis was carried out making use of the student information gathered from about 98,000 students who attended the tertiary education institutions in three cohort years from 1997 to 1999. The generalised logistic regression analysis method was used to model relationships between nominal outcome variables, namely student earnings three years post-study, and the explanatory variables concerning demographic variables and student learning environment.

There are three explanatory variables related to the demographic factors such as age expressed as five-year band groups (10), gender (2) and prioritised ethnicity (6); and eight explanatory variables from study-linked factors, namely qualification level (7), field of study (12), completion status (2), nature of study (2), prior activity (11), provider sub-sector (5), and industry (17) used in the model and the category levels shown in the parentheses. Here the explanatory variable industry is used as a proxy to occupation field, to capture the effects of occupation. EFTS years is the only continuous variable used in this model. The dependent variable, student income, is classified into decile group with a $10,000 range. Logistic regression is fitted to each earning group independently.

Two models were fitted to the data, namely ordinal logistic regression and the generalised logistic regression model. The ordinal regression model was rejected as it did not satisfy the parallel slope assumption. The model statistics and the deviance goodness-of-fit statistics indicate that the model fits the data adequately. The key findings from this study are:

The most important factors that contribute significantly to the variation in the earnings outcome are industry, qualification level, provider sub-sector, and field of study given in the order of merit. However, the interaction effects of qualification by industry, qualification by provider sub-sector, qualification level by prior activity and age by qualification level explain the variations in the earnings more strongly.

The demographic variables are generally not as strongly related to earnings, with the exception of age group, which shows a relatively moderate relationship to student earnings, besides the interaction effect of age by qualification level.

The earnings differential due to sub-degree qualification is about 20 percent lower than the earnings from the bachelors degree student. Those with postgraduate qualifications earned about 58 percent more than the bachelors degree students, keeping all other factors under control.

The premium of completing a qualification assumes significance only when combined with qualification level. For any given level of study, students who complete a qualification earn more than those that don't complete. However, when considered independently of level of study, the premium for completing a qualification is only marginal.

Prior activity influences the earnings significantly but the magnitude of its influence is relatively negligible in comparison to its interaction effect with qualification level. For the student with a prior activity as self-employed and wage or salary worker, the likelihood of earnings in the higher income range is very high.

Recommendations for further investigation

Labour market experience is unobserved in this dataset and is proxied by measures such as age and prior activity. However, individuals with the same number of years of education and potential labour market experience may have substantially different skills - depending on their family environment, their fields of study, their work experience and on-the-job training, and other factors. More generally, education and work experience are inputs into the earnings model, and not direct measures of the outcomes - a set of skills, competencies and knowledge.

Under these circumstances, it is appropriate to include direct measures of skills in an equation explaining earnings or other labour market outcomes. However, it is also appropriate to include traditional human capital variables such as educational attainment and labour market experience, because these control for the influence of unobserved skills.

In continuing research, it will be helpful to include identifying variables that measure the outcomes like skills, competencies and perseverance Adding variables like parental income (using benefit data as a proxy to parental income) as well as other data on occupation and work experience will be also improves the robustness of the estimation. Inclusion of data on students who do not use the student loan scheme would expand the scope of these results. Additional years of income data may also strengthen the results from this study. Furthermore, it is important to note that the results from this type of study may be used as an input in the cost-benefit analysis.

Appendix 1: Generalised multivariate logistic regression model

Logistic regression analysis is often used to investigate the relationship between discrete responses and a set of explanatory variables. Several texts that discuss logistic regression are Collett (1991), Agresti (1990), Cox and Snell (1989), Hosmer and Lemeshow (2000), and Stokes, Davis and Koch (2000). These models were introduced by McFadden (1974) as the *discrete choice* model, and they are also known as *multinomial* models. The logistic model fits linear logistic regression models for discrete response data by the method of maximum likelihood. The maximum likelihood estimation is carried out in this study with the Newton-Raphson algorithm. The logit link function in the logistic regression models is the generalised logit function. The logistic model enables us to specify categorical variables or continuous variables as explanatory variables. We can also specify more complex model terms such as interactions and nested terms in the same way as in the Generalised Linear Model method. Any term specified in the model is referred to as *effect*, whether it is a continuous variable, a categorical variable or an interaction.

Polytomous logistic regression can be divided into two cases: ordinal response (the dependent variable is ordinal; for example, income variables can be grouped - 10,000, 20,000, 30,000, etc), and nominal response (the dependent variable is a nominal categorical variable, for example income band A, income band B, income band C, etc, instead of numeric values). If categories are unordered, the model essentially fits a binary logistic regression to each category, i.e. finds Probability (in category) and Probability (not in category). If categories are ordered, for each category j, say we can find Probability (in a category greater than j) and Probability (in a category less than or equal to j).

Ordinal response

For ordinal response, cumulative logits can be modelled with the proportional odds model. The proportional odds model assumes that the cumulative logits can be represented as parallel linear functions of independent variables; that is, for each cumulative logit the parameters of the models are the same, except for the intercept.

For example, suppose that the dependent variable Y takes three values, namely 1 (10,000), 2 (20,000) and 3 (30,000) and let $\mathbf{p_1 = P(Y=1), p_2 = P(Y=2)}$ **and** $\mathbf{p_3 = P(Y=3)}$. The ordinal logistic regression models the relationship between the cumulative logits of Y, that is: **log(p1/(1-p1))=log(p1/(p2+p3))** and **log((p1+p2)/(1-(p1+p2))) =log((p1+p2)/p3)** in this case, and independent variables. The model assumes a linear relationship for each logit and parallel regression lines,

$$\log(p1/(1-p1)) = \alpha_1 + \beta_1 * X_1 + \beta_2 * X_2 + \ldots + \beta_k * X_k$$
$$\log((p1+p2)/p3) = \alpha_2 + \beta_1 * X1 + \beta_2 * X_2 + \ldots + \beta_k * X_k$$

that is, the intercepts are different, but the remaining regression parameters are the same. It is easy to see that the odds **p1/(1-p1)** and **(p1+p2)/p3** are proportional,

$$(p1/(1-p1)) = e^{\alpha 1} e^{\beta 1 * X1 + \beta 2 * X2 + \ldots + \beta k * Xk,,}$$
$$((p1+p2)/p3) = e^{\alpha 2} e^{\beta 1 * X1 + \beta 2 * X2 + \ldots + \beta k * Xk,,} = const*(p1/(1-p1),$$

where const= $\boldsymbol{\alpha_2 / \alpha_1}$, hence the name the proportional odds model.

Proportional odds imply that odds ratios for Y being 10,000 (1) vs 20,000 or 30,000 (2 or 3) and for Y being 10,000 or 20,000 (1 or 2) vs 30,000 (3) are the same. The maximum likelihood estimation is used to obtain the estimates of the model parameters. After estimators of α1, α2, β1, β2,…,βk are computed, it is easy to compute predicted probabilities using the following formulas derived from the equations above.

$p1 = [e^{\alpha 1 + \beta 1*X1 + \beta 2*X2 + ... + \beta k*Xk}] / [1+ e^{\alpha 1 + \beta 1*X1 + \beta 2*X2+... + \beta k*Xk}]$
$p1+p2 =[e^{\alpha 2+ \beta 1*X1 + \beta 2*X2 + ... + \beta k*Xk}] / [1+ e^{\alpha 2 + \beta 1*X1 + \beta 2*X2 + ... + \beta k*Xk}]$
$p3 = 1-(p1+p2)$.

If parameter βi is positive, then p1 and predicted probability of (Y=1, 10,000), as well as cumulative probability of (Y=1, 10,000, or Y=2, 20,000), p1+p2, are higher for higher values of Xi. If parameter βi is negative, p1 and p1+p2 are lower for higher values of Xi.

Nominal response

If the proportional odds (parallel regression lines) assumption is not satisfied, then the generalised logits approach can be used to model the relationship between the response and the independent variables. The generalised logit regression models are also used when the response variable is nominal. For a categorical variable, the generalised logits are defined as natural logarithm, log, of the probability of each category over the probability of the last response category. These generalised logits are modelled as linear functions of independent variables with different regression parameters for each logit (not only intercepts as in the ordinal logistic regression, but all parameters are different).

For example, suppose that the dependent variable Y takes three category values 1 (income band A), 2 (income band B) and 3 (income band C) and let $p_1 = P(Y=1)$, $p_2 = P(Y=2)$ and $p_3 = P(Y=3)$. The generalised logits model the relationship between the generalised logits, log(p1/p3) and log(p2/p3) in this case, and independent variables. The model assumes a linear relationship for each logit,

$\log (p1/p3) = \alpha_1+ \beta_{11}*X_1 + \beta_{12}*X_2 + \beta_{13}*X_3,$
$\log (p2/p3) = \alpha_2+ \beta_{21}*X_1 + \beta_{22}*X_2 + \beta_{23}*X_3,$

that is, all regression parameters are different. The maximum likelihood estimation is used to obtain the estimates of the model parameters. After estimators of α_1, α_2, β_{11}, β_{12}, β_{13}, $\beta_{21,}$ β_{23} are computed, it is easy to compute predicted probabilities using the following formulas derived from the equations above.
$p_1 = e^{h1} / (1+ e^{h1} + e^{h2}),$
$p_2 = e^{h2} / (1+ e^{h1} + e^{h2}),$
$p_3 = 1- (p1+p2),$

where $h_1 = \alpha_1+ \beta_{11}*X_1 + \beta_{12}*X_2 + ... + \beta_{1k}*X_k$, and $h_2 = \alpha_2+ \beta_{21}*X_1 + \beta_{22}*X_2 + ... + \beta_{2k}*X_k$ These predicted probabilities are estimated using SAS logistic regression procedure with glogit as link function in SAS v.8.3. As the Integrated Dataset on Student Loan Scheme borrowers was analysed it was assumed to follow the proportional odds model. But this assumption is not satisfied and hence it fits the generalised multinomial logistics model. The design of the fitted model is of the form:
log (p_j / p_{k+1}) = β_1 + β_2 age **+ β_3** gender **+ β_4** ethnicity **+ β_5** level of study **+ β_6** field of study **+ β_7** prior activity **+ β_8** provider **+β_9** nature of study **+ β_{10}** completion status **+ β_{11}** industry **+ β_{12}** EFTS years **+ β_{13}** age * level of study **+ β_{14}** age * field of study **+ β_{15}** age * completion status **+ β_{16}** gender * level of study**+ β_{17}** gender * field of study **+ β_{18}** gender * completion

status + $\boldsymbol{\beta}_{19}$ ethnicity * level of study + $\boldsymbol{\beta}_{20}$ ethnicity * field of study + $\boldsymbol{\beta}_{21}$ ethnicity * completion status + $\boldsymbol{\beta}_{22}$ level of study *completion status + $\boldsymbol{\beta}_{23}$ field of study * level of study + $\boldsymbol{\beta}_{24}$ industry * level of study + $\boldsymbol{\beta}_{25}$ industry * field of study + $\boldsymbol{\beta}_{26}$ prior activity * level of study + $\boldsymbol{\beta}_{27}$ prior activity * field of study + $\boldsymbol{\beta}_{28}$ provider * level of study + $\boldsymbol{\beta}_{29}$ provider * field of study

where
age, gender, ethnicity, etc are dummy variables and are called the main effects
age **,*** level of study, gender, * field of study, ethnicity,* and completion status are the interaction effects
β_1 is the log odds for no characteristics involved or constant
β_2, β_3 ... are the increment in log odds for being age, ethnicity, etc from the reference category.
Thus, the parameters defined here are *incremental effects*. The intercept corresponds to a reference group (β_1); the other parameters are incremental effects for the other groups compared with the reference group.

Predicted probabilities

For plotting and interpreting results from logistic regression, it is usually more convenient to express fitted values on the scale of probabilities.

Let the logistic model of the form,

log $(\pi/(1-\pi))$ = log(odds) = logit = $\alpha_1 + \beta_1{*}X_1 + \beta_2{*}X_2 + \beta_3{*}X_3$, then the inverse transformation of the logistic function gives the predicted probability,

hence π=Probability (Y=earnings outcome, given that X=x)

$$\pi = \frac{e^{\alpha + \beta_1 \mathbf{X}_1 + \beta_2 \mathbf{X}_2 + \beta_3 \mathbf{X}_3}}{1 + e^{\alpha + \beta_1 \mathbf{X}_1 + \beta_2 \mathbf{X}_2 + \beta_3 \mathbf{X}_3}}$$

where π is the probability of the outcome of interest.

Diagnostic testing

Any observation with missing values for the response and explanatory variables is excluded from the analysis. The estimated linear predictor and its standard error estimate, the fitted probabilities and confidence limits, and the regression diagnostic statistics are not computed for any observation with missing offset or explanatory variable values. However, if only the response value is missing, the linear predictor, its standard error, the fitted individual and confidence limits for the probabilities can be computed.

For the generalised logit model the income band A is the reference level as displayed in the Response profiles table. The reference category for the explanatory variables is given in the parameter estimation table.

The following generalisation of the coefficient of determination proposed by Cox and Snell (1989) is used here to assess the explanatory power of the model,

$R^2 = 1 - \{L(0) / L(\theta)\}^{[2/n]}$

where $\boldsymbol{L}(\mathbf{0})$ is the likelihood of the intercept-only model, $\boldsymbol{L}(\boldsymbol{\theta}\}$ is the likelihood of the specified model, and $\boldsymbol{n}$ is the sample size. The quantity $\boldsymbol{R^2}$ achieves a maximum of less than one for discrete models, where the maximum is given by $\boldsymbol{R_{max}}^2 = \mathbf{1} - \{\boldsymbol{L}\ (\mathbf{0})\}^{[2/n]}$.An adjusted $\boldsymbol{R^2}$ proposed by Nagelkerke (1991) is given as: $\boldsymbol{R^2} = R^2/(R^2_{max})$.

Testing the parallel lines assumption

For an ordinal response, the logistic model performs a test of the parallel lines assumption, and is labelled as *Score Test for the Proportional Odds Assumption* in the output. The model fitted using the Proportional Odds model showed a statistically significant score test, resulting in the rejection of the parallel assumption. Generalised multinomial model was fitted as an alternative to the ordinal logistic model treating response class as nominal.

Goodness-of-fit test

Model Fit Statistics and Testing Global Null Hypothesis: BETA=0 tables give the various criteria (-2 Log Likelihood, Akaike Information Criterion (AIC), Schwarz Criterion (SC)) based on the likelihood for fitting a model with intercepts only and for fitting a model with intercepts and explanatory variables. The third column of the table gives the chi-square statistics and ***p***-values for the -2 Log L statistics and for the Score statistic. These test the joint effect of the explanatory variables included in the model. For a generalised logit model, iterative maximum likelihood algorithms, namely Newton-Raphson technique is used.

Over-dispersion

For a correctly specified model, the Pearson chi-square statistic and the deviance, divided by their degrees of freedom, should be approximately equal to 1. The fitted generalised model showed that the deviance is equal to one holding the binomial variability indicating that the data is not exhibiting over-dispersion.

The Type III Analysis of effects table gives the Wald chi-square statistic, the degrees of freedom, and the ***p***-value for each effect in the model. The Analysis of maximum likelihood estimates table, which includes parameter name, also indicates the variable level and a response variable column to identify the corresponding logit by displaying the non-reference level of the logit. It also contains the maximum likelihood estimate of the parameter, the estimated standard error of the parameter estimate, the Wald chi-square statistic (Wald statistics - similar to t-statistics), computed by squaring the ratio of the parameter estimate divided by its standard error estimate, and the ***p***-value of the Wald chi-square statistic with respect to a chi-square distribution with one degree of freedom. The Association of predicted probabilities and observed responses table includes a breakdown of the number of pairs with different responses, and four rank correlation indexes: Somers *D*, Goodman-Kruskal Gamma, and Kendall's Tau-*a*, and *c*, confidence intervals for all the parameters.

Appendix 2: Correspondence analysis

Correspondence analysis is a descriptive, exploratory technique designed to analyse simple two-way and multi-way contingency tables containing some measure of correspondence between rows and columns. In a typical correspondence analysis, a cross-tabulation table of frequencies is first standardised to obtain relative frequencies of all cell sums to 1. The objective of this analysis is to identify the distance between rows and columns in a low-dimensional space. Correspondence analysis starts with tabular data, usually two-way cross-classifications, though the technique is generalisable to n-way tables with more than two variables. The variables must be discrete: nominal, ordinal, or continuous variables segmented into ranges. The technique defines a measure of distance between any two points, where points are the values (categories) of the discrete variables. Since distance is a type of measure of association (correlation), the distance matrix can be the input to principal components analysis, just as correlation matrices may be the input for conventional factor analysis. However, where conventional factor analysis determines which variables cluster together, correspondence analysis determines which category values are close together. This is visualised on the correspondence map, which plots points (categories) along the computed factor axes. Some of the characteristics of correspondence analysis are:

Correspondence analysis is an appropriate method for the analysis of categorical data;

- It produces a visual representation of the relationships between the row categories and the column categories in the same space.
- It avoids the unease of using traditional multivariate techniques such as factor analysis on such data.
- The technique is versatile: it can be used with frequency data, with percentages, with data in the form of ratings and with heterogeneous datasets.
- Correspondence analysis can suggest unexpected dimensions and relationships in the tradition of exploratory data analysis even if, in this post-empiricist age, no-one expects theory to emerge automatically from the data.
- Although non-parametric itself, the results of correspondence analysis are often a useful preliminary to a more structured and traditional multivariate modelling of categorical data.
- Correspondence analysis mapping is used as an explanatory device in this context.

Appendix 3: Type III Analysis effects and model fit statistics

Table A3.1: Type III Analysis effects and relative importance

Effect	DF	Chi-Square(Wald)	Pr > Chi-Square
Age Group	81	44961220	<.0001
Gender	9	4806562	<.0001
Ethnicity	45	4894837	<.0001
Year	18	1833180	<.0001
Qualification Level	54	735708082	<.0001
Field of Study	99	414976463	<.0001
EFTS Years	9	71284711	<.0001
Nature of Study	9	1813263	<.0001
Completion Status	9	18816087	<.0001
Prior Activity	108	15075521	<.0001
Industry	162	283340589	<.0001
Provider Sub-sector	36	524622185	<.0001
Age * Qualification Level	475	142883765	<.0001
Gender * Qualification Level _	54	3977713	<.0001
Gender * Completion Status	9	2043719	<.0001
Gender * Ethnicity	45	738850	<.0001
Qualification Level * Field of Study	504	81620786	<.0001
Qualification Level * Completion Status	54	8359698	<.0001
Qualification Level * Ethnicity	270	12504728	<.0001
Qualification Level * Nature of Study	54	14756888	<.0001
Field of Study * Completion Status	99	26802791	<.0001
Ethnicity * Completion Status	45	3872167	<.0001
Qualification Level * Provider Sub-sector	180	923347399	<.0001
Year * Completion Status	18	3452334	<.0001
Qualification Level * Industry	943	96861981	<.0001
Fields of Study * Industry	1754	728034265	<.0001
Prior Activity * Qualification level	630	33850141	<.0001

Table A3.2: Model fit statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	882457	729463
SC	882542	784305
-2 Log L	882439	717893

R-Square=	0.8174	Max-rescaled	R-Square =	0.8175

Deviance and Pearson Goodness-of-Fit Statistics

Criterion	DF	Value	Value/DF	Pr > Chi Sq
Deviance	5.93E+05	260162.153	0.4385	1
Pearson	5.93E+05	1.39E+16	2.35E+10	<.0001

Table A3.3: Contribution of effects to the variation in outcome

Effect	Contribution (%)
Qualification Level * Industry	30.60
Qualification Level * Provider Sub-sector	12.60
Fields of Study * Industry	12.30
Industry	11.50
Qualification Level	10.70
Prior Activity * Qualification Level	9.20
Age Group * Qualification Level	4.50
Qualification Level * Field of Study	2.61
Provider Sub-sector	1.95
Field of Study	1.12
Qualification Level * Ethnicity	1.08
Qualification Level * Nature of study	0.38
Age Group	0.34
Field of Study * Completion Status	0.25
Qualification Level * Completion Status	0.21
Completion Status	0.16
Gender * Qualification Level * Completion Status	0.14
Prior Activity	0.11
Ethnicity	0.07
Year * Completion Status	0.06
Gender	0.04
Year	0.02
Gender * Completion Status	0.02
Nature of study	0.01
Gender * Ethnicity	0.01
Ethnicity *Completion Status	0.01

Appendix 4: ANZSIC codes and NZSCED codes

ANZSIC codes

Division A - Agriculture, Forestry, Fishing and Hunting
Division B - Mining
Division C - Manufacturing
Division D - Electricity, Gas and Water Supply
Division E - Construction
Division F - Wholesale Trade
Division G - Retail Trade
Division H - Accommodation, Cafes and Restaurants
Division I - Transport and Storage
Division J - Communication Services
Division K - Finance and Insurance
Division L - Property and Business Services
Division M - Government Administration and Defense
Division N - Education
Division O - Health and Community Services
Division P - Cultural and Recreational Services
Division Q - Personal and Other Services

NZSCED fields of study codes (broad and narrow levels)

Natural and Physical Sciences

Mathematical Sciences
Physics and Astronomy
Chemical Sciences
Earth Sciences
Biological Sciences
Other Natural and Physical Sciences

Information Technology

Computer Science
Information Systems
Other Information Technology

Engineering and Related Technologies

Manufacturing Engineering and Technology
Process and Resources Engineering
Automotive Engineering and Technology
Maritime Engineering and Technology
Mechanical and Industrial Engineering and Technology
Civil Engineering
Geomatic Engineering
Electrical and Electronic Engineering and Technology
Aerospace Engineering and Technology
Other Engineering and Related Technologies

Architecture and Building

Education

Teacher Education
Curriculum and Education Studies
Other Education

Management and Commerce

Business and Management
Sales and Marketing
Tourism
Office Studies
Banking, Finance and Related Fields
Other Management and Commerce

Society and Culture

Political Science and Policy Studies
Studies in Human Society
Human Welfare Studies and Services
Behavioural Science
Law
Justice and Law Enforcement
Librarianship, Information Management and Curatorial Studies

Architecture and Urban Environment
Building

<u>Agriculture, Environmental and Related Studies</u>

Agriculture
Horticulture and Viticulture

Forestry Studies
Fisheries Studies
Environmental Studies
Other Agriculture, Environmental and Related Studies

<u>Health</u>

Medical Studies

Nursing
Pharmacy
Dental Studies
Optical Science
Veterinary Studies

Public Health
Radiography
Rehabilitation Therapies
Complementary Therapies
Other Health

Language and Literature
Philosophy and Religious Studies
Economics and Econometrics
Sport and Recreation

Other Society and Culture

<u>Creative Arts</u>

Performing Arts
Visual Arts and Crafts
Graphic and Design Studies

Communication and Media Studies
Other Creative Arts

<u>Food, Hospitality and Personal Services</u>

Food and Hospitality
Personal Services

<u>Mixed Field Programme</u>

General Education Programme
Social Skills Programme
Employment Skills Programme
Other Mixed Field Programme

<u>Note: n.e.c. = not elsewhere classified</u>

Appendix 5: Predicted probability of qualifications by industry

Figure A5.1a: Predicted probability of earnings of sub-degree qualifications by 21-45 year age groups

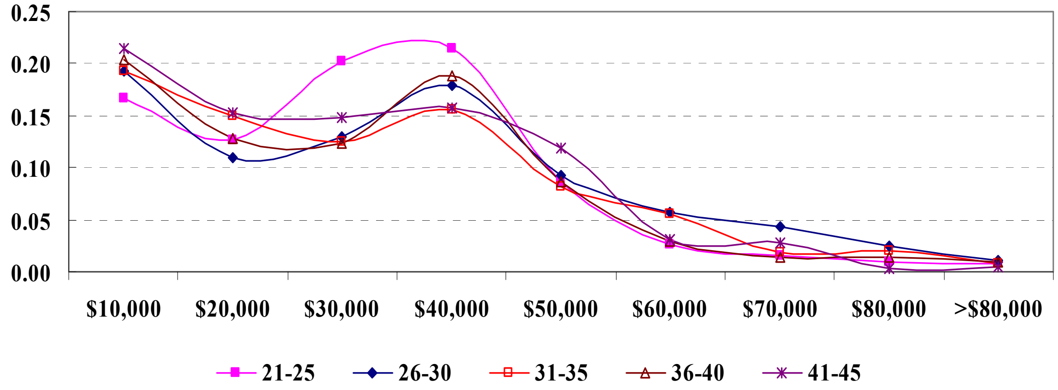


Figure A5.1b: Predicted probability of earnings of bachelors degree qualifications by 21-45 year age groups

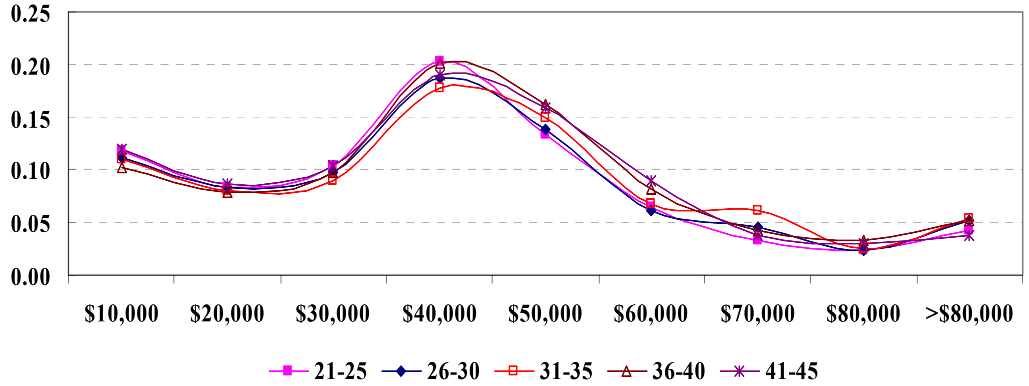


Figure A5.1c: Predicted probability of earnings of postgraduate qualifications by 21-45 year age groups

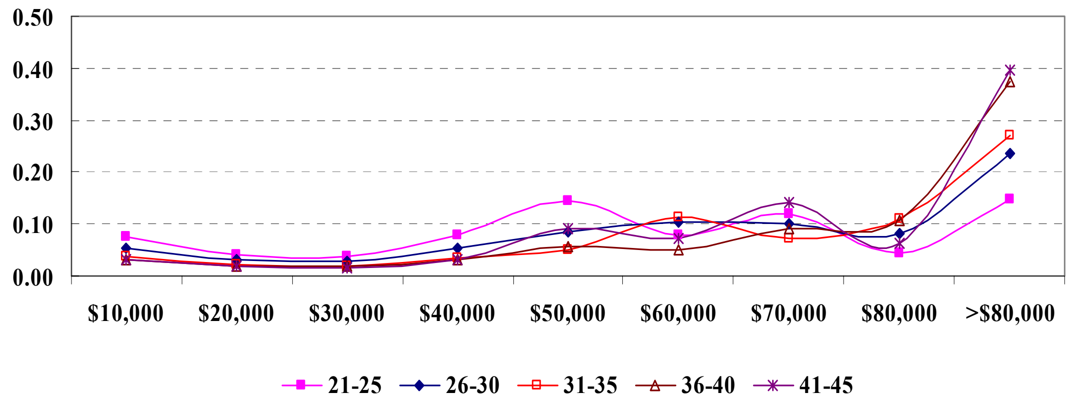


Figure A5.1d: Predicted probability of earnings of sub-degree qualifications by top and bottom age groups

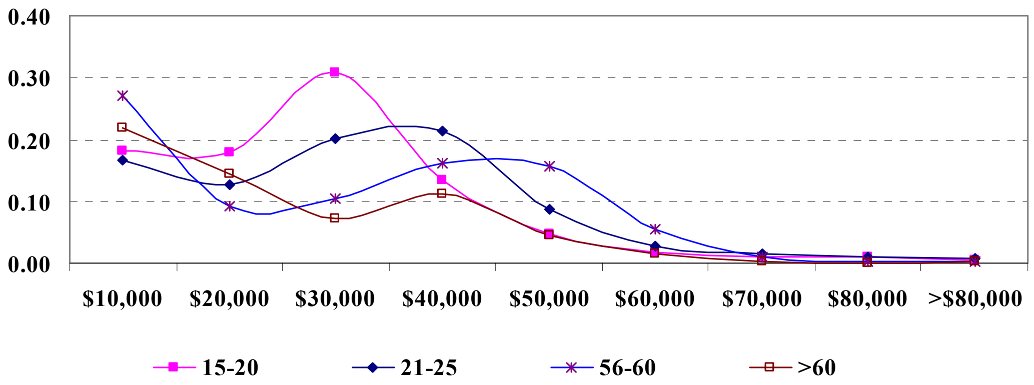


Figure A5.1e: Predicted probability of earnings of bachelors degree qualification by top and bottom age groups

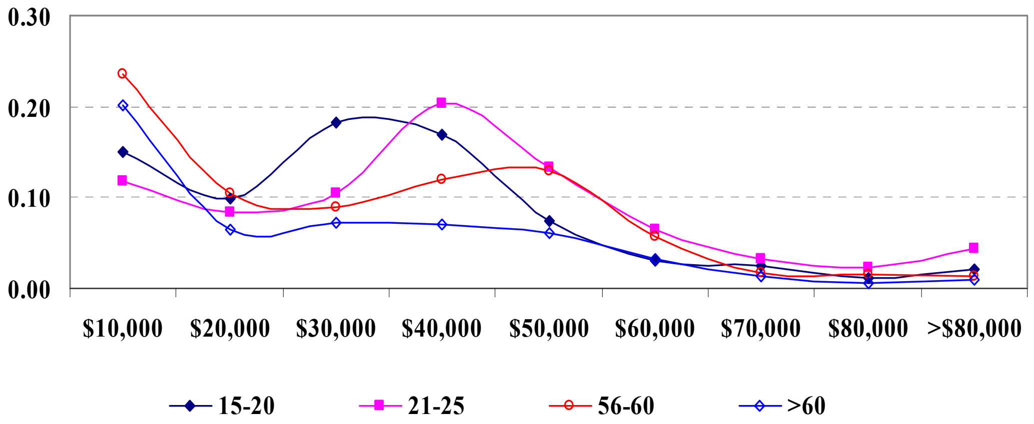


Figure A5.1f: Predicted probability of earnings of postgraduate qualifications by top and bottom age groups

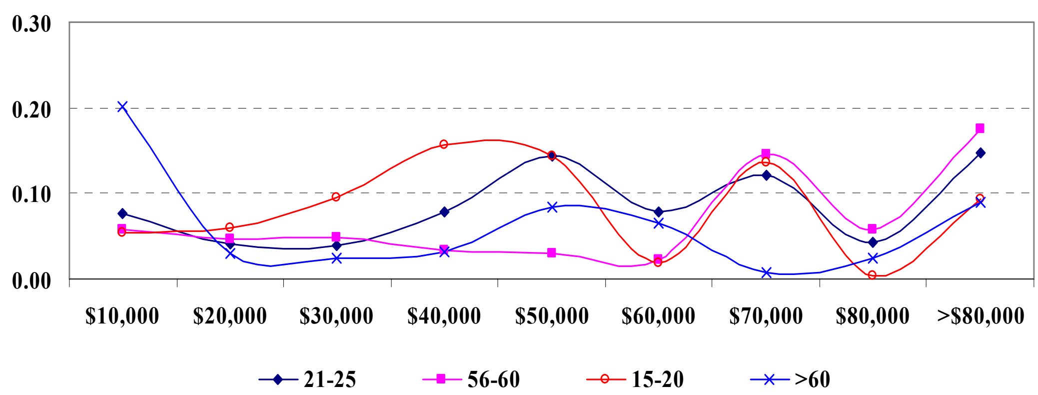


Figure A5.2a: Predicted probability for sub-degree qualifications by industry -1

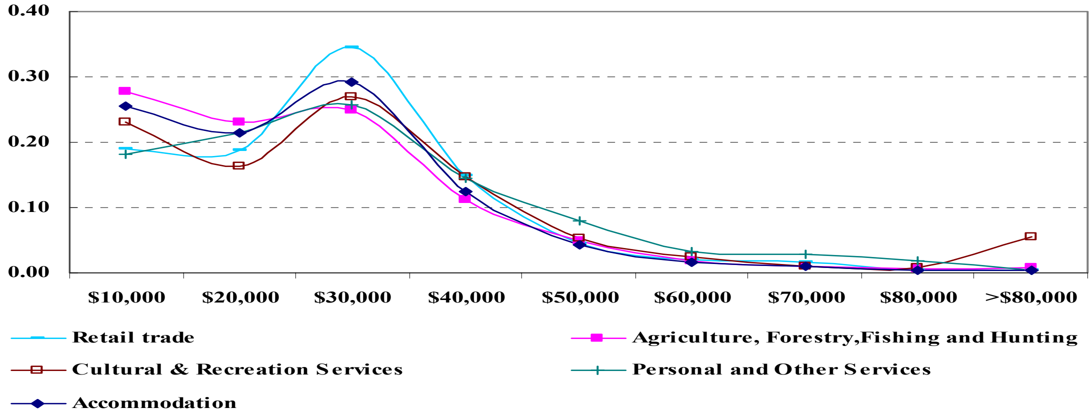


Figure A5.2b: Predicted probability for sub-degree qualifications by industry -2

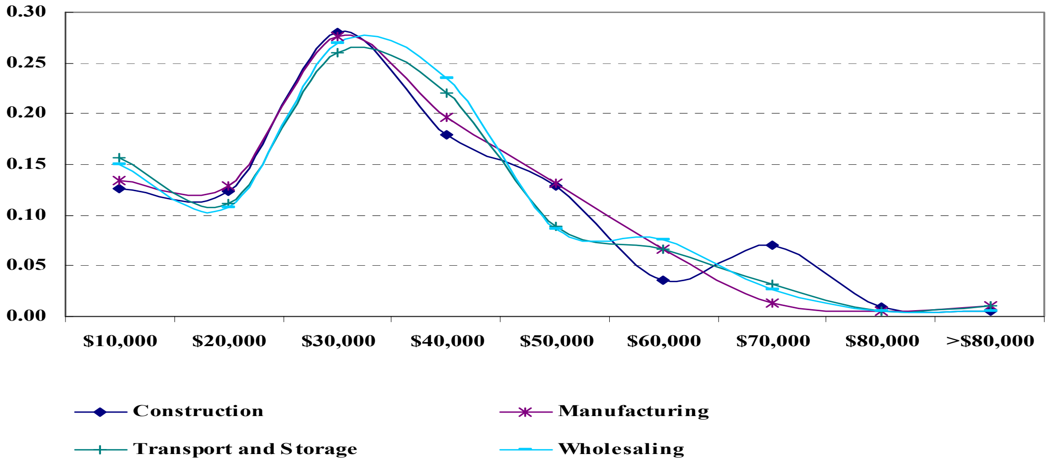


Figure A5.2c: Predicted probability for sub-degree qualifications by industry -3

Figure A5.2d: Predicted probability for bachelors degree qualification by industry – 1

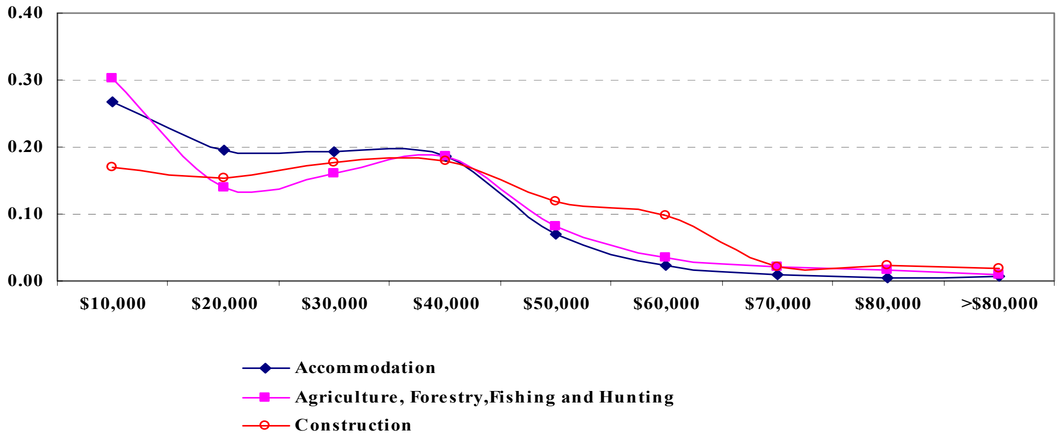


Figure A5.2e: Predicted probability for bachelors degree qualification by industry - 2

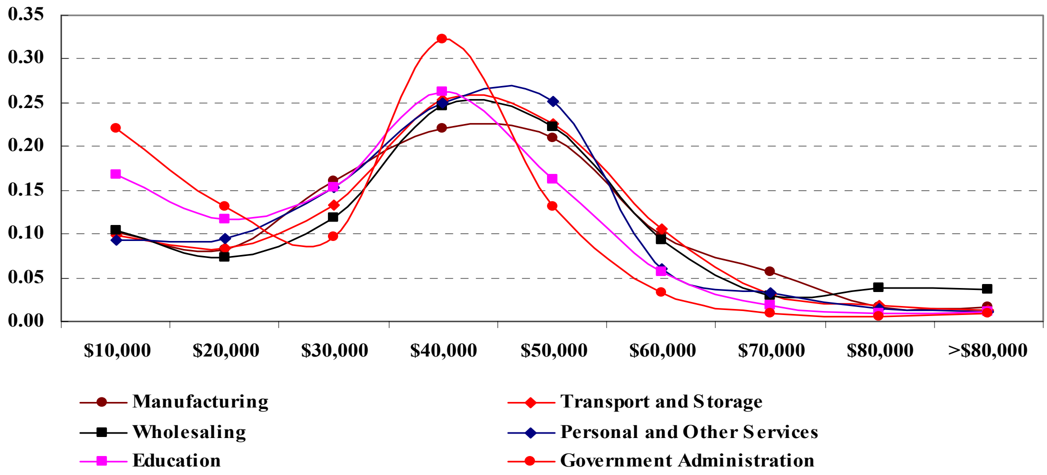


Figure A5.2f: Predicted probability for bachelors degree qualification by industry -3

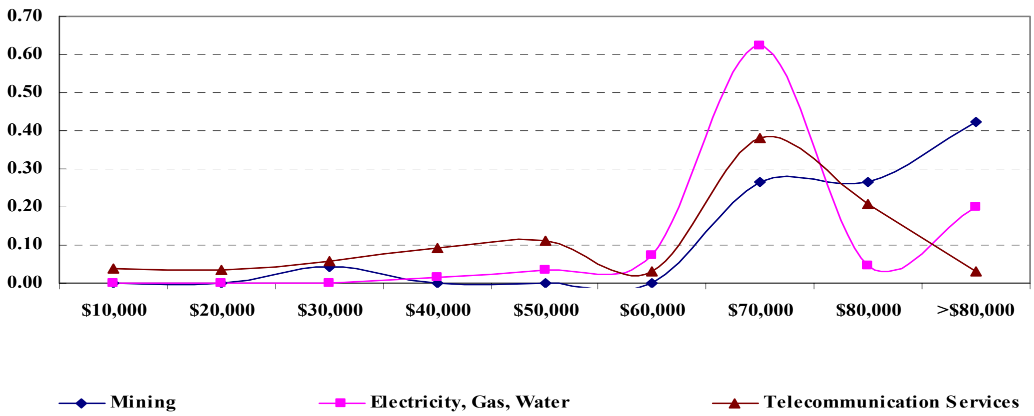


Figure A5.2g: Predicted probability for postgraduate qualifications by industry – 1

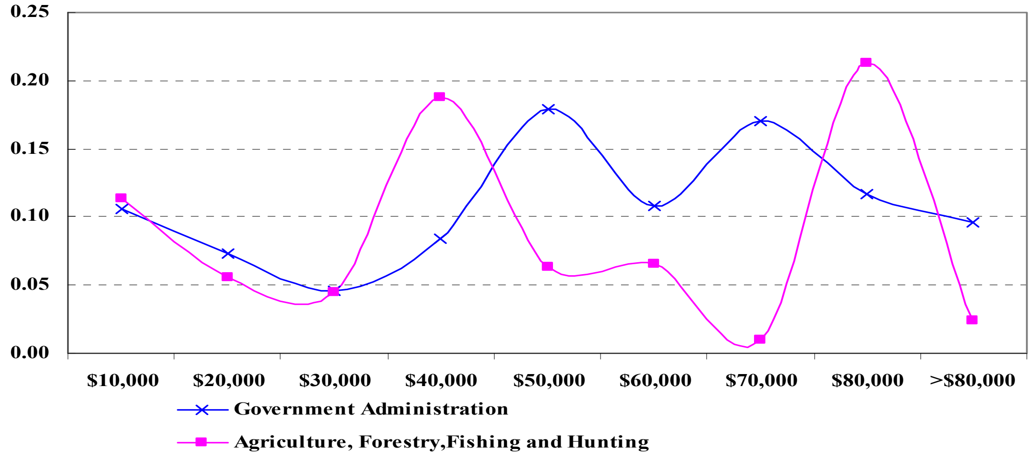


Figure A5.2h: Predicted probability for postgraduate qualifications by industry – 2

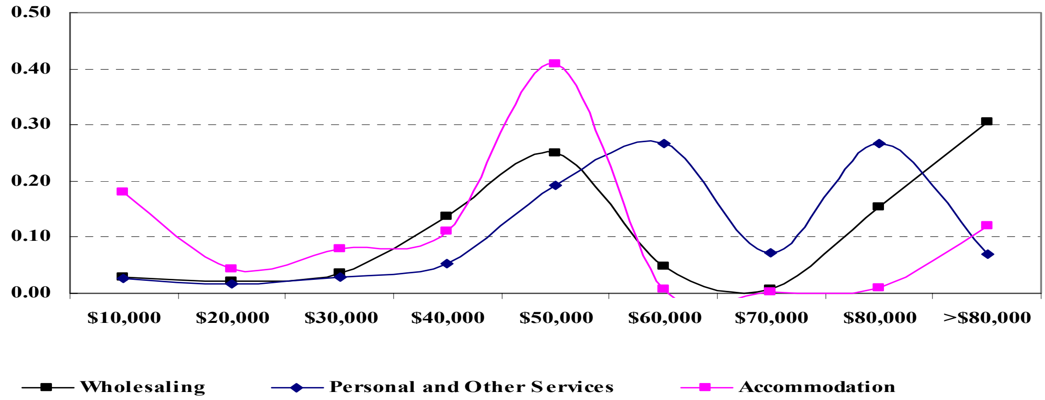
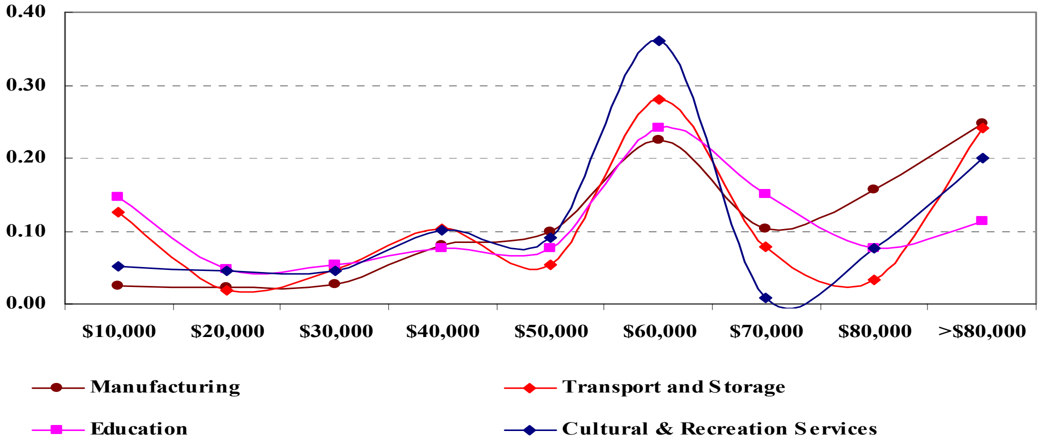


Figure A5.2i: Predicted probability for postgraduate qualifications by industry - 3

Figure A5.2j: Predicted probability for postgraduate qualification by industry - 4



Appendix 6: Correspondence between ethnicity and qualification level

Figure A6.1: Correspondence between ethnicity and qualification level

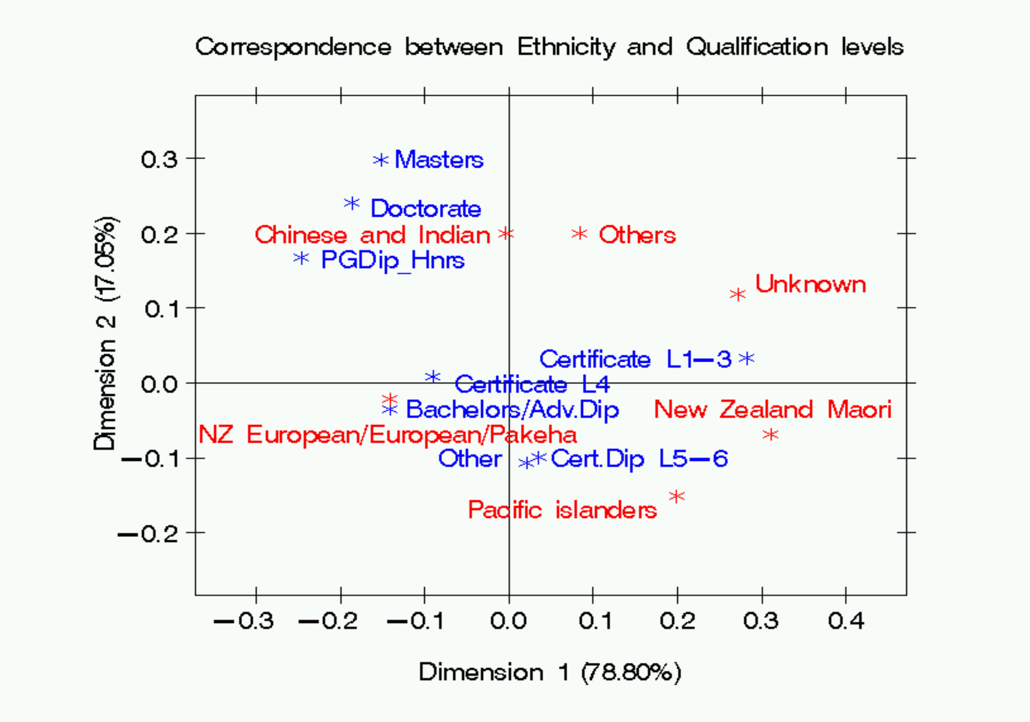


Figure A6.1a: Predicted probability of earnings in bachelors degree by prior activity

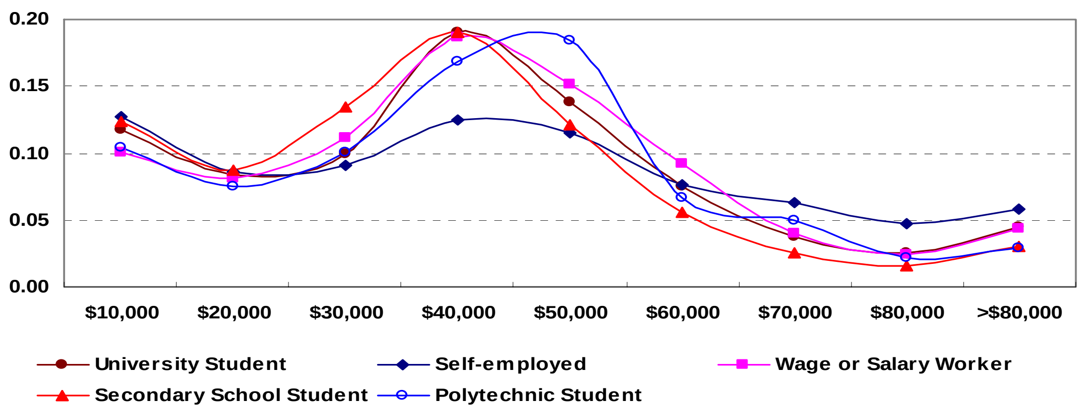


Figure A6.1b: Predicted probability of earnings in postgraduate by prior activity

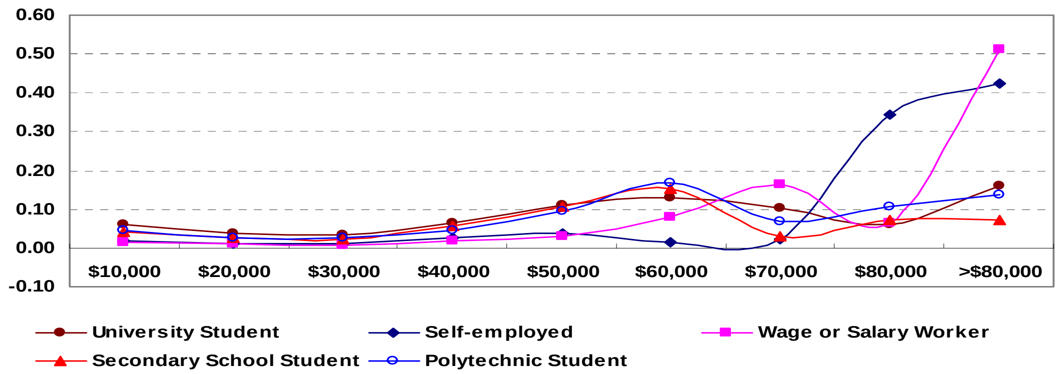
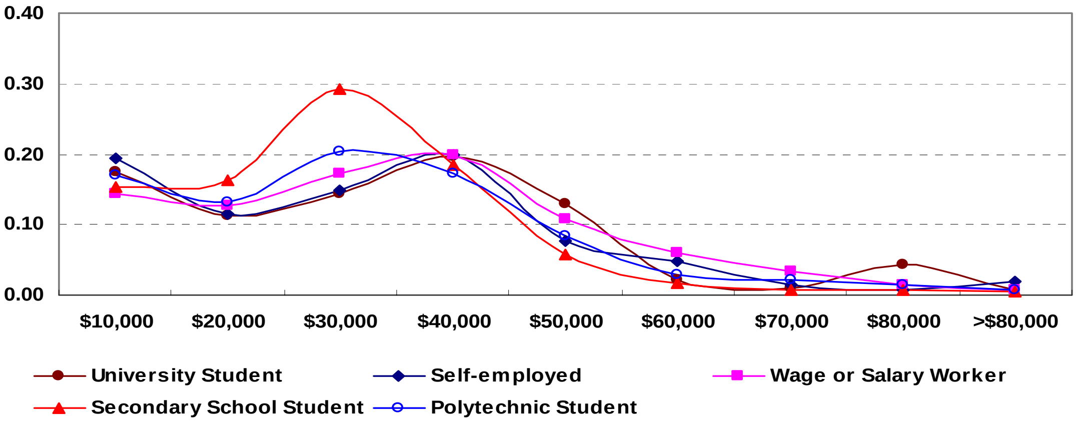


Figure A6.1c: Predicted probability of earnings in sub-degree by prior activity



Appendix 7: Predicted average earnings by field of study and employer industry

Table A7.1: Predicted average earnings by field of study and employer industry

Industry / Fields of Study → ↓	Agriculture ,environmental and related Studies	Architecture and Building	Creative Arts	Education	Engineering and Related Technologies	Food, Hospitality Services	Health	Information Technology	Management and Commerce	Mixed Field Programmes	Natural and Physical Sciences	Society and Culture
Accommodation	16,269	24,603	15,134	23,140	25,676	21,350	19,541	16,110	1,335	15,131	21,010	18,278
Agriculture, Forestry, Fishing and Hunting	25,535	23,530	16,751	24,299	39,707	16,111	20,379	18,685	3,555	9,589	23,555	13,544
Construction	23,183	30,460	25,125	32,529	40,976	23,313	38,893	24,078	9,715	18,657	31,707	24,163
Cultural and Recreation Services	19,739	31,484	25,928	28,823	51,035	20,476	43,635	29,718	51,142	31,015	30,183	27,329
Education	19,117	29,760	25,003	33,074	21,128	20,871	37,457	28,966	37,970	37,330	20,829	27,503
Electricity, Gas and Water Supply	75,000		54,997	41,606	69,357	41,487	55,000	47,149	78,994	65,000	67,700	66,638
Finance and Insurance	74,438	65,001	37,381	34,427	76,092	30,177	41,243	52,938	55,371	62,409	48,373	36,753
Government Administration	21,071	29,067	15,248	32,906	23,889	15,362	27,069	16,129	25,591	14,816	21,365	22,397
Health and Community Services	36,031	74,955	19,387	26,127	31,282	17,321	52,659	29,394	35,356	27,573	29,807	27,606
Manufacturing	29,833	30,488	25,871	28,017	44,408	21,733	44,795	27,893	34,470	23,980	35,177	24,949
Mining	44,976	24,984	54,269	55,000	73,000			65,000	78,335	24,984	68,966	43,852
Personal and Other Services	26,164	43,036	24,480	29,546	26,463	18,799	41,667	22,347	32,329	23,651	35,212	32,070
Property and Business Services	30,159	34,012	25,921	29,507	43,936	20,950	39,796	39,927	39,615	29,315	42,409	38,171
Retail Trade	20,686	22,979	22,182	25,785	28,915	20,486	62,403	21,176	26,770	18,183	23,757	22,071
Telecommunication Services	25,009	43,468	40,683	30,327	74,838	25,567	33,140	39,918	60,188	24,943	72,898	41,102
Transport and Storage	30,783	34,130	23,934	35,832	47,880	26,890	25,957	30,612	36,440	22,656	27,634	28,289
Wholesaling	34,107	31,846	24,949	27,452	40,359	21,225	36,772	34,778	38,066	23,930	34,306	27,615

Table A7.2: Distribution of students three years post-study employed in different employer industries (%)

Industry /Fields of study → ↓	Agriculture	Architecture and Building	Creative Arts	Education	Engineering and Related Technologies	Food and Hospitality	Health	Information Technology	Management and Commerce	Mixed Field	Natural and Physical Sciences	Society and Culture
Accommodation	2.10	1.00	12.60	5.34	1.52	32.82	2.13	2.22	14.23	3.07	4.08	18.89
Agriculture, Forestry, Fishing and Hunting	26.46	1.52	13.28	5.00	2.53	7.40	2.02	2.16	11.83	2.79	6.12	18.90
Construction	4.63	5.81	34.07	2.32	12.52	4.83	0.71	2.91	12.50	1.83	5.89	11.97
Cultural and Recreation Services	1.57	2.52	22.05	7.57	1.67	5.84	1.94	2.44	13.24	1.87	8.08	31.22
Education	0.85	0.85	6.99	27.42	2.03	1.56	2.75	3.55	10.78	3.06	10.54	29.61
Electricity, Gas and Water Supply	3.30	0.00	5.42	2.24	16.39	5.78	1.18	6.13	23.82	1.53	15.33	18.87
Finance and Insurance	2.61	0.50	3.99	4.34	2.67	5.21	0.74	4.22	43.91	1.51	7.76	22.54
Government Administration	1.38	1.25	6.92	28.25	1.53	2.73	1.70	3.30	11.64	4.15	7.55	29.61
Health and Community Services	2.62	0.27	2.87	10.26	0.37	1.58	40.28	1.03	6.50	2.82	4.62	26.79
Manufacturing	4.55	1.31	26.25	2.28	10.39	4.88	1.49	4.00	18.14	3.56	8.73	14.43
Mining	2.77	2.77	25.21	2.77	14.40	0.00	0.00	2.77	24.38	2.77	16.90	5.26
Personal and Other Services	2.73	1.09	7.80	8.89	0.54	8.24	3.41	2.73	14.20	2.20	5.81	42.38
Property and Business Services	2.65	3.32	9.93	4.61	6.21	4.11	2.03	5.48	26.13	2.58	9.26	23.70
Retail Trade	3.27	1.24	20.10	4.72	2.31	11.10	4.08	4.37	19.05	3.64	6.66	19.47
Telecommunication Services	1.79	1.13	7.50	4.95	7.06	4.80	1.20	6.50	30.41	3.02	8.24	23.39
Transport and Storage	3.41	1.64	10.04	3.73	5.15	17.34	0.57	3.44	23.83	2.88	5.47	22.51
Unknown	4.14	1.98	11.98	4.83	5.52	6.93	6.61	4.05	22.60	2.80	8.74	19.83
Wholesaling	4.28	1.84	15.67	3.28	6.64	4.63	1.39	5.67	31.04	2.49	8.30	14.77
Total	3.35	1.62	10.96	12.17	3.67	6.08	6.17	3.59	17.36	3.09	7.70	24.24

Appendix 8: Correspondence between level of study and income

Figure A8.1: Correspondence between level of study and income

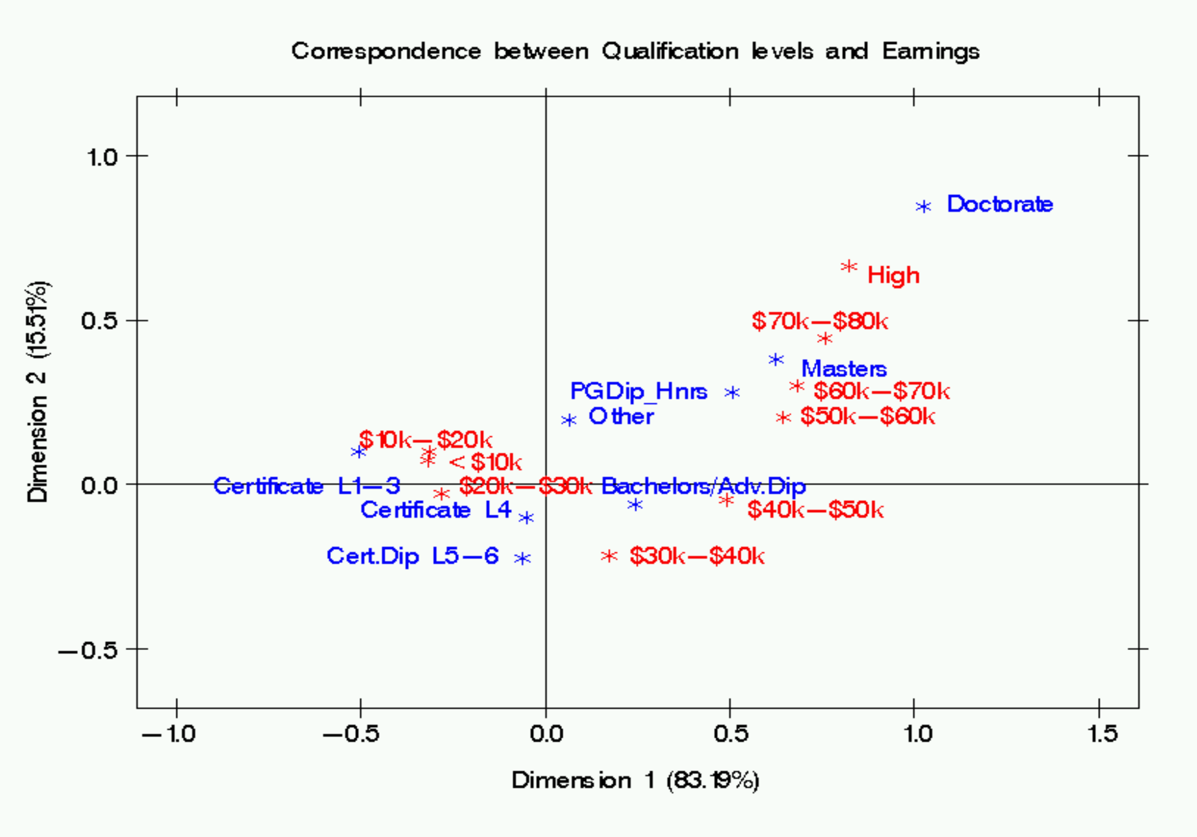


Figure A8.2: Correspondence between field of study and level of study

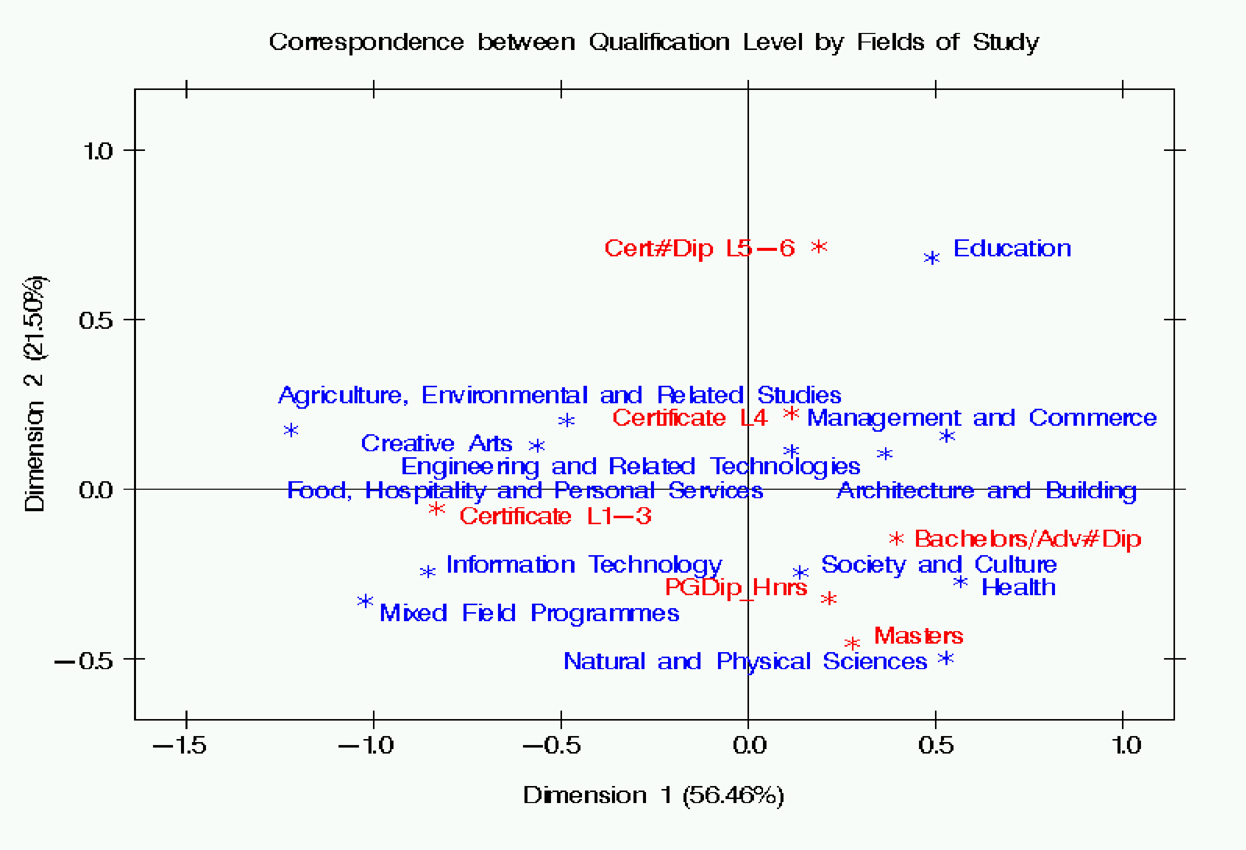


Figure A8.3: Correspondence between field of study and industry

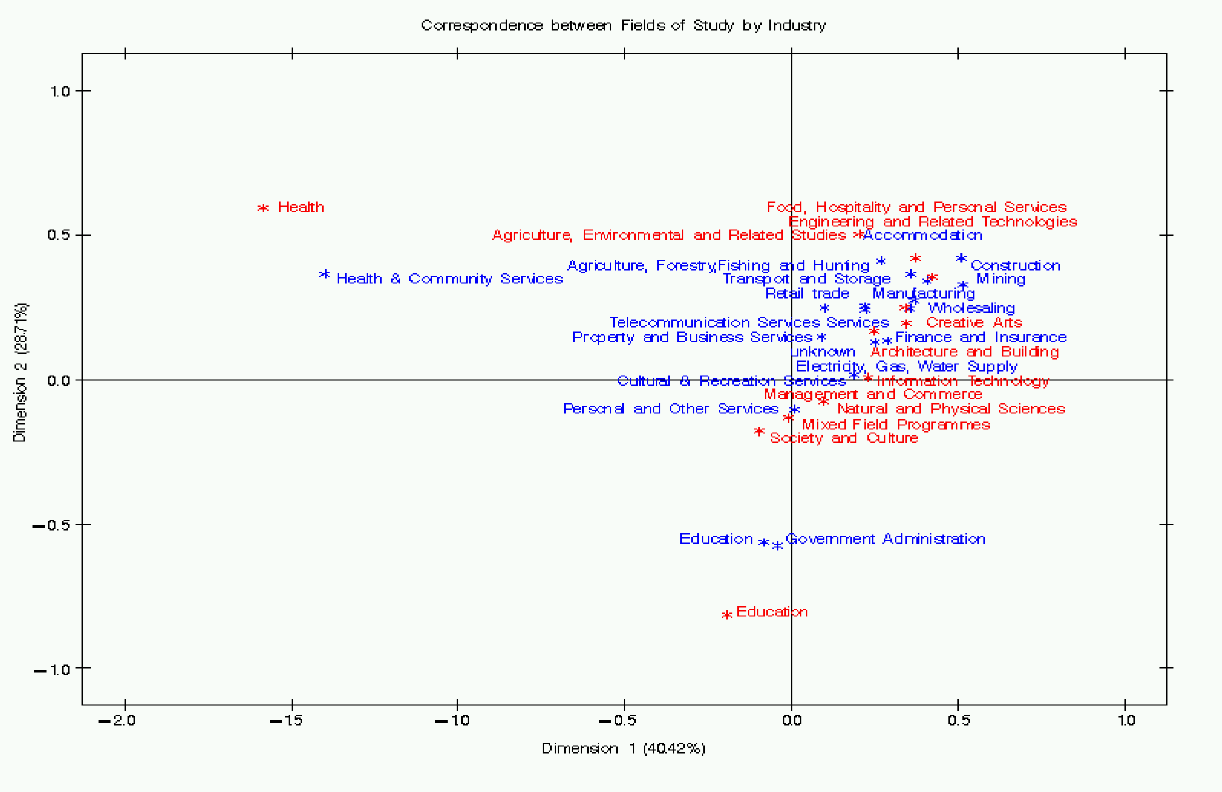
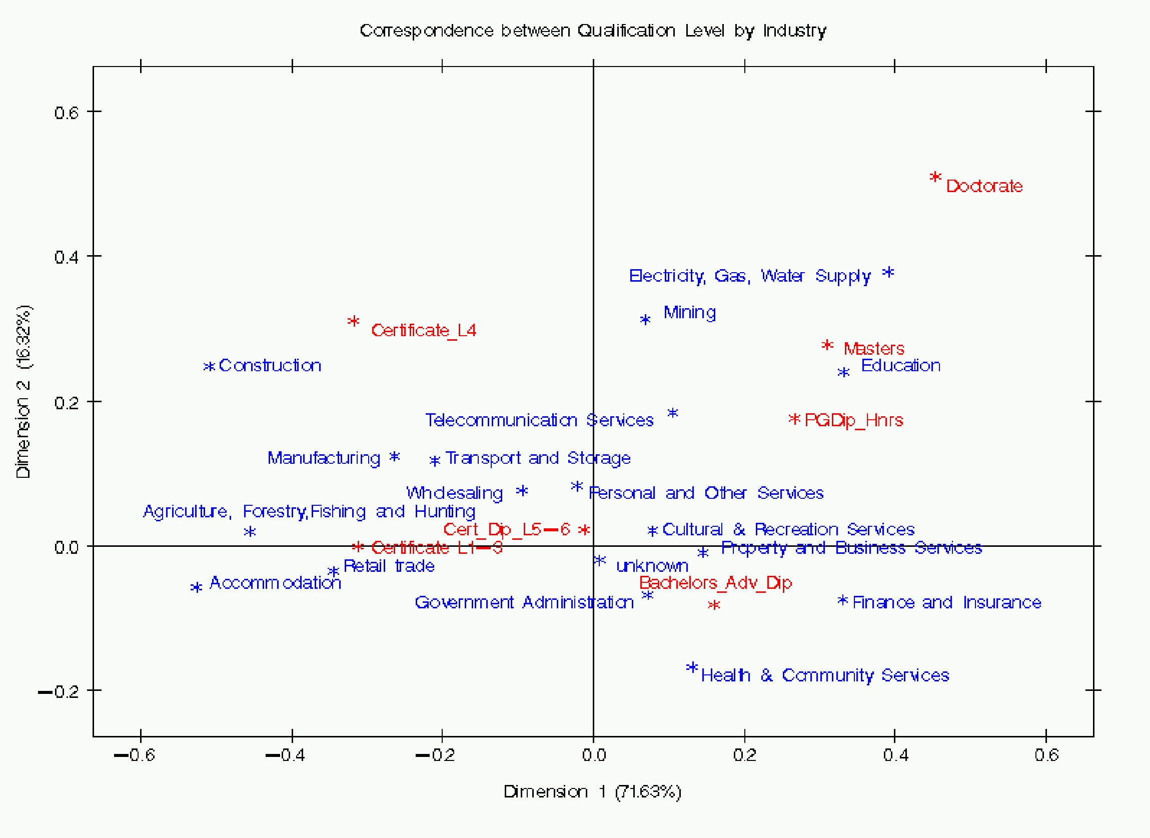


Figure A8.4: Correspondence between level of study and industry



Appendix 9: Predicted probability of earnings in sub-degree qualifications by provider sector

Figure A9.1: Predicted probability of earnings in undergraduate qualifications by provider sector

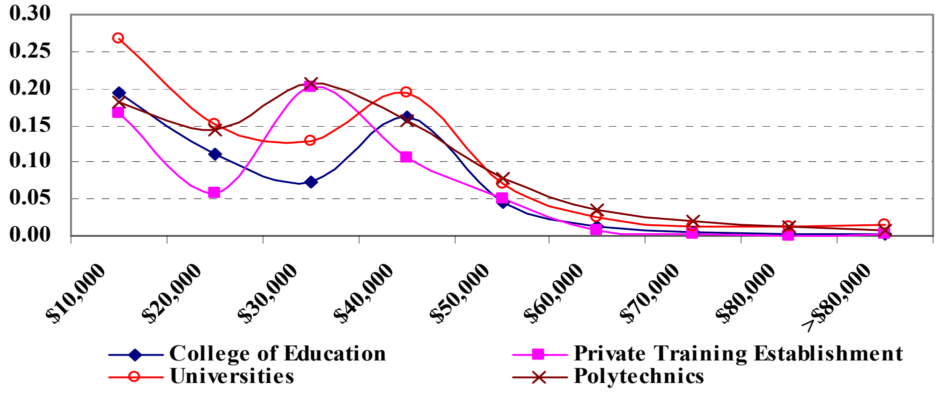


Figure A9.2: Predicted probability of earnings in bachelors qualifications by provider sector

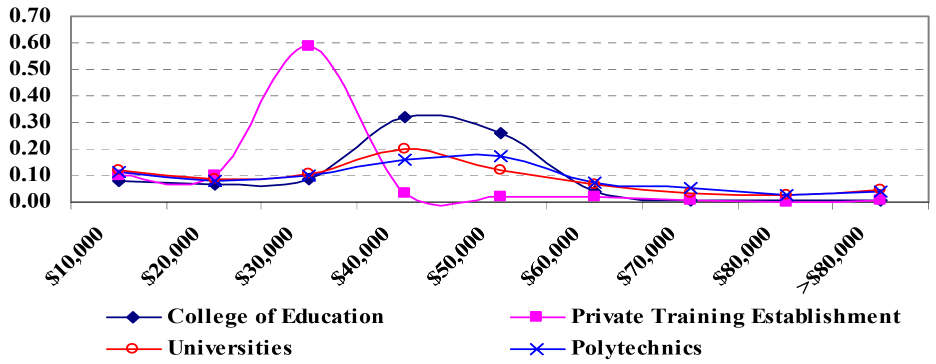
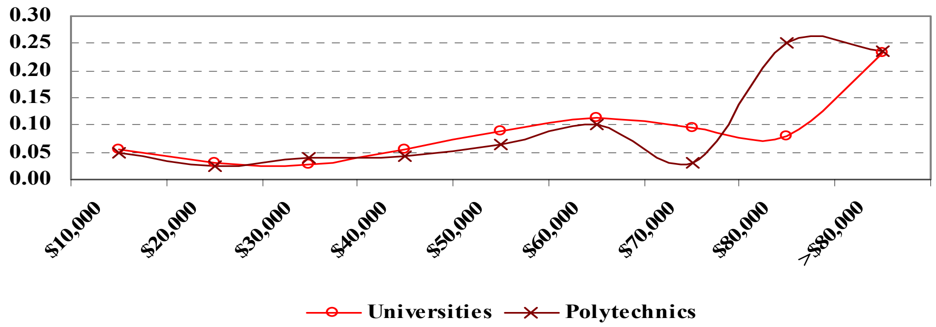


Figure A9.3: Predicted probability of earnings in postgraduate qualifications by provider sector



References

Adelman, C. (2006). *The Toolbox Revisited: Paths to Degree Completions from High School through College*. Washington, DC: US Department of Education.

Boden, Jr., R. J. (1999). "*Flexible working hours, family responsibilities, and female self-employment*". American Journal of Economics and Sociology 58: 71-83.

Boissiere, M., J. B. Knight and R. H. Sabot (1985). "*Earnings, schooling, ability, and cognitive skills*". The American Economic Review 75, 5 (December): 1016-1030.

Borland, J. (2000). *Economic Explanations of Earnings Distribution Trends in the International Literature and Application to New Zealand*. Treasury Working Paper 00/16.

Card, D. (1999). "*The causal effect of education on earnings*". In Handbook of Labor Economics, Vol. 3A. Edited by O. Ashenfelter and D. Card. Amsterdam: Elsevier Science.

Charette, M. F. and R. Meng (1998). "*The determinants of literacy and numeracy, and the effect of literacy and numeracy on labour market outcomes*". Canadian Journal of Economics 31, 3 (August): 495-517.

Gibson, J. (2000). "*Sheepskin effects and the returns to education in New Zealand: do they differ by ethnic groups?*" New Zealand Economic Papers, December 01.

Griliches, Z. (1977). "*Estimating the returns to schooling: some econometric problems*". Econometrica 45, 1 (January): 1-22.

Heckman, J. (1978). "*A partial survey of recent research on the labor supply of women*". American Economic Review 68: 200-207.

Heijden, P. G. M. and J. Leevw (1985). "*Correspondence analysis using complementary to loglinear analysis*". Psychometrika 50, 4 (December).

Hosmer, D. and S. Lemeshow (2000). *Applied Logistic Regression*. Second edition, New York: Wiley & Sons.

Hyatt, J., P. Gini and R. Smyth (2005). *Income of Student Loan Scheme Borrowers*. Wellington: Ministry of Education.

Hyatt, J. and R. Smyth (2006). *How Do Graduates' Earnings Change over Time*? Wellington: Ministry of Education.

Le, T., J. Gibson and L. Oxley (2005). *Measures of Human Capital*: A Review of the Literature. http://www.treasury.govt.nz/workingpapers/2005/wp05-10.asp

Maani, S.A. and T. Maloney (2004). *Returns to Post-school Qualifications: New Evidence Based on the HLFS* Income Supplement (1997-2002). Report to the Labour Market Policy Group. Wellington: Department of Labour.

Mare, D. and L. Yun (2006). *Labour Market Outcomes for Young Graduates*. Presentation at Department of Labour, Wellington, March 2006.

Murnane, R. J., J. B. Willett and F. Levy (1995). "*The growing importance of cognitive skills in wage determination*". The Review of Economics and Statistics 77, 2: 251-266.

Penny, N. (2005). *The Approach to Measuring the Returns to Secondary and Tertiary Qualifications in New Zealand*: An Investigation and Update Using Data from the 2001 Census. Unpublished thesis.

Riddell, W. C. and S. Arthur (2006). *The Impact of Education on Economic and Social Outcomes*: An Overview of Recent Advances in Economics. Workshop on An Integrated Approach to Human Capital Development, Ottawa.

Rosen, S. (1977). "*Human Capital*: *A Survey of Empirical Research*". In Research in Labor Economics: An Annual Compilation of Research, Vol. 1. Edited by Ronald G. Ehrenberg. Greenwich, CT: JAI Press.

Smart, W. (2004). *Resourcing Trends in New Zealand Tertiary Education Institutions 1991-2003*. Wellington: Ministry of Education.

Statistics New Zealand (2003). *Human Capital Statistics 2003*, Part 4 – Participation in Education and Training. http://www.stats.govt.nz/analytical-reports/human-capital-statistics/part-4-participation-in-education-and-training.htm

Steiner, M. and Teszler, N. (2005). *Multivariate Analysis of Student Loan Defaulters at Texas A&M University*. Texas Guaranteed Student Loan Corporation Report. Austin, TX.

Stiglitz, J. (1975). "*The theory of 'screening', education, and the distribution of income*". American Economic Review 65: 282-300.

Wermuth, N. and D. R. Cox (2001*). Separation and Joint Response Graphs Induced by Triangular Systems*. Research report. Australian National University. http://www.maths.anu.edu.au/research.reports

Willis, R. J. (1986). "*Wage determinants: a survey and reinterpretation of human capital earnings functions*". In Handbook of Labor Economics, Vol. 1. Edited by O. Ashenfelter and R. Layard. Amsterdam: Elsevier Science.

Vella, F. (1994). *Gender roles and human capital investment: the relationship between traditional attitudes and female labour market performance*. *Economica*, New Series, Vol. 61, No. 242, pp. 191-211.